A Physical Deterministic Inverse Method for Operational Satellite Remote Sensing: An Application for Sea Surface Temperature Retrievals

Prabhat K. Koner, *Member, IEEE*, Andrew Harris, and Eileen Maturi

***Abstract*—We propose a new deterministic approach for remote sensing retrieval, called modified total least squares (MTLS), built upon the total least squares (TLS) technique. MTLS implicitly determines the optimal regularization strength to be applied to the normal equation first-order Newtonian retrieval using all of the noise terms embedded in the residual vector. The TLS technique does not include any constraint to prevent noise enhancement in the state space parameters from the existing noise in measurement space for an inversion with an ill-conditioned Jacobian. To stabilize the noise propagation into parameter space, we introduce an additional empirically derived regularization proportional to the logarithm of the condition number of the Jacobian and inversely proportional to the L2-norm of the residual vector. The derivation, operational advantages and use of the MTLS method are demonstrated by retrieving sea surface temperature from GOES-13 satellite measurements. An analytic equation is derived for the total retrieval error, and is shown to agree well with the observed error. This can also serve as a quality indicator for pixel-level retrievals. We also introduce additional tests from the MTLS solutions to identify contaminated pixels due to residual clouds, error in the water vapor profile and aerosols. Comparison of the performances of our new and other methods, namely, optimal estimation and regression-based retrieval, is performed to understand the relative prospects and problems associated with these methods. This was done using operational match-ups for 42 months of data, and demonstrates a relatively superior temporally consistent performance of the MTLS technique.**

***Index Terms*—Condition number of matrix, ill-conditioned inverse methods, regularization, satellite remote sensing, sea surface temperature (SST), total error, total least squares (TLS).**

I. Introduction

PARAMETER estimation from remotely sensed satellite measurements can be broadly categorized into two groups: a) the direct method, which directly correlates the measurements with geophysical parameter/s without considering a proper physical model; and b) multivariable parameter estimation using some approximation of radiative transfer (RT) physics, which is usually referred to as the physical method. A direct method (e.g., regression) is generally used when the number of spectral measurements is limited, where the effects of RT physics are heavily approximated. This is often inadequate for characterizing the full-range of variability of global geophysical conditions (atmospheric and oceanic states). Thus, some corrections on the measurement often become necessary to improve such result (by means of altering the measurement values) that is done either statistically or by RT modeling using *a priori* ancillary data [1]. The geophysical parameters are too dynamic, and *a priori* data-based corrections may introduce a high error in the retrievals, which are further used as another set of *a priori* data for other applications. As a consequence, there can be a cumulative effect of these errors on higher level products.

The main challenge of multivariable parameter estimation employing RT physics is that it is an ill-posed inverse problem. Conceivably, there exists no unique solution according to the review of retrieval theories in remote sensing (e.g., [2]–[4]). A large number of approaches have been developed to find a solution, all of which have their inherent limitations and assumptions. Most of the "physical" techniques for satellite remote sensing applications are in fact stochastic/probabilistic approaches. Such techniques differ from each other both in the procedure for solving a set of spectrally independent RT equations (e.g., matrix inversion, numerical iteration) and in the choice of ancillary data. These ancillary data are used to constrain the solution (e.g., atmospheric covariance statistics and *a priori* estimate of the retrieved parameters), which may introduce errors from ancillary sources. Therefore, it is critically important to establish and implement a physical deterministic inverse method for operational satellite applications.

There are two distinct schools of thoughts in the development of physical inverse model. i) One school believes that there is a true value of all individual retrieved parameter, but associated with error root-mean-square error (RMSE) and these methods can be derived at individual measurement points. ii) The other school assumes that all retrieved parameter values are uncertain and as a result these parameters are stated in the expected value and uncertainty of retrieval standard deviation (SD). As per

Manuscript received August 19, 2014; revised January 24, 2015 and March 5, 2015; accepted April 8, 2015. This work was supported by the NESDIS Product Systems Development and Implementation program through NOAA Grant NA09ES4400006 (Cooperative Institute for Climate and Satellites) at the University of Maryland/ESSIC.

P. K. Koner and A. Harris are with the NOAA Satellite and Information Service (NESDIS), Center for Satellite Applications and Research (STAR), College Park, MD 20740 USA, and also with Earth System Science Interdisciplinary Centre, Cooperative Institute of Climate Sciences, University of Maryland, College Park, MD 20740 USA.

E. Maturi is with the NOAA Satellite and Information Service (NESDIS), Centre for Satellite Applications and Research (STAR), College Park, MD 20740 USA.

Digital Object Identifier 10.1109/TGRS.2015.2424219

0196-2892 © 2015 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the present literature survey, we shall refer to the first one as "deterministic" and the second one as the "stochastic" approach [5]–[7]. This work compares the outcomes of these two different approaches after bringing them to a common platform, i.e., the deterministic approach. Using the deterministic paradigm, the expected value of the parameter in the stochastic approach is combined with the mean error and true state of the parameter. The mean error is often known as the bias term. Therefore, $\mathrm{RMSE} = \sqrt{(\mathrm{bias}^2 + \mathrm{SD}^2)}$, which is the reported total error (RMSE) in the deterministic approach.

One of the traditional deterministic approaches is least squares (LS) method, which is well understood and was first introduced by Legendre in1805 [8]. However, it does not work well in real life problems (linear or nonlinear) when the problem is ill-conditioned. For an ill-conditioned problem the error propagated from measurement space to state space is multiplied by the condition number of the matrix [9], [10]. Many real life applications are ill-conditioned problems. The need to constrain the problem by other assumptions has resulted in a number of different inverse methods. The bases of real life applied deterministic and stochastic approaches are regularization/constrained (e.g., [11]–[17]) at a single measurement point and Bayesian probability theory (e.g., [2], [18]–[20]) for a set of measurement, respectively. Parameter estimation, in general, can be viewed as an optimization problem, in which an objective function representing data misfit (measurement error) is minimized in a given norm [12]. From a deterministic viewpoint, the two-norm (i.e., a quadratic objective function) is mathematically attractive because the minimum can be explicitly written in a closed form. In the stochastic approach, this choice of two-norm is statistically the most likely solution if data are normally distributed, however, this estimate is typically less accurate if the data are not normally distributed or have outliers [20].

A significant difference between methods and their solutions lies in how the weight (regularization strength) is chosen. It is often stated that the stochastic approach is more accurate than the deterministic approach because it better weights the data and parameter misfits using corresponding covariance matrices. On the other hand, most traditional deterministic approaches such as positivity constraints or regularization only use constraints or simple weights on the parameter misfit [20], but do not include the data misfit information. Thus, in practice, solving atmospheric inverse problems is dominated by optimal estimation method (OEM), which is in stochastic family. However, the weight of the data misfit embedded in the mathematical formulation of OEM for inherently nonlinear RT problems presents its own issues. For example, the weight of the residual by the data misfit covariance matrix in this approach may not improve optimization results because the contribution from the nonlinear term is also weighted. In contrast, recent deterministic approaches have considered the fidelity of the data or the data misfit criterion [10], [14], [15] on the determination of the optimal regularization strength. However, some deterministic methods have been established on an experimental basis using simulations, and there are no popular deterministic techniques implemented for an operational environment. One of the primary aspects of this paper is to establish the benefit of the physical deterministic method in overcoming many operational hurdles.

One of the inputs for stochastic methods is the measurement error covariance matrix, and it is difficult to construct a properly representative one for any satellite retrieval because of errors in the instrument, the forward model, RT spectral coefficients, and due to nonlinearity. Additionally, near-real-time processors in remote sensing frequently use high approximations of the Jacobian matrix in order to speed up the calculation [21], [22]. Among the existing deterministic and stochastic methods, Total Least Squares (TLS) is the only one, which inherently accounts for the Jacobian error in its optimization, since it calculates the data misfit criterion and feeds it into its regularization scheme [23]–[26]. In addition to the theoretical discussion about inverse problems, we will discuss in this paper various operational problems as well as demonstrate an evidence of successful sea surface temperature (SST) retrievals from GOES-13 (Geostationary Operational Environmental Satellite) measurements using a modified TLS (MTLS) method.

Although the MTLS method can generally be applied for any inversion problem, we have chosen satellite SST retrievals to demonstrate the technique because of the availability of suitable high-quality reference data sets (buoy temperatures) required for validation. Most SST retrievals from satellite measurements are still performed using regression-based methods and, over the years, there has been only limited progress. Such approaches were justifiable in the interest of time and computational resources when they were formulated [27], [28] at the cost of highly approximated RT physics. With the availability of improved computational facilities, render the development of a deterministic physical approach, such as the MTLS method, both feasible and desirable. In order to demonstrate its advantages, this paper also compares retrievals from the proposed MTLS, OEM and the traditional regression-based formulation.

II. Theoretical Background of MTLS

Although TLS has first been used only in recent decades [17], [29], this fitting method has a longer history in the statistical literature and has been known under different nomenclature, e.g., orthogonal regression or errors-in-variables. For instance, in 1877 Adcock [30] employed this approach for univariate problem. However, despite its long use in statistical inversion, a deterministic version of TLS has only recently been derived using linear algebra, which will be recapped here. The TLS method [31] is based on minimization of the errors in the measurement $(\delta \boldsymbol{y})$ and the Jacobian $(\delta \boldsymbol{K})$ at the individual pixel level as follows:

$$\min_{\|\delta \mathbf{K}\|, \|\delta \mathbf{y}\|, \mathbf{x}} \|\delta \boldsymbol{K}\|^2 + \|\delta \boldsymbol{y}\|^2 \text{ subject to}$$

$$(\boldsymbol{K} - \delta \boldsymbol{K})x = \boldsymbol{y}_\delta - \delta \boldsymbol{y} \quad (1)$$

where $\|.\|$ is the second-order mathematical norm and $\boldsymbol{x}$ is the vector of state space parameters. The terms K, $\boldsymbol{y}_\delta$, $\delta \boldsymbol{K}$ and $\delta \boldsymbol{y}$ are the Jacobian and measurement vector (both including error), and errors in the Jacobian and the measurements, respectively. The basic assumptions are that any real-world inverse problem comprises errors in Jacobian and measurement, and that the theoretical equation $\boldsymbol{y}_\delta - \delta\boldsymbol{y} = (\boldsymbol{K} - \delta\boldsymbol{K})\boldsymbol{x}$ can be formed after subtraction of these errors from $\boldsymbol{K}$ and $\boldsymbol{y}_\delta$. It involves three-parameter minimization processes of $\delta\boldsymbol{K}$, $\delta\boldsymbol{y}$ and $\boldsymbol{x}$. First, we consider the inner minimization with two parameters $\delta\boldsymbol{K}$ and $\delta\boldsymbol{y}$ in the norm space to reduce it to a one-parameter minimization problem

$$\min_{\|\delta\mathbf{K}\|,\|\delta\mathbf{y}\|} \|\delta\boldsymbol{K}\|^2 + \|\delta\boldsymbol{y}\|^2 \text{ subject to}$$
$$\|(\boldsymbol{K} - \delta\boldsymbol{K})\boldsymbol{x}\| = \|\boldsymbol{y}_\delta - \delta\boldsymbol{y}\| \quad (2)$$

The Lagrangian function of (2) can be written as

$$L(\delta\boldsymbol{K}, \delta\boldsymbol{y}, \lambda) = \|\delta\boldsymbol{K}\|^2 + \|\delta\boldsymbol{y}\|^2 + \lambda\left\{\|(\boldsymbol{K} - \delta\boldsymbol{K})\boldsymbol{x}\| - \|(\boldsymbol{y}_\delta - \delta\boldsymbol{y})\|\right\} \quad (3)$$

where L is the Lagrangian function and λ is the Lagrange multiplier. Applying Karush-Kuhn-Tucker (KKT) conditions [32], [33] to (3), it can be formulated as

$$2\|\delta\boldsymbol{K}\| - \lambda\|\boldsymbol{x}\| = 0 \ \left(\nabla_{\|\delta\boldsymbol{K}\|} L = 0\right) \quad (4)$$
$$2\|\delta\boldsymbol{y}\| + \lambda = 0 \ \left(\nabla_{\|\delta\boldsymbol{y}\|} L = 0\right) \quad (5)$$
$$\|(\boldsymbol{K} - \delta\boldsymbol{K})\boldsymbol{x}\| = \|\boldsymbol{y}_\delta - \delta\boldsymbol{y}\| \ (\nabla_\lambda L = 0) \quad (6)$$

where ∇ is the partial derivative of the Lagrangian function. Combining (4) and (5), the relation of $\delta\boldsymbol{K}$ and $\delta\boldsymbol{y}$ can be derived as

$$\|\delta\boldsymbol{K}\| = -\|\delta\boldsymbol{y}\| \, \|\boldsymbol{x}\|. \quad (7)$$

Similarly, combining (6) and (7), δy can be separated out as

$$\|\delta\boldsymbol{y}\| = \frac{\|\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x}\|}{1 + \|\boldsymbol{x}\|^2} \quad (8)$$

$\|\delta\boldsymbol{K}\|$ can be determined by substituting the expression of $\|\delta\boldsymbol{y}\|$ in (8) into (7)

$$\|\delta\boldsymbol{K}\| = -\frac{\|(\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x})\| \ \|\boldsymbol{x}\|}{1 + \|\boldsymbol{x}\|^2}. \quad (9)$$

We now replace the $\|\delta\boldsymbol{y}\|$ [(8)] and $\|\delta\boldsymbol{K}\|$ [(9)] into (2) and the three-parameter minimization reduces to a one-parameter minimization as

$$\min_{\|\delta\boldsymbol{K}\|,\|\delta\boldsymbol{y}\|,\boldsymbol{x}} \|\delta\boldsymbol{K}\|^2 + \|\delta\boldsymbol{y}\|^2 = \left\{-\frac{\|(\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x})\| \, \|\boldsymbol{x}\|}{1 + \|\boldsymbol{x}\|^2}\right\} + \left\{\frac{\|\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x}\|}{1 + \|\boldsymbol{x}\|^2}\right\}^2 = \min_{\boldsymbol{x}} \frac{\|\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x}\|^2}{1 + \|\boldsymbol{x}\|^2}. \quad (10)$$

The well-known first-order necessary condition [34], [35] for an exact solution for linear function $(f(\boldsymbol{x}))$ of (10) can be formulated as

$$\boldsymbol{x} = \left(\boldsymbol{K}^T\boldsymbol{K} + g(\boldsymbol{x})\boldsymbol{I}\right)^{-1} \boldsymbol{K}^T\boldsymbol{y}_\delta \quad (11)$$

where, $g(\boldsymbol{x}) = ((\|\boldsymbol{y}_\delta - \boldsymbol{K}\boldsymbol{x}\|^2)/(1 + \|\boldsymbol{x}\|^2))$ and $\boldsymbol{I}$ is the identity matrix. Using a first-order Taylor series approximation at kth iteration $f(\boldsymbol{x}_k) = f(\boldsymbol{x}_{k-1}) + \boldsymbol{K}(\boldsymbol{x}_{k-1})\boldsymbol{S}_k$ and the residual vector $\Delta\boldsymbol{y}_\delta = \boldsymbol{y}_\delta - f(\boldsymbol{x}_{k-1})$, (11) can be rewritten as

$$\boldsymbol{S}_k = \left(\boldsymbol{K}^T\boldsymbol{K} + g(\boldsymbol{x}_{k-1})\boldsymbol{I}\right)^{-1} \boldsymbol{K}^T\Delta\boldsymbol{y}_\delta \quad (12)$$

where $\boldsymbol{S}_k$ is the update of an iterative inverse method. The constant $g(\boldsymbol{x}_{k-1})$ is dependent on the state space parameters and it cannot be uniquely determined by using any analytical solution without knowing the true state vector. Thus, this problem is solved by eigenvalue analysis and considering the error in variables as described in Markovsky and Huffel [36] who demonstrated that the adoption of the lowest singular value of the matrix $[\boldsymbol{K} \ \Delta\boldsymbol{y}_\delta]$ is the appropriate value of $g(\boldsymbol{x}_{k-1})$ for both single and multiple iterative solutions. This allows factorizing the matrix $[\boldsymbol{K} \ \Delta\boldsymbol{y}_\delta]$ using singular value decomposition (SVD), as follows:

$$[\boldsymbol{u} \ \boldsymbol{\sigma} \ \boldsymbol{v}] = [\boldsymbol{K} \ \Delta\boldsymbol{y}_\delta] \quad (13)$$

where $\boldsymbol{\sigma}$ represents the singular values. The lowest value of $\boldsymbol{\sigma}$ is the appropriate regularization strength (σ_{end}) for an iterative solution [36]. Equation (12) can now be reformulated to its final form for a single iteration as

$$\boldsymbol{x}_{tls} = \boldsymbol{x}_{ig} + \left(\boldsymbol{K}^T\boldsymbol{K} + \sigma_{end}^2\boldsymbol{I}\right)^{-1} \boldsymbol{K}^T\Delta\boldsymbol{y}_\delta \quad (14)$$

where, $\boldsymbol{x}_{ig}$ is the initial guess information for the parameter space. This method yields a good retrieval when the condition number of the Jacobian may be considered low for a given problem. Since the RT problem is inherently ill-conditioned, the regularized TLS (RTLS) or truncated TLS (TTLS) are most commonly used (e.g., [23], [24], [31], [37], [38]). The number of state space variables of the present problem, which will be introduced in the next section, is only two. It is not feasible to develop a first derivative operator for such a problem, which is the minimum requirement for the RTLS method. The present problem will be constructed with three measurements and two parameters and there will be only two singular values. A study was made using truncation of second singular value from the solution space but the solution produces erroneous results due to excessive regularization. Thus, we proposed an alternate way to regularize the TLS solution.

As reported by Koner *et al.* [24] using a mathematical derivation, the error realization into the state space parameters for any ill-conditioned linear inversion is proportional to the condition number of the Jacobian (κ) and all errors associated with this inversion, which can be written as

$$\|e\| \leq \ \kappa \, \Sigma\|E_i\| \quad (15)$$

where e is the realization of the error into the retrieved parameters and E_i represents errors associated in the inversion, including errors in Jacobian, forward model, measurements, and ancillary data. Since TLS formulation does not account for the error propagation due to an ill-conditioned matrix, we have modified the TLS to minimize enhancement of errors from the measurement to the state space. We have empirically modeled this for our specific application and found that the lowest singular value multiplied by $\log(\kappa)$ and divided by L2-norm of the normalized residual vector in a step function is the optimal regularization strength, which we refer to as "modified total least squares" (MTLS). A similar type of logarithmic constraint ($\log(n)$ instead of $\log(\kappa)$, where n is the cardinality) is assumed for theoretical ill-conditioned inverse problems in [39], where they argued that a logarithmic constraint has important optimality properties.

The basic understanding of the ill-conditioned inversion is that the regularization strength would be high when the problem is highly ill-conditioned and/or high error. It is very difficult to estimate the noise embedded in a particular retrieval in an operational environment. The residual is the sum of the noise and signal. Therefore, using some knowledge of the problem, it is possible to approximate the signal-to-noise ratio (SNR) value. Most satellite inverse operational problems have a wide range of SNR, as signal varies widely depending on the variation of the target parameters that are dynamic. When the residual values are low, the information content is also low and noise is higher compared with the signal. Since the truth is close to initial guess when the signal is low, the inversion error may be comparable to or higher than the improvement of the state space knowledge by the inversion. Thus, we increase the regularization strength to avoid relatively higher inversion error compared with the information gain, in the case where the information content is very low. The L2-norm of the normalized residual $(\Delta \boldsymbol{y}_\delta^n)$ is applied to increase the regularization strength for low residual cases in a step function. This normalization is based on the amount of error in the problem, e.g., if the expected error is 10–50 (of some arbitrary unit), the residual is divided by 100 and similarly if the expected error 0.1–0.5 (of the some arbitrary unit), the residual is divided by 1. The residual constrained regularization is a step function with respect to SNR, which is given as

$$\begin{aligned} \gamma_{snr} &= \frac{\|\Delta \boldsymbol{y}_\delta^n\|}{\sqrt{m}}; \quad \text{when SNR} \leq \beta \\ \gamma_{snr} &= 1; \quad \text{when SNR} > \beta \end{aligned} \tag{16}$$

where m is the number of measurements and β is the threshold of SNR, which must be proportional to the condition number of the Jacobian. The basic concept is that high signal is required if problem is highly ill-conditioned. The implemented additional regularization is the combined concepts of the SNR and residual norm-based optimization [40], [41]. We need more research on this to identify the proper value of β for generalization. The final formulation of MTLS is

$$\boldsymbol{x}_{mtls} = \boldsymbol{x}_{ig} + \left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}^2}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T\Delta\boldsymbol{y}_\delta. \tag{17}$$

The empirical information content for such a problem can be determined [2], [10], [16] in terms of degree of freedom in retrieval (DFR) by taking the trace of the model resolution matrix ($\boldsymbol{M}_{rm}$). The expression for $\boldsymbol{M}_{rm}$ and DFR as follows:

$$\begin{aligned} \boldsymbol{M}_{rm} &= \left[\left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T\boldsymbol{K}\right] \\ \text{DFR} &= \text{trace}(\boldsymbol{M}_{rm}). \end{aligned} \tag{18}$$

Equation (18) gives an assessment of the amount of information coming from the measurements into the retrieved parameters.

A. Error Analysis

When a problem is solved using (17), there are errors from two major sources embedded into the retrieved parameter [10]. Data error, which penetrates from the measurement space to the state space is driven by the condition number of the inverted matrix after regularization. The inserted regularization for reduction of the condition number of the original Jacobian contributes so-called regularization/smoothing error. The resultant retrieved error at individual pixel level can be derived using a simple algebra for a regularized inversion as

$$e = \sqrt{e^2} = \sqrt{\langle\|\boldsymbol{x}_{mtls} - \boldsymbol{x}_{true}\|^2\rangle}. \tag{19}$$

Combining (1), (17), and (19) and omitting the square root yields

$$\begin{aligned} \|e\| = \Bigg\langle \Bigg\| &\left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T \\ &(\Delta\{\boldsymbol{K}\boldsymbol{x}_{true} - \delta\boldsymbol{K}\boldsymbol{x}_{true}\} + \delta\boldsymbol{y}) \\ &- \boldsymbol{I}(\boldsymbol{x}_{true} - \boldsymbol{x}_{ig}) \Bigg\| \Bigg\rangle. \end{aligned} \tag{20}$$

Using the assumption that the statistical errors in the components of $\delta\boldsymbol{y}$ are uncorrelated and replacing the $\boldsymbol{M}_{rm}$ from (18), the (20) can be rewritten as

$$\begin{aligned} \|e\| = &\langle\|(\boldsymbol{M}_{rm} - \boldsymbol{I})(\boldsymbol{x}_{true} - \boldsymbol{x}_{ig})\|\rangle \\ &+ \Bigg\langle \Bigg\| \left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T \\ &(\Delta\boldsymbol{y}_\delta - \boldsymbol{K}(\mathbf{x}_{true} - \boldsymbol{x}_{ig})) \Bigg\| \Bigg\rangle. \end{aligned} \tag{21}$$

The first term of (21) is independent of statistical term and referred to as the systematic error. From the stochastic point of view, it is a bias for a fixed regularization scheme in a given data set. However, we use dynamic regularization, and thus, the error from this term varies for retrievals at different pixels. The second term in (21) contains the statistical error $\delta\boldsymbol{y}$ and, under the aforementioned conditions discussed, (21) can be further simplified to

$$\begin{aligned} \|e\| = &\|(\boldsymbol{M}_{rm} - \boldsymbol{I})(\boldsymbol{x}_{true} - \boldsymbol{x}_{rig})\| \\ &+ \left\| \left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T \right\| \\ &\times \langle\|(\Delta\boldsymbol{y}_\delta - \boldsymbol{K}(\boldsymbol{x}_{true} - \boldsymbol{x}_{ig}))\|\rangle. \end{aligned} \tag{22}$$

The requirement for accurate error estimation in (22) is knowledge of the true state space parameter ($\boldsymbol{x}_{true}$) and the function is perfectly linear. Thus, (22) is a useful tool for the optimization of the different errors for establishing validity of linear regularized inverse methods during development using simulation studies. However, most real life problems are nonlinear and $\boldsymbol{x}_{true}$ is never available. Thus, we substitute here $\boldsymbol{x}_{true}$ by the value of $\boldsymbol{x}_{rtv}$ for the approximated analytical error (AE) calculation. The values of this analytical error are subsequently used to determine the quality of retrievals in our study. This also helps to assess our assumption that the retrieved value for $\boldsymbol{x}$ is a valid (useful) substitute for this problem.

III. Datasets for Forward Models

It is a common practice to use data of match-up between satellite measurements, *in situ* references and other ancillary information to validate and further improve a product in most operational environments. In the present study, we have used BTs from the GOES-13 imager and National Centers for Environmental Prediction (NCEP) surface and upper-air forecast fields. The satellite SST and buoy match-ups were operationally generated at the Office of Satellite Products and Operations (OSPO) at NOAA for both day and night scenes. The match-up window was set to ± 30 min for buoys coincident with satellite pixels, yielding about 100 000 matches (irrespective of cloud cover) for each month in this match-up database (MDB) and the data are geographically well distributed. Subsequently, the *in situ* data in our MDB are quality controlled using corresponding quality flags from NOAA *iQUAM* [42].

Upper-air forecast fields within 3 h of each match were obtained from NCEP. The NCEP data with a latitude and longitude resolution of 1 degree are also stored in our MDB during original data processing and comprise profiles of atmospheric temperature and relative humidity at specified pressure levels, and surface pressure and temperature. In this validation exercise, the nearest profile is associated with each match. Despite the fact that mature RT forward models exist in IR and near IR, a fast forward RT model, the Community Radiative Transfer Model (CRTM), is used in the operational environment to reduce computational cost. Simulated Brightness temperatures (BT) were calculated employing the CRTM v2.0 [43] with NCEP data as input for the study of MTLS and OEM. We have also used the CRTM-derived partial derivatives (Jacobians) of the channel BTs with respect to surface temperature ($\delta y_\lambda / \delta s$, where $y = \mathrm{BT}$ and $s = \mathrm{SST}$) and logarithm of total column water vapor (TCWV) ($\delta y_\lambda / \delta \log(w)$ where $w = \mathrm{TCWV}$), required to solve (17) and (25). (Note that, even for retrospective processing we have used NCEP forecast data rather than reanalyzed fields in order to assess accuracy in the operational setting.)

In addition we have also used climatological aerosol profiles as input to the CRTM. This is mainly to aid in mitigation of nighttime aerosol contaminated pixels, which is not feasible employing any spectral difference available with the limited GOES-13 channel set. Thus, we use the CMIP5 [44] climatological aerosol profiles [RCP Database (Version 2.0.5)] to reduce nighttime retrieval errors only, which will serve to partly alleviate the issue of aerosol contamination.

In this paper, using fast-forward CRTM in the interest of operational efficiency, our goal is to implement a deterministic method to improve physical SST retrievals. Speaking strictly scientifically, the study of day-night together is not a significant issue as the RT physics for both day and night is mature. However, approximations of RT physics are used in any fast-forward RTM, including the CRTM and the confidence for 3.9-μm daytime simulations is not yet well established. To separate this effect from an assessment of the true performance of MTLS and two other comparator methods (regression and OEM), we have restricted this study to night-only retrievals. However, we have also done a comparative study for daytime data (not shown in this paper) which reaches the same conclusion for relative performance of the methods.

The retrieval problem is initially assumed to be linear within the range of BTs corresponding to NCEP field errors. This means we anticipate that a reasonable solution may be obtained in one step with no iteration. It is also assumed that, in this limited BT range, only the leading two terms affecting BT, i.e., s, w, need to be considered as variables to solve the inverse problem for the limited number of radiometric channels in the GOES-13 imager.

The Bayesian cloud screening method reported in Merchant *et al.* [45] is used operationally at NOAA and was therefore considered in this study. The threshold probability of clear sky was assumed to be 0.98. Ideally, satellite SSTs are best validated against *in situ* radiometric SSTs, which are also skin measurements [46]–[49]. However, such reference data have a limited availability and routine validation in the full satellite domain remains unfeasible. Therefore, satellite SST is customarily validated against *in situ* SSTs (e.g., buoys), including in this study. A constant offset of -0.17 K to account for the skin bulk SST differences was used as a first-order approximation. This is in agreement with the typical value for the skin effect at night [46].

IV. Other Prevalent SST Retrieval Methods

The theoretical basis of a multiple-channel sea surface temperature estimation using satellite infrared data was developed in the 1970s [50]. An early history of this development is given elsewhere [27], [28], [51]. The atmospheric correction term of such methods was formulated by taking advantage of the differential absorption between 11 and 12 μm channels using approximated RT physics. However, the 12 μm channel is replaced by one at 13.4 μm for GOES-12 and beyond. Very little of the signal in the 13.4 μm channel emanates from the surface when the water vapor loading is high and it provides suboptimal corrections for 10–20% of cases in our study. Thus, we use a linear combination of regression formulation, similar to the equations in references [52]

$$s_{regb} = a_1 y_{3.9} + a_2 y_{11} + a_3 y_{13.4} + (\sec(sza) - 1)\,(a_4 y_{3.9} + a_5 y_{11} + a_6 y_{13.4}) + C. \quad (23)$$

For comparison purposes, in addition to the MTLS retrievals, we have also included regression SST based on calibration against buoys (REGB). The coefficients are calculated using the data for the month of June 2010, which are subsequently used to derive REGB for all months.

The spatial and temporal distributions of buoys were very low compared with global satellite measurements and may not have been statistically significant [53]. The results may be ambiguous due to the fact that these retrievals are usually validated against the same set of buoys used to calculate the regression coefficients. To overcome these problems, RT coefficient-based regression SST retrieval method has been developed. The coefficients are calculated using simulated BTs by considering a large number of atmospheric profile data selected to spatially and temporally representative of conditions within the GOES-13 field of view. For comparison purposes, we have also included the GOES-13 SST product generated (OSPO), which is derived using the following regression-based equation [52]:

$$s_{ospo} = a_1 + a_2 y_{3.9} + a_3 y_{11} + (\sec(sza) - 1)(a_4 + a_5 y_{3.9} + a_6 y_{11}). \quad (24)$$

We have also compared the results of MTLS with those from the well-known optimal estimation method (OEM), which is widely used in different satellite retrieval problems [2], including for SST retrievals [54], [55]. In the implementation of OEM for SST retrievals the same reduced state vector

$$\boldsymbol{x} = \begin{bmatrix} s \\ \log(w) \end{bmatrix}$$

is used, where, as before, s is SST and w is total column water vapor. The OEM equation may be stated as

$$\boldsymbol{x}_{oe} = \boldsymbol{x}_a + \left(\boldsymbol{K}^T \boldsymbol{S}_e^{-1} \boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1} \boldsymbol{K}^T \boldsymbol{S}_e^{-1} (\boldsymbol{y} - f(\boldsymbol{x}_a)) \quad (25)$$

where $\boldsymbol{x}_a$ is the *a priori* reduced state vector, $\boldsymbol{S}_a$ is the *a priori* error covariance matrix and $\boldsymbol{S}_e$ is the combined measurement and RT error covariance matrix. For characterizing the measurement error in this case, the instrumental noise equivalent differential temperature (NEDT) values are assumed to be: $\varepsilon_{3.9} = 0.05$, $\varepsilon_{11} = 0.053$, $\varepsilon_{13.4} = 0.06$ K at 300 K as stated by the satellite operators, where ε_λ denotes NEdT for a channel at λ μm. In this paper, the NEDTs are considered constant in radiance space (the underlying assumption is that the NEDT is governed by photon noise only) and are calculated according to the measured brightness temperatures (BT) employing Planck's law. To characterize the RT error for a particular sensor, e.g., GOES-13, the exact channel-specific errors of the fast-forward CRTM are not available. Thus, based on the literature about CRTM error for the AIRS instrument [56], we assumed CRTM errors to be 0.25, 0.15, and 0.15 K for the aforementioned GOES-13 channels, respectively. It is very difficult to do a perfect forward model for any science problem, and more so in any operational environment when fast models require approximations to the full physical processes, thus the estimation of error in the forward model is potentially ambiguous, however, it is a required input for OEM. These uncertainties are added to the instrumental noises to obtain values equal to the observational errors $(\boldsymbol{S}_e)$ in OE. The radiometric measurements of 3.9, 11, and 13.4 μm are assumed to be mutually independent, and thus, a diagonal covariance matrix is used for observation error. A diagonal covariance matrix is also used for S_a, with values for *a priori* errors in SST ($\sim$1 K) and 15% variance in TCWV.

The information content in terms of the degree of freedom of signal (DFS) [2] is

$$\text{DFS} = \text{trace}\left[\left(\boldsymbol{K}^T \boldsymbol{S}_e^{-1} \boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1} \boldsymbol{K}^T \boldsymbol{S}_e^{-1} \boldsymbol{K}\right]. \quad (26)$$

Retrieval results of these methods are discussed in Section V and beyond.

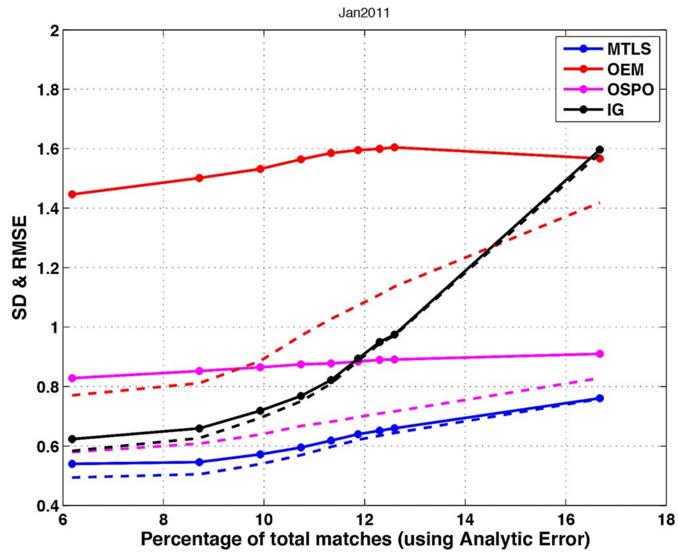

Fig. 1. Total errors in SST retrievals from three different methods and in first guess SST for January 2011. RMSE is shown as solid lines and SD is shown as dashed lines. MTLS: modified total least square; REGR: regression; OEM: optimal estimation method; IG: NCEP initial guess surface temperature.

V. Results and Discussions Using CRTM 2.0

Here, we present comparative retrieval results between three methods, namely, MTLS, OEM and OSPO regression, which are somewhat related via RT modeling. The comparative performances of all methods with a forward model calculation employing CRTM 2.1 for MTLS and OEM will be discussed in Section VI and beyond. The reason why we are showing both sets of results is because it demonstrates the relative sensitivity of the two physical retrieval methodologies (MTLS & OEM) to RTM accuracy. Another objective is to capture different operational retrieval problems at different times, which may not be available in a particular month of match-ups. Thus, different months will be chosen for the discussions of different operational problems. Finally, in Sections VII and VIII, information content analysis and the time-series analysis of 42 months will be presented.

Fig. 1 shows the retrieval errors with reference to buoy temperatures (SST_b) for three different retrieval methods for the month of January 2011. The percentage of the total matches shown in abscissa of Fig. 1 is based on the value of total analytical error in MTLS. It is not possible to do the binning based on the fixed value of AE due to the practical constraint of the problem, because the ranges of analytical errors are different for different months. Thus, we first sort the analytical errors in increasing order and select the value of analytical error at 20% of total cloud-free pixels as the starting point. Then we divided the range of AE values (from first point to the maximum of AE) into eight bins. If a bin does not get at the least 5% of cloud-free data, then it is combined with the subsequent bin. For each bin, the percentage of total matches is based on the cumulative analytical errors.

An initial guess (IG) of the parameter to be retrieved is required for any RT inversion problem. Note that, *a priori* and IG are not necessarily the same, i.e., any *a prior* may serve as

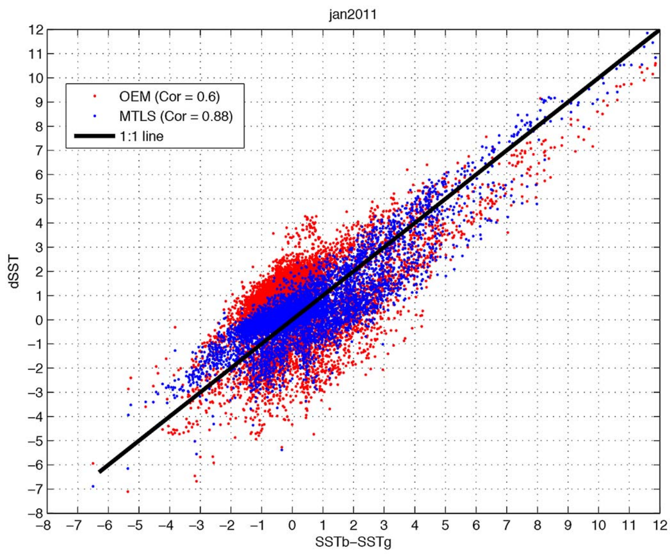

Fig. 2. Scatter plot of the true versus calculated SST innovation for a single iteration retrieval for both OEM and MTLS. "corr" stands for correlation factor.

the IG, but the IG does not require *a priori* knowledge, i.e., it could be far from truth and could be of a different shape for profile retrieval. We have assumed that the IG and *a priori* are identical for this study. The OEM requires *a priori* by the nature of its own derivation and the retrievals must be influenced by *a priori*, but the MTLS does not require *a priori* by its own definition and the retrievals are expected to be independent of *a priori*. If the retrievals are independent of *a priori* using OEM, then it is indicative that these retrievals are following a different inverse theory because the mathematical form of most of the inverse methods are identical. We have used SST and TCWV in the $1^\circ \times 1^\circ$ global forecast system (GFS) data as the IG for our SST retrievals. Since our main focus on SST retrievals, the statistical difference between the truth (buoy temperature) and the IG/*a priori* of SST (SST_g) is also shown in Fig. 1, primarily to quantify the improvement of *a priori* knowledge after inversion of the satellite BTs measurements using the different retrieval methods. It is a desirable property of any retrieval that the addition of the satellite information increases accuracy over the IG/SST_g.

Fig. 2 shows the adjustment to the initial guess SST calculated by the two physical retrieval approaches (MTLS and OEM) compared with the "true" innovation (i.e., SST_b-SST_g). The most desirable result is for points to lie on the 1 : 1 line. The first thing to note is that the range of SST innovations is rather large (up to $\sim$10 K) although the vast majority lies within a couple of degrees of zero. The significant vertical spread of points for the OEM, particularly where the required SST innovation is small, indicates that there is some cloud leakage in the Bayesian cloud detection. Note that the underlying assumption is that the shape of the WV profile (and corresponding atmospheric temperature) is correct when this RT-based inversion is considered. However, there is ambiguity in the NCEP generated shape of the WV profile. Furthermore, if the total column water vapor is drastically different from IG, it is impossible to find a solution using only a single iteration due to nonlinearity. The last bin is likely to contain a substantial population of "bad" retrievals, i.e., caused by cloud-leakage or other errors in ancillary information including WV profile shape error. This is implemented using appropriate threshold conditions based on the physical understanding of the problem on the output of SST and TCWV from MTLS retrieval. The detection of night-time fractional cloud or low level cloud is very difficult when using only a limited number of imager measurements and RT output. This will produce low BTs for all measurements, as a result, MTLS solution will produce a relatively large negative value of SST increment as well as high TCWV to compensate for the negative value of the residual. This offers the prospect that such conditions can be detected at solution time, by determining a suitable threshold. Thus, another advantage of MTLS solution is to facilitate improved quality indexing (QI). It is notable that all retrieval methods closely follow the trend determined by the QI that is based only on the computed analytical total error using MTLS method. This implies that the employed QI is capable of providing an independent assessment of the inherent errors in this ill-posed inversion problem, and that the primary sources of error (channel noise, residual cloud, water vapor, etc.) have broadly similar effects on all retrievals. It should be noted that AE is unable to account for certain aspects of the statistics obtained with respect to *in situ* matches, such as surface effects, buoy error and representativity (point-to-area) error.

The errors in SST retrievals are shown in Fig. 1 using two different metrics: a) RMSE, and b) SD as following the two schools of thoughts. (Here, RMSE is defined as the square root of sum of squares of SD and bias.) One may argue that the bias is also an important term to report, however, it is irrelevant in the scope of this paper. Bias is produced in a solution due to model approximation and it is important for controlled experiments to improve the model. In this present study, bias may result from the approximation errors of fast forward model, inverse model and instrument model as well as cloud leakage and errors in the reference and ancillary data, and these are difficult to be separated out in an operational environment. The composite bias has no additional scientific significance unless the source is identified and we stay away from this in this paper. Additionally, it is often observed that low bias (difference between RMSE and SD) and relatively higher SD are typical for any global data set (e.g., satellite SST). However, the bias term may increase when the same data set is analyzed on the regional scale due to local retrieval conditions being different from those used to derive the global average. This is generally the case for regression-based methods, which are unable to account for the full range of retrieval conditions in only a few coefficients. It can be argued that the most representative single statistic to indicate the quality of any retrieval is the total error (RMSE), as well as it is suggested by Rodgers [2], particularly for easy intercomparison purposes. The errors in the different retrieval methods, as shown in Fig. 1 are discussed next.

A. *Analysis of MTLS Results*

As shown in Fig. 1, the RMSE of MTLS retrievals (MTLS) is always lower than that of initial guess. This confirms that the MTLS retrieval scheme is using the satellite measurements to improve knowledge in the parameter space. It is also able to analytically calculate the total error at individual pixel level. The value of RMSE of MTLS is less than 0.65 for up to 13% of matches and less than 0.55 for up to 10% of matches as well as the bias is small (i.e., the RMSE and SD values are close to each other). This improvement is possible due to the fact that MTLS method inherently calculates the optimum regularization without requiring extra information, e.g., covariance matrices of *a priori* and observational errors. The remaining error could be due to the fact that presently implemented MTLS is constrained by CRTM error, which will be discussed later as well as the ancillary atmospheric data, which has inherent errors. (These data are generally produced by the solution of some optimization problems, where conventionally, a Bayesian inversion technique is used.). Also, as already discussed, a single iteration in MTLS may not be fully adequate in certain circumstances, and a second iteration may improve the result by reducing the functional residual error.

The MTLS technique is a regularization method. It is often argued that regularization methods reduce the degrees of freedom of the solution), as a result, the information content of the solution is low [57]. However, Fig. 2, which shows the update after a single iteration, confirms that MTLS can approach the truth even when the magnitude of the required innovation is as much as 10 K. This implies that the issue of "loss of information" for conventional fixed regularization inversion can be overcome using a TLS-based method, where regularization is determined by residual where various errors are embedded. Fig. 2 shows that the correlation coefficient of MTLS retrieval is 0.88, which is much higher than for OEM (0.6), at least for CRTMv2.0. The accuracy of CRTMv2.0 will be discussed later.

This analysis (Fig. 2) shows that SST retrievals can depart significantly from *a priori* (e.g., SST error can be beyond $\sim$2 K of the numerical weather prediction (NWP) analysis/forecast model). Match-up data sets may exist where "$\mathrm{SST_b}$-$\mathrm{SST_g}$" is within 2 K due to the fact that only quality controlled drifter measurements are included and coastal moorings may be excluded due to high deviation from the background SST. However, our MDB shows that the difference "$\mathrm{SST_b}$-$\mathrm{SST_g}$" may be up to 10 K of NWP models because coastal mooring are included in our database and not excluded purely on the basis of deviation from analyzed SST. Two key inferences can be drawn from this: a) the ocean SST may be as high as 10 K from NWP analysis/forecast model SST; and b) these are still solvable using MTLS.

B. Analysis of Operational Regression Results

The previous operational GOES SST retrieval scheme at OSPO used only the 3.9 and 11 μm channels. Fig. 1 shows that the SD of OSPO retrievals (dashed pink line) decreases with decreasing number of matches. What is important here is the increased separation between the values of RMSE (solid magenta line) and SD, which implies increased bias for data expected to be of higher quality. This does not come as a surprise since regression SST retrieval is based on a set of statistical coefficients derived from RT and may be biased with respect to truth unless the agreement between modeled and observed clear-sky BT is "perfect." However, the RMSE of regressed retrievals is better than that of IG SSTs considering the whole set of data (up to 12% of matches), which means that regression retrieval is generally improving the accuracy of the parameter relative to *a priori* knowledge (i.e., NWP analyzed SST). The performance of the two-channel SST retrieval using RT coefficient-based regression retrieval, which was implemented in OSPO, is inherently limited by the number measurements and shows poorer performance than MTLS. In such retrievals there are two parameters: SST and the atmospheric correction term. Since there is noise in any measurement, it is necessary to use at least three channels because a minimum of three points is required to draw a line that accounts for the noise and allow a valid inference for a two-parameter retrieval. The present physical retrieval model assumes that TCWV is single parameter, which is a heavy approximation in RT physics in the IR region, and SST retrieval will still be somewhat erroneous due to the high approximation of the WV profile in RT physics and nonlinearity.

C. Analysis of OEM Results

Compared with $\mathrm{MTLS_{rtv}}$, the OEM retrievals ($\mathrm{OEM_{rtv}}$) are less than satisfactory (see Fig. 1) in this particular example using CRTMv2.0. RMSE and SD values $\mathrm{OEM_{rtv}}$ are always much higher compared with any other method and also with respect to the IG. $\mathrm{OEM_{rtv}}$ is the only method in this comparison study that is based on Bayes' theory and requires an *a priori* state and its associated error covariance. Fig. 1 shows that the *a posteriori* contains more error than the *a priori*, as implied by the higher RMSE and SD values in $\mathrm{OEM_{rtv}}$ compared with the IG. This suggests that by using $\mathrm{OEM_{rtv}}$ we lose confidence in improving knowledge of a system after measurement.

Since this result is unexpected using OEM, despite a low condition number of Jacobian ($\sim$5) and the problem being fairly linear, it requires further investigation. The retrieval problem has been analyzed from a different angle considering the deterministic paradigm for pixel level. Recalling (15), the retrieval error is proportional to the condition number of the matrix and total error associated with inversion. Thus, the retrieval error will be zero if the error in measurement is zero until the Jacobian is rank deficient (i.e., condition number exceeds 10^{16} for double precision calculations). The inversion of such a problem, where observational is accurately known, can be easily formulated as

$$\Delta \boldsymbol{y}_\delta - \delta \boldsymbol{y} = \boldsymbol{K} \Delta \boldsymbol{x}. \tag{27}$$

The left side of (27) has been formulated in a way that if the observational error is known accurately, subtracting it from the real-life measurement can generate an effectively "perfect" measurement. After rearranging we have

$$\frac{\Delta \boldsymbol{y}_\delta}{\Delta \boldsymbol{x}} = \boldsymbol{K} + \frac{\delta \boldsymbol{y}}{\Delta \boldsymbol{x}} \; \boldsymbol{or} \; \Delta \boldsymbol{x} = \left(\boldsymbol{K} + \frac{\delta \boldsymbol{y}}{\Delta \boldsymbol{x}} \right)^{-1} \Delta \boldsymbol{y}_\delta. \tag{28}$$

Equation (28) is valid for a square matrix and a direct inversion, where the numbers of measurements and state space parameters are equal.

In this inversion, ill-conditioned error propagation does not feature, but, as before, it is not feasible to determine the exact errors with sign associated with an observation. Even if the magnitude of error could be assumed accurately known for simulation purposes, the sign of the random errors should not be confirmed in any scientific experiment. To minimize these errors in an inversion, least squares (LS) optimization is necessary. Using matrix algebra and employing LS minimization, in a few steps (not shown for brevity), (28) can be rewritten for a rectangular Jacobian matrix as

$$\Delta \boldsymbol{x} = \left(\boldsymbol{K}^T \delta \boldsymbol{y}^{-2} \boldsymbol{K} + \Delta \boldsymbol{x}^{-2}\right)^{-1} \boldsymbol{K}^T \delta \boldsymbol{y}^{-2} \Delta \boldsymbol{y}_{\delta}. \quad (29)$$

Equation (29) is deterministically derived using simple linear algebra for the single pixel. Equations (29) and (25) for the OEM formulation are identical, and both can produce reasonable retrievals if the expected values of $\delta \boldsymbol{y}$ and $\Delta \boldsymbol{x}$ can be accurately specified in the inversion. A simple numerical experiment using accurately specified $\delta \boldsymbol{y}$ and $\Delta \boldsymbol{x}$ on an individual retrieval for a random linear function shows that the retrieval contains error of the order of average SD of $\delta \boldsymbol{y}$. This confirms that (29) can retrieve the state space parameter with the order of measurement error if measurement error and truth are perfectly known. However, we find that the retrieval error is much higher than measurement error if one of the errors or both are not accurately known and the error enhancement is proportional to the condition number of inverted matrix. In practice, stochastically derived estimates of these errors are used in OEM for a set of measurement because obtaining per-pixel exact value is unfeasible, and will therefore generally be either an overestimate or underestimate of the true error in any given retrieval situation.

VI. RT Sensitivity for MTLS and OEM

The error $\delta \boldsymbol{y}$ in (29) is the collective error due to instrument noise, errors in ancillary atmospheric data, unaccounted model parameters and approximation error in the forward model. In order to analyze and effectively understand the forward model error, we have performed an offline study. We have observed a mean difference of 2.5 K between CRTM 2.0 fast forward model and MODTRAN 4.2 [58] band model simulations for 13.4 μm channel. According to Rodgers [2], all errors can be included in the error covariance to implement OEM. Thus, we made an investigation by replacing the specified CRTM error variance value in 13.4 μm channel of ~0.15 with 2.5 K. Using a 2.5 K value overall improves the results (i.e., SD and RMSE are reduced) compared with using 0.15 K but the difference between these two parameters is still high (figure is not shown). This is expected because high error covariance produces high regularization, which increases bias toward the IG in the solution in any regularization scheme. What is somewhat surprising is that this observation holds, although 13.4 μm channel is in the nonwindow region, i.e., does not contribute directly to the SST innovation. Additionally, this also shows that the aforementioned experiment does not follow the notion of OEM theory, rather it illustrates the underlying regularization principle.

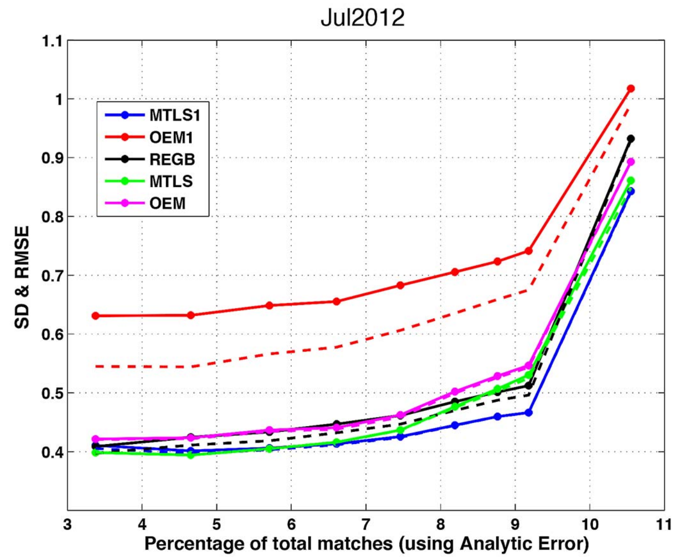

Fig. 3. Plots of retrieval errors for MTLS and OEM with error variance in channel 13.4 of 0.5K using CRTM2.0 (SD is "dashed" and RMSE ='*solid line') and MTLS1 and OEM1 using CRTM2.1 as well as REGB and OSPO (SD is "dot solid" and RMSE is "circle solid") for July 2012.

The SD of "simulation *minus* observation" for the 13.4-μm channel is less than 0.8 in our monthly MDB. It is generally accepted that bias should not be more than SD for any meaningful analysis, so the model needs to be improved in such a situation. Results improve if we assume a more reasonable error covariance of 0.5 K for the 13.4 μm channel instead of 2.5 K (cf. Fig. 3). As already noted, perfect forward modeling is close to impossible in any branch of science, particularly in an operational environment. Therefore, it is beneficial for a retrieval scheme to be as robust as possible to RT errors. To reduce the effect of a 2.5 K error in the 13.4 μm channel in CRTM2.0 for the MTLS solution, two pseudomeasurements, which are the difference ($BT_{3.9}$-BT_{11}) and average (($BT_{3.9} + BT_{11}$)/2) of the 3.9 and 11 μm as per a linear approximation (including Jacobians with difference and average) are included for comparison study in the following section.

A. *Effect of RTM Versions: CRTM 2.0 Versus CRTM2.1*

We have tested our results using the older and newer versions of CRTM (v2.0 and v2.1). As previously mentioned, analysis thus far was carried out using forward model BTs calculated with CRTM 2.0. The previously observed offset of 2.5 K in CRTM 2.0 was addressed in CRTM 2.1 [59] by the developers of this model. The difference was primarily due to a post-launch revision of the spectral response function for the 13.4-μm channel [60]. The comparative retrieval results for different methods and the two versions of CRTM are shown in Figs. 3 and 4 for two different months (July 2012 and February 2013). These illustrate some interesting observations, as described in the following. First of all, the small difference between RMSE and SD of CRTM v2.0-based MTLS shown in Fig. 1 is essentially nonexistent in Figs. 3 and 4, which utilize two additional pseudochannel measurements, aforementioned. In addition, both the RMSE and SD values in Figs. 3 and 4 are lower than that without using pseudomeasurements. It should

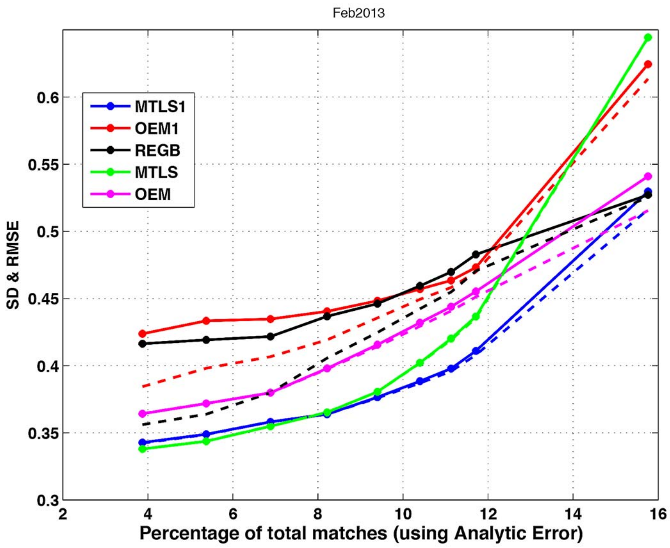

Fig. 4. Plots of retrieval errors for MTLS and OEM with error variance in channel 13.4 of 0.5K using CRTM2.0 (SD is "dashed" and RMSE = '*solid line') and MTLS1 and OEM1 using CRTM2.1 as well as REGB and OSPO (SD is "dot solid" and RMSE is "circle solid") for February 2013.

be also mentioned that a dramatic reduction in SD and RMSE in OEM is observed, just by using tuned error variance value (0.15 to 0.5 K).

The observed sensitivity of OEM to RTM error in the nonwindow 13.4 μm channel yields a key insight into a weakness of the OEM. One may argue that the RTM (CRTM 2.0) is responsible for introducing the high error in OEM (see Fig. 1). However, can we develop a perfect forward model for any science problem? If it is true, why we are not hunting for alternate method, which can be robust for the error in nonwindow channel. Moreover, in operational systems, although infrequently, an error of order 2.5 K for the 13.4 μm channel may also originate from other non-RTM sources, such as incorrect total column water, water vapor profile or ice-contamination in the sensor [61]. Additionally, as in this case, an offset in modeled BT may also originate due to spectral response function shifts after launch and sometimes identification of such issues takes a long time. Therefore, the error specification required for OEM may not be done in a timely manner, e.g., such an issue with GOES-12 was detected and reported after completion of the mission life [62]. Therefore, specifying $\boldsymbol{S}_e$ value for OEM based on prelaunch information will always be speculative and unclear whether the specified values are representative or not. The overall statistics will be significantly degraded even if a few points out of millions of retrievals contain such a high error. In addition, the detections of such retrievals are difficult. Validation statistics can be improved by excluding outliers for those points where references are available, but it is difficult to exclude outliers in an operational product and for all retrieval points.

The aforementioned observation also emphasizes the importance of obtaining an optimal regularization in a weighted regularization scheme. This also implies that a large error can be propagated into the solution in a weighted regularization scheme due to small errors in the essential parameters of $\delta\mathbf{y}$ and $\Delta\mathbf{x}$ in OEM as well as in (29). Since the error assigned to the 3.9-μm channel is greater than that for the 13.4 μm channel, the information (SST and WV) coming from the 13.4 μm channel will actually be greater in the OEM solution, considering only the channel weights in the solution. However, this issue is complex since it depends on the channel weighting function and the amount of water vapor. Recall that SST information content in 13.4 μm channel is very low and will not be influenced as much as the TCWV information due to the variable weights of $\boldsymbol{S}_e$. Thus, the error in TCWV is expected to be high due to RT errors (CRTM2.0) which manifest as large increments in TCWV ($> 100\%$) in the MTLS solution. Whether or not it leads directly to a higher error in the SST requires further investigation.

Another interesting observation from Figs. 3 and 4, supporting our aforementioned arguments, is the sensitivity of OEM to a change in the version of CRTM. The RMSE values for OEM retrievals using CRTM 2.0 (with 0.5 K error covariance) are significantly lower than those using CRTM 2.1 (with 0.15 K error covariance) except last bin. This implies that, for OEM, a more accurate forward model with reasonable error specification does not necessarily outperform retrievals with a less accurate RT model but with a more tuned error specification. We observed this behavior for many months, although some months show that the RMSE of OEM using CRTM 2.1 is better than CRTM2.0 with 0.5 K error covariance. This demonstrates that an OEM solution may display inconsistent results with respect to CRTM versions, because it is difficult to exactly specify the errors in an operational environment, i.e., a more accurate model and supposedly correct noise values do not necessarily produce the most accurate result.

As we mentioned in Section V, bad retrievals identified on the basis of retrieved WV increment, shown in the last bin, are mainly due to cloud leakage and/or high error in WV profile and total column. Large errors are observed in the WV retrievals using MTLS and CRTMv2.0 due to high simulation error in the 13.4 μm channel. As result, a large number of pixels ($\sim$4%) were consigned to the last bin (see Fig. 1) without significantly improving the results. However, a drastic error reduction for retrieving WV using CRTMv2.1 is observed, which is manifested by significant reduction of error at the cost of fewer discarded observations (see Fig. 3). Errors are drastically reduced with same trend for all retrieval methods (REGB, MTLS, OEM) for the last-but-one bin in Fig. 3. This implies that high cloud leakage is the primary cause for these pixels that are consigned to the last bin. It can be confirmed by the argument that REGB does not depend on the forward model (CRTM+NCEP) and it will be only affected by cloud in a similar way. We have separately studied this issue and have found that there was high cloud leakage of Bayesian cloud detection method for this month. (We have completed a study on the issue of Bayesian cloud detection problem and found that there is significant cloud leakage and $\sim$50% pixels are discarded for which good SST retrievals may be obtained, which will be a separate future publication.)

Fig. 3 shows that the reduction of RMSE of 0.85 to 0.47 for the MTLS solution is possible by discarding only 1% of the total pixels that are present in the last bin. This is based on the interpretation of simultaneous MTLS solutions for TCWV and SST. On the other hand, the gradients of the lines for the last bin of Fig. 4 are different for the various retrieval methods. The gradient for REGB is lower but that retrieval only depends on measured BTs. The gradients for RT-based OEM and MTLS are higher but have different magnitudes for the two versions of the CRTM. The lower gradient of REGB solution in the last bin indicates that cloud detection is not the only cause for this month, but rather high RT errors due to imperfect WV profiles. These profiles also contribute to the population of the last bin, at least for the physical retrieval methods (MTLS and OEM).

It should be also noted that the monthly average RMSE of REGB (regression against buoy) excluding the last bin is slightly lower than the same for OEM irrespective of the CRTM version for July 2012 (see Fig. 3). The RMSE of REGB is very slightly lower than that of MTLS using CRTMv2.0 (see Fig. 3) for the set of all data excluding last bin but more up to the good match of ∼8.5%. However, the RMSE of REGB is always higher than MTLS RMSE when CRTMv2.1 was employed for MTLS, for the all data including the last bin. Reassuringly, MTLS produces the best results using CRTMv2.1, which demonstrates that, with improvements to our RTM and ancillary data, the performance of MTLS will further improve. As already mentioned, the accuracy of a forward model in an operational environment is a significant issue.

The most interesting result in Fig. 4 is that the RMSE of MTLS solutions are close to 0.35 K for almost 6% of match-ups. The RMSE of REGB for the monthly average is higher than for the other methods, but REGB does perform better than OEM (CRTMv2.1) for the best ∼8% match-ups. The RMSE of REGB does not increase much in the last bin, which implies that the cloud leakage of the Bayesian cloud detection software is relatively low for this month. The significant increase for the last bin of all the forward-model-based retrievals illustrate that high forward model error is the likely culprit for this month, presumably due to somewhat degraded quality of ancillary data, since the CRTM is invariant.

B. Analysis of Simulation Minus Observation Bias Removal

As aforementioned, realistic estimation of the various errors in an operational environment is a challenging task, however, OEM requires these errors as input parameters. Thus, some recent studies introduced measures to estimate these errors from the same measurements [54], [63], [64].

Some studies have used these estimated errors in error covariance as suggested by Rogers [2], whereas others use them in *simulation minus observation* (S-O) bias removal. The use of estimated errors in the covariance matrix is safer as it regularizes the solution or weighted the measurements without altering the physics of the problem. The main problem with an overall S-O bias removal is that it is not fully objective, as the sources of bias (differences) are not completely identifiable. To estimate this, some assumptions are required, which raise certain ambiguities. Put another way, the bias that is observed may be due to an error in the initial guess, which the retrieval is designed to correct for. We see no strong scientific argument supporting the use of the same measurements for both bias

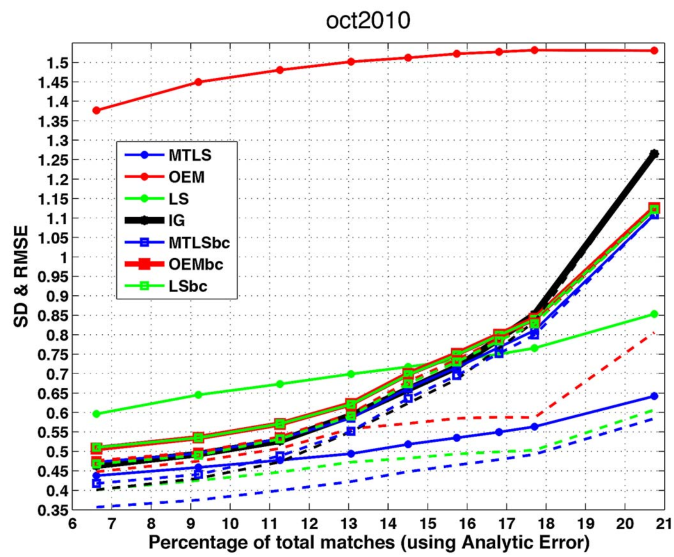

Fig. 5. Plots of retrieval errors for MTLS, LS and OEM without RBC using CRTM2.0 (SD is "dashed" and RMSE ="*solid line") and including RBC (SD is "square solid" and RMSE is "square dashed") for October 2010 including IG error.

correction and inversion to access more information than using these measurements in an appropriate retrieval method. Using spatially and temporally averaged S-O to enforce a potentially ambiguous zero-mean captures some information from adjacent pixels, rendering the retrievals not to be fully independent. From the deterministic point of view, radiance bias correction (RBC) may result in alteration of functional physics relationships, Jacobian mismatch error and information/residual loss that will degrade the retrieval. To investigate this, we conducted an experiment using RBC retrieval, which approximates the method of Merchant *et al.* [54]. Their assumption of RBC is that the mean value of the "forward model BT *minus* observed BT" for 2.5° spatial resolution of image space and for a period of 21 days is zero. We cannot consider 2.5° spatial resolution in our study as we are using only *in situ* match-ups. Thus, we considered the monthly mean of each buoy point forward model simulation to be zero for the cloud-free observations. We consider this to be a reasonable approximation of the method of Merchant *et al.* [54] because moored buoys are in a fixed location and the assumption of drifter movement within ∼2.5° spatial grid is not that unreasonable. We have used the (S-O) bias correction for MTLS without pseudo channel and OEM without tuned error covariance where RTM error is high using CRTMv2.0, as shown in Fig. 5. In addition, since the current problem consists of low ill-conditioned (condition number ∼5) Jacobian and it is expected from the theory of ill-conditioned inversion that LS can produce reasonable retrieval if the total errors are within ∼0.5 K. Therefore, we have also additionally included LS solution to gain additional insight into fundamental problem in different inverse methods. The LS formulation is simply

$$\boldsymbol{x}_{ls} = \boldsymbol{x}_{ig} + \left(\boldsymbol{K}^T - \boldsymbol{K}\right)^{-1} \boldsymbol{K}^T \left(\boldsymbol{y} - f(x_{ig})\right). \tag{30}$$

We included both LS and IG errors in this study as baseline indicators to understand the comparative performances of OEM and MTLS when an RBC is implemented. Fig. 5 shows that the LS solution without RBC produces a RMSE (solid green line with circles) of 0.76 and SD (dashed green line) of 0.5, for all matches excluding those in the last bin. The RMSE value is much higher than the SD value, which confirms that CRTMv2.0 introduces bias in simulated BT for this case. An interesting observation here is that, as compared with LS, the RMSE values of OEM is much higher than the same of SD values. Since OEM is regularized cf. LS in deterministic point of view, this result may seem surprising, as one might expect the regularization to mitigate the effect of the biased channel. However, since the error assigned to the 3.9 μm channel is greater than that for the (biased) 13.4 μm channel, the weight assigned to the latter will actually be greater in the OEM case than for LS, thus the effect of the channel bias is increased. This implies that externally supplied error assignment for weighting the measurements poses a potential risk in an operational environment. As we have shown (Figs. 1 and 5) the OEM solution appears to be degraded with respect to its *a priori* knowledge, whereas the LS solution is an improvement with respect to *a priori* knowledge for the whole data set (see Fig. 5), even in the presence of a high error in simulated BT in a nonwindow channel.

Perhaps more importantly, Fig. 5 illustrates an intrinsic strength of MTLS. It is understandable that the SD of MTLS (dashed blue line) improves cf. LS due to optimal regularization of the ill-conditioned problem, but it also noticeably reduces the bias compared with LS using CRTMv2.0 without RBC as indicated by the RMSE curve (solid blue line with circles). One might anticipate inherently high regularization due to high error in CRTMv2.0 in MTLS solution, leading to bias toward the IG, but there is little evidence of this in Fig. 5, i.e., the bias remains constant across the range of MTLS solutions, whereas it changes for the IG. The constant bias of the MTLS solution (in Fig. 5) may have originated from other sources (reference, or the ancillary NCEP data) because it is not distinctly observed in Fig. 1. Moreover, the MTLS error does not track the steep upward trend of IG error, which confirms that it is largely independent of the IG. This confirms that MTLS can function adequately in the presence of significant forward-model bias, at least if it is in nonwindow channel. Furthermore, MTLS produces best results without RBC, at least of the form developed and applied in this study.

In contrast, we see that the RMSE of OEM is drastically reduced from 1.53 to 0.85 K for the best ~18% of matches using RBC (see Fig. 5). However, it is still much higher than RMSE of MTLS (0.56 K) and LS (0.78 K) without RBC. According to Fig. 5, the large reduction of OEM error is highlighted for the success of RBC due to gross error RTM (CRTMv2.0), but the error of 0.56 K can be achieved using deterministic method MTLS without RBC, at least when RT error is not in a window channel. Moreover, the MTLS error increased to 0.8 from 0.56 K by applying RBC, and OEM error of 0.85 K with RBC is close to IG error. This implies that the solutions are no longer dependent on the characteristics of the inverse method when radiance level bias correction is implemented and the error of all methods is close to IG error.

The SD values for all three forward-model-based retrievals deteriorate using RBC, including OEM. The OEM results

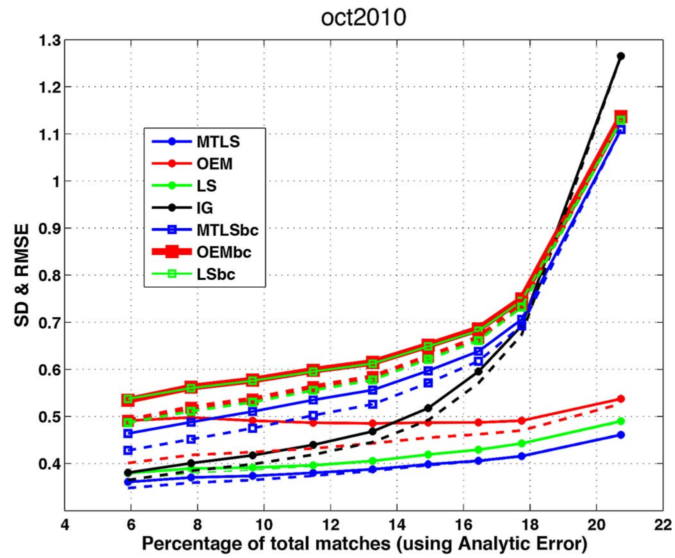

Fig. 6. Plots of retrieval errors for MTLS, LS and OEM without RBC using CRTM2.1 (SD is “dashed” and RMSE =“*solid line”) and including RBC (SD is “square solid” and RMSE is “square dashed”) for October 2010, including IG error.

in particular display unfortunate characteristics, i.e., whereas RMSE improves with RBC, SD increases. This implies that the functional physics relationships have been degraded somewhat by the RBC. This is borne out by the results for MTLS, where both SD and RMSE are increased when using RBC. Despite the fact that the LS solution is biased without RBC, the RMSE of LS is higher for data set as a whole after implementing RBC, whereas a subgroup of 16% of matches are comparable both with and without RBC. Another interesting result here is that all three retrievals produce almost identical results and close to *a priori* error under RBC, all of whose errors are significantly higher than MTLS without RBC. This study shows that RBC may compensate for some of the drawbacks of OEM in an operational environment, but is not a correct scientific approach in deterministic framework.

Fig. 6 shows results similar to that shown in Fig. 5, but using the newer version of CRTM, where the primary source of bias (13.4-μm channel) has been eliminated. All SD and RMSE for the three forward-model-based retrievals are degraded when using the previously described RBC methodology. The most interesting result found from this study is that the retrieval results of all three methods using CRTM2.1 under RBC are noticeably higher than the IG error (Fig. 6), excluding last bin. RBC may reduce OEM error in the presence of RT gross error, as is the case with CRTM2.0 and the 13.4 μm channel, but there may be price to be paid in terms of integrity of functional physics. The solutions of all forward-model-based retrieval methods without RBC get better due to the improved version of CRTM (see Fig. 6). Thus, correction in top level (radiance) is not a good choice as opposed to fixing the problem at the root level (forward model), at least in this study. Of course the latter is more difficult, but, with a sound inversion methodology, valid retrievals may still be possible in the presence of radiance bias, and accuracy will improve automatically as forward modeling improvements are made without need to adjust the inputs. Therefore, it can be concluded from this study (using Figs. 3–6) that it may be better if a forward model bias term is incorporated in error covariance matrix (e.g., [2]) for OEM to obtain reasonable results as opposed to using RBC.

It is also observed from this study that even using the improved version of RTM (CRTMv2.1), RMSE of OEM is more than IG error for more than 70% of retrievals under clear sky, excluding the last bin. Moreover, the OEM retrieval (see Fig. 6) also shows a noticeable bias under CRTMv2.1, but MTLS and LS retrievals are almost bias-free, although some bias is present in the IG. This implies that OEM has a tendency to produce biased retrieval along with higher SD compared with MTLS and LS, because it is impossible to correctly specify all the required error covariances in an operational environment. Despite the fact that RT physics is well understood, RT error is inevitably present in any operational RTM due to approximations of RT physics for reducing computational costs. Although we now have reasonable RTM (CRTM 2.1) at this time for this application, there are many satellite inversion problems where RT error is still a major issue. Thus, the lesson learned from this study may be useful for other satellite inversion problem and the study based on comparison of CRTM 2.0 with CRTM 2.1 is valuable to understand some of these issues. From this point on, only results from CRTM 2.1 will be discussed.

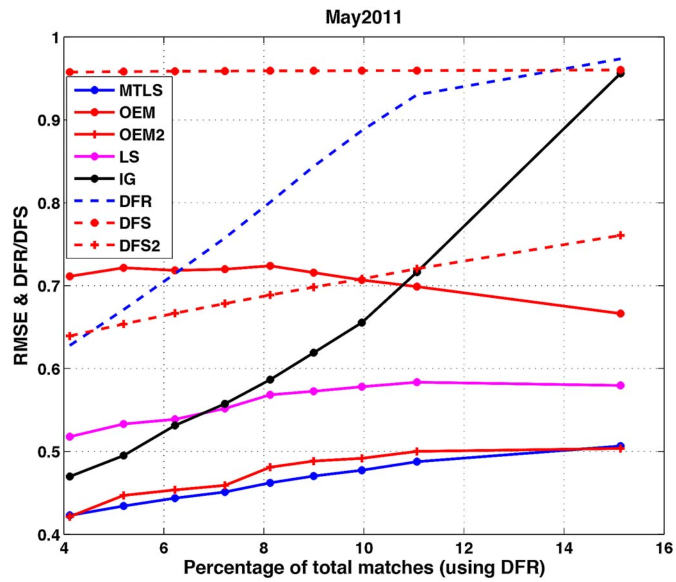

Fig. 7. Plots of retrieval errors for MTLS, LS, IG, and OEM using CRTM2.1 (RMSE ="*solid line") and information content in terms of DFS and DFR ("dashed"), as well as OEM2 (RMSE "+solid line") and DFS2 ("+dashed") using correct *a priori* error for OEM for December 2013.

VII. Information Analysis of Physical Retrieval

In the stochastic approach, according to Rodgers [2], the averaging kernel $(\boldsymbol{A})$ is the "$\mathrm{dx_{rtv}/dx_{true}}$." In the deterministic approach, this is known as the model resolution matrix $(\boldsymbol{M}_{rm})$ [12]. The analysis of $\boldsymbol{A}/\boldsymbol{M}_{rm}$ can only give some understanding of how much regularization is imposed on the inverse problem. The "trace" of the $\boldsymbol{A}/\boldsymbol{M}_{rm}$ matrix is the so-called degree of freedom from signal (DFS; stochastic) [2] or degree of freedom in retrieval (DFR) in the deterministic approach [16], which have been formulated before. The normalized DFR and DFS are plotted in Fig. 7. We have also included the least squares (LS) solution in this comparison along with MTLS and OEM because LS solutions have a sensitivity of unity.

The x-axis of Fig. 7 shows cumulatively binned data based on DFR. The reason for showing a cumulative binning rather than ordered data is that the number and value of minimum sensitivity of MTLS varies from month to month due to varying amounts of error and condition number of the matrix, depending on the data-driven regularization strength. For example, the lowest sensitivity of this month (see Fig. 7) is 0.62, but for some months we found this to be 0.4. Thus, to plot in absolute (discrete) scale is not necessarily informative and also the limited number of matches in regions of low sensitivity is another concern. Furthermore, the sensitivity statistics in discrete bins is not helpful for comparison purposes due to a wide variation in the number of matches. Thus, we have partitioned the data into 10 equidistant points between high sensitivity value of MTLS (close to 1) and the lowest sensitivity. We consider this to be appropriate primarily because we are investigating the sensitivity of the MTLS retrieval. Furthermore, it is reasonable to expect the sensitivity to display a greater range for MTLS since its regularization strength (the primary governing factor) is dynamically determined. In contrast, the sensitivity of LS is one, and is approximately 0.96 for the substantial majority of OEM retrievals, due to the static nature of the error covariance matrices. It should be noted that some decrease in sensitivity is possible in OEM since the water vapor uncertainty is initially defined as a percentage (15%) and this will result in stronger regularization when the absolute value of water vapor is low.

This calculation was performed with the latest version of CRTM 2.1. The OEM solution is worse for any subselection of data (based on the DFR of MTLS), as compared with both LS and MTLS, and does not improve for lower bins. (Note that this analysis shows inconsistent outcome for OEM in different months. Detailed time series sensitivity will be discussed in a future paper. Here, we only show one case study.)

Recall that the sensitivity of LS for all subselections is one. On the other hand, the error of MTLS is lower than those of LS and OEM, because MTLS is able to make optimal regularization for all sets of subselections. The sensitivity of MTLS varies from 0.62 to 0.98 due to data driven regularization and MTLS produces better quality SST retrieval in terms RMSE from 0.51 to 0.42. The main objective of MTLS is that it increases the regularization when the problem tends to be either highly ill-conditioned or has high error in measurement space or both. This implies that the solution stays close to *a priori*/IG when the problem is subject to severe ill conditioning and observation noise but is not required to correct a large *a priori* error, and thus, there is no scope to wrongly introduce more error than a prior error. Degrading *a priori* knowledge by use of an inappropriate retrieval method that has a high sensitivity to measurement is undesirable. One may now raise the question, based on this experiment, what such measures of sensitivity are really showing, and whether a sensitivity of $\sim$1 is actually desirable in all circumstances, given the concomitant risk of increase in noise and therefore total retrieval error. For example, one can choose DFR/DFS close to 1 in this experiment but get high retrieval error when the inverse problem has high noise and/or high ill conditioning. Conversely, one may have a lower sensitivity to measurement but have less retrieval error.

It is understandable that LS solution is deteriorating with the reduction of DFR because the problem approaches to either high ill-condition or high error or both, as discussed earlier. An interesting observation in this experiment is that the OEM solution deteriorates with a higher gradient than the same for LS and more than 70% of OEM error is higher than *a prior* error, whereas it is about 45% for LS. This again proves that even without significant RT bias (i.e., using CRTM 2.1), the OEM solution results in degraded *a prior* knowledge after measurement more than 70% of cloud free retrievals without last bin.

One may now question the OEM type of formulation that is derived using an optimization based on the *a priori* constraint of a cost function J of the form [64] as

$$J = (\boldsymbol{y}-\boldsymbol{K}\boldsymbol{x})^T \boldsymbol{S}_e^{-1}(\boldsymbol{y}-\boldsymbol{K}\boldsymbol{x}) + (\boldsymbol{x}-\boldsymbol{x}_a)^T \boldsymbol{S}_a^{-1}(\boldsymbol{x}-\boldsymbol{x}_a). \quad (31)$$

In such an assumption, the *a priori* error should not be less than the OEM retrieval error. It is not easy to objectively assign millions of different numbers in the *a priori* and measurement error covariance matrices; often only one number is assigned for the variance. Where this number matches the real life measurement/s [see (29)], in conjunction with $\boldsymbol{S}_e^{-1}$, it will satisfy the aforementioned argument, which will therefore provide the "optimum" regularization and hence solution. All points follow regularization theory, however, and the error propagated to state space when observation noise and/or condition number of the Jacobian will be unnecessarily high if the IG/*a priori* is close to truth (which will be the case for the ∼68% of data that lie within $\pm 1\ \sigma$ of the true value, even when the *a priori* uncertainty is correctly described). This may explain why the OEM error is more than *a priori error* for 72% of the data as aforementioned.

When the value of DFR of MTLS for a subselected data set is less than 0.72, implying high regularization, *a prior* error of those subselected data is still lower than for the other two retrieval methods. This behavior is anticipated by regularization theory, since the MTLS retrieval inherently regularizes the problem when the problem is ill conditioned and has high error in measurement space (including forward model, data errors), which is desirable. The most remarkable result with respect to regularization theory is that the OEM error for different subselected data is higher than that of LS, which serves as a basic reference for inverse problems with low condition number matrix. We find that the condition number of the inverted matrix $(\boldsymbol{K}^T\boldsymbol{S}_e^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1})$ of OEM is sometimes higher than for $(\boldsymbol{K}^T\boldsymbol{K})$, which is the inverted matrix in LS, thus the anticipated regularization does not always happen when applying OEM methodology to this problem. Furthermore, the term $\boldsymbol{K}^T\boldsymbol{S}_e^{-1}$ is outside of the inverted matrix in the formulation of OEM. This means that any error in measurement space is also multiplied by $\boldsymbol{S}_e^{-1}$. The effective measurement error therefore increases if the value of error variance is less than 1, which is true in our case.

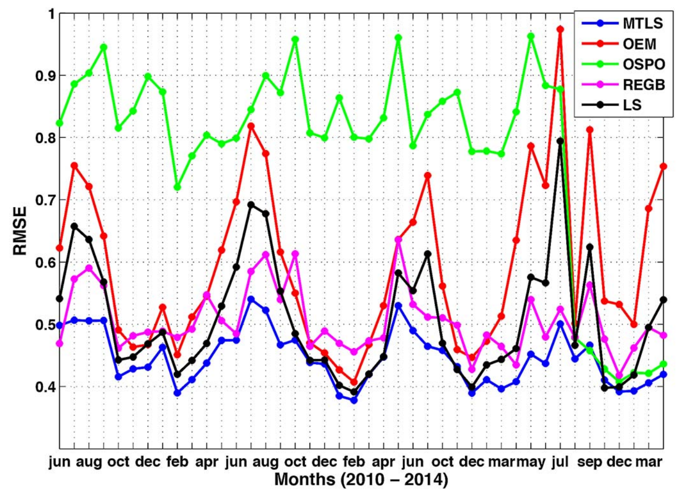

Fig. 8. Plots of RMSE for MTLS, REGB, LS, OSPO and OEM except last bin for the time series of 42 months (2010–2014).

As discussed before, there are difficulties associated with the estimation of the required OEM input observational error in an operational environment, and it requires expert skill to obtain a good estimate of *a priori* error to implement OEM and get reasonable results. There is always a debate for assigning the *a priori* error for OEM-based solutions, and even the *a priori* value itself may be debatable due to the dynamicity of the geophysical parameters. In order to ameliorate the second limitation, we have done a study where *a priori* variance for SST is defined accurately. This is done by subtracting IG values from the reference (*in situ*) values for each individual pixel, to understand the behavior of the OEM formulation under more ideal conditions (see OEM2 and DFS2 in Fig. 7). Under these circumstances, the OEM solution shows a performance comparable with MTLS, but DFS2 is significantly lower than the DFR of MTLS, particularly where *a priori* error is high. There is no practical application for such a method because the true *a priori* error is not generally available (again, there is no requirement for an inversion if the truth is known), but it provides further insight and shows that this problem does not yield a DFS close to 1 when the error covariance for OEM is optimized. Thus, the observation of relatively low DFR values ∼0.62 for MTLS when *a priori* error is low is not a matter of potential concern. It would be more of an issue when the initial guess error is high (i.e., a large increment is needed) but in this case the sensitivity of MTLS is very close to 1 (and slightly better than for the OEM).

VIII. Time Series Results

To compare retrieval performances of various methods, we have plotted time-series RMSE for 42 months in Fig. 8, where each point represents monthly average statistics excluding the last bin where excessive water vapor adjustments and cloud leakage were identified. As can be seen, MTLS consistently performs better, as compared with any other method. The second best choice in terms of retrieval performance is REGB and the operational OSPO shows the worst performance (until the last few months), which is in agreement with results presented in Figs. 1, 3, and 4. [The improved performance of OSPO for last seven months is discussed later in the paragraph]. The lowest RMSE value of MTLS is less than 0.4. The highest RMSE of MTLS in this time series is 0.55 for the month July 2012, which is mainly because of the failure of the cloud detection scheme. The cloud detection scheme is responsible for the high error of MTLS (July 2012), as well as the seasonal variations. (We have completed a separate study on the issue of the implemented Bayesian cloud detection scheme in OSPO and the cloud leakage of this scheme is rather unpredictable, at least in the operational setup. There is essentially no seasonal variation in the RMSE of MTLS under a new cloud detection scheme, which will be published in a separate paper.) One of the interesting results we found in this study is that the RMSE of OEM is always higher than LS and more than 60% of the cases of REGB. However, the good performance of LS solution is possible due to the fact that the condition number of the Jacobian is generally less than 10. Furthermore, it can be seen that the RMSE of OEM is 50% higher than MTLS error for some of the months.

The dramatic improvement in RMSE of the operational OSPO SST for the last few months coincides with the operational implementation of MTLS at OSPO in August 2013. Some minor differences are observed in this period between offline and operational MTLS studies (Fig. 8) that need to be further investigated. This present work is the first step to proceed from theoretical inverse problem to an operational environment, along with three years of offline validation and comparison against other results.

IX. Conclusion

We have demonstrated in this work the advantage of the MTLS, which is the family of the deterministic inverse methods, for producing SST retrievals compared with other prevailing methods. In addition, it is noteworthy that MTLS does not require additional error information, e.g., well-specified errors in observational and *a priori* information. This may provide a significant advantage for climate-based applications where retrievals should be as independent of external error sources as possible. The MTLS retrieval is improved by using the newer version of CRTM, which implies that more accurate forward models and ancillary data can further reduce the remaining MTLS error. This package can also calculate a metric relating to the total retrieval error and automatic QI at individual pixel level. Apart from the QI, MTLS is also capable of identifying the most difficult retrievals due to cloud contamination or high WV profile error. The sensitivity analysis confirms that MTLS solution is independent of *a priori*/IG error. The data driven dynamic regularization property of MTLS regularizes solutions toward the IG when the problem is either highly ill-conditioned or has high observation error or both to keep the solution below the *a priori* error.

It is found that OEM retrieval, at least as implemented for this problem, is worse than the LS solution, and sometimes worse than the *a priori* error, irrespective of the version of CRTM. OEM is the most popular choice for physically based operational retrievals due to the assumption that *a priori* based constraining of an ill-posed inversion should still yield reasonable results under conditions where there may be unaccounted for parameters or unforeseen errors, as may be the case in real-world retrieval problems. However, these results suggest that this view may be based more on perception of idealized Bayesian statistics rather than comparative scientific study with respect to alternative methods. This study has also demonstrated that the sensitivity of OEM retrievals under practical circumstances renders it more vulnerable to noise than MTLS retrievals. Even by employing dynamic error covariance matrices, OEM is unable to produce the best retrieval for a fairly linear and moderately ill-conditioned problem of SST retrieval. Moreover, the estimation of error of the errors, which is a prerequisite for OEM, is rather difficult in practice, which perhaps explains why OEM results do not match the expectation from the theory of adding to/constraining by *a priori* knowledge.

To date, operational SST retrievals are dominated by regression (REGB), which highly simplifies RT physics. Mostly, it does produce reasonable results (SD) due to the fact that the global variance of SST fields itself is not very high (e.g., compared with gaseous distributions) and the atmospheric attenuation for 3.9-μm channel is rather low, but such methods are still subject to biases on a spatial and temporal basis, with seasonal variations, and has no inherent means of correcting for them.

This derivation of MTLS is based on linear algebra. However, this paper illustrates that a deterministic classical mathematics approach can produce better retrievals for real-world RT problems compared with more recent probability-based mathematics that solve ill-posed problems using covariance matrices. The MTLS retrievals outperform the OEM retrievals due to the fact that the regularization in MTLS is data driven. As opposed to OEM that uses regularization from user-defined *a priori* knowledge of measurement error and forward model error, as well as *a priori* knowledge error of the retrieved target parameter. A reliable estimation of both the errors in an operational environment is very difficult due to the highly dynamic atmosphere, fast forward model error, including NCEP data, as well as error in the measurements. An alternate effort toward error estimation using simulation minus observation (S-O) bias correction leads to further ambiguities and may potentially mislead our fundamental science understanding. With the advent of newer sensors with improved multispectral capabilities (e.g., the Visible and Infrared Imaging Radiometer Suite and the future Advanced Baseline Imager), employing a deterministic physical method for simultaneous retrieval of SST and WV (critical for weather and climate studies), such as the MTLS package, has the potential to provide substantial improvements in the use of satellite data and derived products.

Acknowledgment

The authors would like to thank the members of the SST Team (J. Mittaz, J. Sapper, G. Rancic, and B. Potash) for helpful discussions. The authors also thank K. Saha, CIRA Postdoctoral Research Scientist for technical assistance. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or U.S. Government position, policy, or decision.

REFERENCES

[1] P. Le Borgne, H. Roquet, and C. J. Merchant, "Estimation of sea surface temperature from the spinning enhanced visible and infrared imager, improved using numerical weather prediction," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 55–65, Jan. 2011.

[2] C. D. Rodgers, *Inverse Methods for Atmospheric Soundings: Theory and Practice*. Singapore: World Scientific, 2000.

[3] S. Twomey, *An Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. New York, NY, USA: Elsevier, 1977.

[4] J. T. Houghton, F. W. Taylor, and C. D. Rodgers, *Remote Sounding of Atmospheres*. Cambridge, U.K.: Cambridge Univ. Press, 1984.

[5] J. L. Mead, "Parameter estimation: A new approach to weighting a priori information," *J. Inv. Ill-Posed Probl.*, vol. 16, no. 2, pp. 1–21, 2007.

[6] R. Zhang, M. K. Sen, S. Phan, and S. Srinivasan, "Stochastic and deterministic seismic inversion methods for thin-bed resolution," *J. Geophys. Eng.*, vol. 9, no. 5, pp. 611–618, Oct. 2012.

[7] M. A. Aguiló, L. Swiler, and A. Urbina, "An overview of inverse material identification within the frameworks of deterministic and stochastic parameter estimation," *Int. J. Uncertainty Quantification*, vol. 3, no. 4, pp. 289–319, 2013.

[8] A. Legendre, "Nouvelles méthodes pour la détermination des orbites des comètes," in *New Methods for the Determination of the Orbits of Comets (in French)*, Paris, France: F. Didot, 1805.

[9] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis*. New York, NY, USA: McGraw-Hill, 1978.

[10] P. K. Koner and J. R. Drummond, "A comparison of regularization techniques for atmospheric trace gases retrievals," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 109, no. 3, pp. 514–26, Feb. 2008.

[11] A. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Dokl. Akad. Nauk SSSR*, vol. 151, no. 3, pp. 501–504, 1963.

[12] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*. San Diego, CA, USA: Academic, 1989.

[13] M. A. Hanke, "Regularizing Levenberg-Marquardt scheme with applications to inverse groundwater filtration problems," *Inv. Probl.*, vol. 11, no. 13, pp. 79–95, Feb. 1997.

[14] A. Doicu, F. Schreier, and M. Hess, "Iterative regularization methods for atmospheric remote sensing," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 83, no. 1, pp. 47–61, Jan. 2004.

[15] C. Riddell, I. Buvat, A. Savi, M. C. Gilardi, and F. Fazio, "Iterative reconstruction of SPECT data with adaptive regularization," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 5, pp. 2350–54, Oct. 2002.

[16] T. Steck, "Methods for determining regularization for atmospheric retrieval problems," *Appl. Opt.*, vol. 41, no. 9, pp. 1788–96, Mar. 2002.

[17] G. Golub, "Some modified matrix eigenvalue problems," *SIAM Rev.*, vol. 15, no. 2, pp. 318–344, Apr. 1973.

[18] G. Demoment, "Image reconstruction and restoration: overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2024–36, Dec. 1989.

[19] J. Scales and L. Tenorio, "Prior information and uncertainty in inverse problem," *Geophysics*, vol. 66, no. 2, pp. 389–397, Mar./Apr. 2001.

[20] A. Tarantola, *Inverse Problems Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: SIAM, 2005.

[21] Y. Chen, Y. Han, P. Van Delst, and F. Weng, "On water vapor Jacobian in fast radiative transfer model," *J. Geophys. Res.*, vol. 115, no. D12, Jun. 2000, Art. ID. D12303.

[22] J. Li, "Temperature and water vapor weighting functions from radiative transfer equation with surface emissivity and solar reflectivity," *Adv. Atmos. Sci.*, vol. 11, no. 4, pp. 421–426, 1994.

[23] P. K. Koner and J. R. Drummond, "Atmospheric trace gases retrievals using nonlinear regularized total least square method," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 109, no. 11, pp. 2045–59, Jul. 2008.

[24] P. K. Koner, A. Battaglia, and C. Simmer, "A rain rate retrieval algorithm for attenuating radar measurement," *J. Appl. Meteorol. Climatol.*, vol. 49, no. 3, pp. 381–93, Mar. 2010.

[25] S. Chandrasekaran, G. H. Golub, M. Gu, and A. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–52, 1998.

[26] S. V. Huffel and J. Vanderwalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA, USA: SIAM, 1991.

[27] L. M. McMillin, "Estimation of SSTs from two infrared window measurements with different absorption," *J. Geophys. Res.*, vol. 80, no. 36, pp. 5113–5117, 1975.

[28] C. C. Walton, W. G. Pichel, J. F. Sapper, and D. A. May, "The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites," *J. Geophys. Res.*, vol. 103, no. C12, pp. 27 999–28–012, Nov. 1998.

[29] G. Golub and C. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.

[30] R. Adcock, "Note on the method of least squares," *Analyst*, vol. 4, no. 6, pp. 183–184, 1877.

[31] A. Beck and A. Ben-tal, "On the solution of the Tikhonov regularization of the total least squares problem," *SIAM J. Optim.*, vol. 17, no. 1, pp. 98–118, 2006.

[32] W. Karush, "Minima of functions of several variables with inequalities as side constraints," M.Sc. dissertation, Dept. Math., Univ. Chicago, Chicago, IL, USA. 1939.

[33] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in *Proc. 2nd Berkeley Symp. Math. Statist. Probl.*, 1951, pp. 481–492.

[34] R. Fletcher, *Practical Methods of Optimization*. Hoboken, NJ, USA: Wiley, 1988.

[35] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 1999.

[36] I. Markovsky and S. V. Huffel, "Overview of total least squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, Oct. 2007.

[37] J. Lampe and H. Voss, "On a quadratic eigenproblem occurring in regularized total least squares," *Comput. Stat. Data Anal.*, vol. 52, no. 2, pp. 1090–102, Oct. 2007.

[38] R. D. Fierro, G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Regularization by truncated total least squares," *SIAM J. Sci. Comput.*, vol. 18, no. 4, pp. 1223–1241, 1997.

[39] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[40] E. Haber and D. Oldenburg, "A GCV based method for nonlinear ill-posed problems," *Comput. Geosci.*, vol. 4, no. 1, pp. 41–63, 2000.

[41] B. Yegnanarayana, C. Avendano, H. Hermansky, and P. S. Murthy, "Speech enhancement using linear prediction residual," *Speech Commun.*, vol. 28, no. 1, pp. 25–42, May 1999.

[42] F. Xu and A. Ignatov, "In situ SST Quality Monitor (iQuam)," *J. Atmos. Ocean. Technol.*, vol. 31, no. 1, pp. 164–180, Jan. 2014.

[43] CRTM: v2.0 User Guide, Joint Center for Satellite Data Assimilation, Camp Springs, MD, USA, 2010

[44] D. Van-Vuuren *et al.*, "Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs," *Clim. Change*, vol. 81, no. 2, pp. 119–159, Mar. 2007.

[45] C. J. Merchant, A. R. Harris, E. Maturi, and S. N. Macallum, "Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval," *Q. J. R. Meteorol. Soc.*, vol. 131, no. 611, pp. 2735–2755, Oct. 2005.

[46] C. J. Donlon *et al.*, "Toward improved validation of satellite sea surface skin temperature measurements for climate research," *J. Clim.*, vol. 15, no. 4, pp. 353–369, Feb. 2002.

[47] P. J. Minnett *et al.*, "The marine-atmospheric emitted radiance interferometer: A high-accuracy, seagoing infrared spectroradiometer," *J. Ocean. Atmos. Technol.*, vol. 18, no. 6, pp. 994–1013, Jun. 2001.

[48] P. J. Minnett, "Radiometric measurements of the sea-surface skin temperature: The competing roles of the diurnal thermocline and the cool skin," *Int. J. Remote Sens.*, vol. 24, no. 24, pp. 5033–5047, Jan. 2003.

[49] P. J. Minnett, M. Smith, and B. Ward, "Measurements of the oceanic thermal skin effect," *Deep Sea Res. II—Top. Stud. Oceanogr.*, vol. 58, no. 6, pp. 861–868, Mar. 2011.

[50] D. Anding and R. Kauth, "Estimation of sea surface temperature from space," *Remote Sens. Environ.*, vol. 1, no. 4, pp. 217–220, 1970.

[51] P. J. Minnett, "A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space," *J. Geophys. Res.*, vol. 91, no. C7, pp. 8509–852, Jul. 1986.

[52] C. J. Merchant *et al.*, "Sea surface temperature estimation from the Geostationary Operational Environmental Satellite-12 (GOES-12)," *J. Ocean. Atmos. Technol.*, vol. 26, no. 3, pp. 570–81, Mar. 2009.

[53] P. J. Minnett, "The regional optimization of infrared measurements of sea surface temperature from space," *J. Geophys. Res.*, vol. 95, no. C8, pp. 13 497–13 510, Aug. 1990.

[54] C. J. Merchant, P. Le Borgne, H. Roquet, and A. Marsouin, "Sea surface temperature from a geostationary satellite by optimal estimation," *Remote Sens. Environ.*, vol. 113, no. 2, pp. 445–457, Feb. 2009.

[55] B. Petrenko, A. Ignatov, N. Shabanov, and Y. Kihai, "Development and evaluation of SST algorithms for GOES-R ABI using MSG SEVIRI as a proxy," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3647–3658, Dec. 2011.

[56] S. Ding *et al.*, "Validation of the community radiative transfer model," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 112, no. 6, pp. 1050–1064, Apr. 2011.

[57] J. A. Scales, P. Docherty, and A. Gersztenkorn, "Regularisation of nonlinear inverse problems: imaging the near-surface weathering layer," *Inv. Probl.*, vol. 6, no. 1, pp. 115–131, Feb. 1990.

[58] A. Berk *et al.*, *MODTRAN4 Version 3 Revision 1 User's Manual*. Bedford, MA, USA: Air Force Res. Lab., Hanscom Air Force Base, 2003.

[59] CRTM: v2.1. [Online]. Available: http://ftp.emc.ncep.noaa.gov/jcsda/CRTM/REL-2.1/

[60] X. Wu and F. Yu, "Correction for GOES imager spectral response function using GSICS—Part I: Theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1215–1223, Mar. 2013.

[61] T. J. Hewison and J. Müller, "Ice contamination of Meteosat/SEVIRI IR13.4 channel implied by inter-calibration against metop/IASI," in *Proc. IEEE IGARSS*, 2012, pp. 7197–7199.

[62] T. J. Hewison *et al.*, "GSICS inter-calibration of infrared channels of geostationary imagers using Metop/IASI," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1160–1170, Mar. 2013.

[63] J. Susskind, C. D. Barnet, and J. M. Blaisdell, "Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB Data in the presence of clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 390–409, Feb. 2003.

[64] S. A. Boukabara *et al.*, "MiRS: An all-weather 1Dvar satellite data assimilation and retrieval system," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3249–3272, Sep. 2011.

Prabhat K. Koner (M'15) received the Ph.D. degree in solar photovoltaics engineering from the Indian Institute of Technology, Delhi, India. His thesis work involved the design and optimization of solar photovoltaic arrays for water pumping in addition to a detailed socio/techno economic analysis and analog and digital electronics controller.

He has worked as a Postdoctoral Fellow with the European Union Joint Research Centre, Italy, and as a Research Scientist with the Queensland University of Technology, Brisbane, Qld., Australia on data analysis, instrument development, and quality issues of electrical grid systems. However, in the last seven years he has been actively involved in the field of satellite remote sensing retrievals. Over the last decade, he has been deeply involved in research and application of remote measurements applying physically deterministic methods. During this period, he carried his research work in different countries at reputed academic institutes such as University of Toronto, University of Bonn, and Dalhousie University. He has also worked on the NASA/ESA Mars Science Orbiter mission, reviewed articles for IEEE journals, and has given invited talks at the IEEE society. Since 2010, he has been jointly affiliated with the University of Maryland, College Park, MD, USA, and NOAA, USA.

Andrew Harris received the B.Sc. degree in physics and astronomy and the Ph.D. degree from University College London, London, U. K.

Since November 2002, he has been an Assistant Research Scientist with the Earth System Science Interdisciplinary Center of the University of Maryland, College Park, MD, USA. His research is focused in three fields, i.e., the generation of sea surface temperature products from satellite data, fundamental instrument calibration and characterization in the thermal IR, and development of high-resolution global sea surface temperature (SST) analyses. and postdoctoral work was conducted at the Mullard Space Science Laboratory, where he was involved in the development and testing of calibration and ground support equipment for the European Along-Track Scanning Radiometer (ATSR), as well as subsequent algorithm development. His work at the U.K. Meteorological Office included the development of a system to account for biases in the ATSR due to surface effects, which was adopted by the world-renowned Hadley Centre for Climate Change Research. His recent work has focused on the development of methods to work with the current generation of geostationary and polar-orbiting weather satellites operated by NOAA, as well as the development and implementation of a new high-resolution global SST analysis. He is a member of the Group for High Resolution Sea Surface Temperature (GHRSST) Science Team, chairman of their Estimation And Retrievals Working Group, and is a member of the NASA SST Science Team.

Eileen Maturi received the B.S. degree in physics from George Mason University, Fairfax, VA, USA, and the M.S. degree in oceanography from Old Dominion University, Norfolk, VA, USA.

She is an Oceanographer for NOAA/NESDIS and currently works with the National Center for Climate Prediction, College Park, MD, USA. She managers the sea surface temperature team, which generates geostationary, blended, and microwave sea surface temperatures products and oceanic heat content products.