Remote Sensing of Environment 174 (2016) 266–278

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Hybrid cloud and error masking to improve the quality of deterministic satellite sea surface temperature retrieval and data coverage

Prabhat K. Koner [a,b,*], Andrew Harris [a,b], Eileen Maturi [a]

[a] NOAA NESDIS, Center for Satellite Applications and Research (STAR), NCWCP E/RA3, College Park, MD 20740, USA
[b] Earth System Science Interdisciplinary Centre, Cooperative Institute of Climate Sciences, University of Maryland, College Park, MD 20740, USA

ARTICLE INFO

Article history:
Received 8 June 2015
Received in revised form 30 October 2015
Accepted 10 December 2015
Available online 30 December 2015

Keywords:
Cloud detection
Spectral differences method
Radiative transfer model
Bayesian cloud detection
Modified total least squares
Sea surface temperature
GOES-13 imager

ABSTRACT

In the infrared region, the quality of sea surface temperature (SST) retrievals critically depends on the cloud detection scheme. More than 5 million matchups, where the surface and top of atmosphere measurements are available, have been carefully analyzed to understand clouds related errors and to develop the advanced cloud detection scheme for improvement of satellite SST quality. The effectiveness of a Bayesian cloud detection (BCD) scheme, operationally implemented at the NOAA Office of Satellite Product Operations (OSPO) for the GOES-Imager, has been examined using an experimental filter and it is found that this scheme is not optimal. Thus, a new algorithm for cloud and error masking (CEM) scheme is proposed for physical SST retrievals. This is based on a quasi-deterministic approach combined with an approximated radiative transfer model and the functional spectral differences at pixel level. Although, traditionally the validation of cloud detection algorithms have often been reported qualitatively using visual inspection of imagery, we have made a quantitative validation of the cloud algorithm for its intended purpose by determining the quality of satellite SST retrievals against *in situ* data. Results show that CEM can reduce the *root mean square error* in SST by an average of 22% while increasing the data coverage by an average of 38% compared to the operationally implemented BCD at OSPO, as assessed over a period of fifty months.

1. Introduction

Reliable detection and elimination of cloudy pixels is a prerequisite for attempting meaningful infrared (IR) sea surface temperature (SST) retrievals. Deriving accurate cloud-masks from geostationary and polar orbiting satellite data has been a topic of research since the launch of the television infrared observation (TIROS)-1 weather satellite in 1960, *e.g.* Jedlovec (2009). Consequently, many cloud detection algorithms have been developed, based on cloud spectral properties (quasi-deterministic) and statistical (stochastic) theories. For example, a popular cloud detection technique, known as the AVHRR Processing scheme over cLouds, Land and Ocean (APOLLO), was developed for NOAA AVHRR data (see Kriebel, Gesell, Kästner, & Mannstein, 2003; Saunders & Kriebel, 1988). This method performs several threshold tests on data from visible (VIS) and IR AVHRR channels to determine if pixels are cloudy or cloud-free. Subsequently, the cLouds from Advanced Very high resolution Radiometer (CLAVR)-1 algorithm, also based on spectral and textural differences, was developed that classifies pixels in 4-km resolution images into clear, mixed and cloudy categories (Stowe et al., 1991; Stowe, Davis, & McClain, 1999). Quasi-deterministic spectral difference methods are popular, because there are some shortcomings in statistical methods, which limit their performance and prevent them from being used globally, *c.f.* Dybbroe, Karlsson, and Thoss (2005), and because of computational efficiency. For example, the neural network approach needs training sets which are region-based and Bayesian methods need information on the distribution of the data, which is often assumed to be Gaussian while it may vary regionally, Zhenglong, Li, Menzel, Schmit, and Ackerman (2007). In contrast, the quasi-deterministic threshold methods, if they are sufficiently physics-based, need not be affected by location and the distribution of the data, which makes them suitable for global use.

Standard cloud screening methods exploit, by and large, five basic properties of passive satellite imagery, Karlsson, Johansson, and Devasthale (2015): a) Clouds appear bright (*i.e.*, having high Top of Atmosphere reflectance) in VIS and near-IR channels as opposed to ice-free water surfaces and vegetation-covered Earth surfaces, b) Clouds consisting of liquid cloud particles (not ice crystals) reflect strongly in short wavelength IR and medium wavelength IR channels while Earth surfaces appear dark, c) Clouds are generally colder than Earth surfaces, d) Thin cirrus clouds have a higher transmissivity in IR channel 11 μm than in channel 12 μm which enables cirrus detection when using the split-window IR brightness temperature (BT) difference, and

* Corresponding author at: NOAA NESDIS, Center for Satellite Applications and Research (STAR), NCWCP E/RA3, College Park, MD 20740, USA.
E-mail address: Prabhat.Koner@noaa.gov (P.K. Koner).

http://dx.doi.org/10.1016/j.rse.2015.12.015

e) Broken clouds give rise to a scattered pattern or texture in images over otherwise homogeneous surfaces (especially ice-free ocean). Some of the tests are derived to exploit the abovementioned understanding of reflectance-based properties. These are not applicable at night and often fail at large satellite zenith angles and in the glint area. This is one of the reasons that many cloud detection techniques do not perform consistently well in operation. Extensive use of reflectance-based tests degrades day/night consistency and spatial uniformity of clear-sky masking results, Petrenko, Ignatov, Kihai and Heidinger (2010). Thus, we look to use a consistent cloud algorithm for both day and night, however, the performance may be expected to vary slightly because of different physics (primarily the scattering of solar radiation in the 3.9 μm channel).

The primary alternative to spectral differences is the Bayesian approach (see Heidinger, Evan, Foster, & Walther, 2012; Merchant, Harris, Maturi, & MacCallum, 2005; Murtagh, Barreto, & Marcello, 2003 and Uddstrom, Gray, Murphy, Niles, & Murray, 1999). The Bayesian cloud detection (BCD) algorithm of Merchant et al. (2005), which became operational for GOES-13, uses radiative transfer model (RTM) simulations and numerical weather prediction (NWP) information to construct the *a posteriori* probability of the pixel being clear-sky as a function of observed brightness BT and a local texture parameter. This algorithm reduces to a single test in which the above probability is calculated and compared against a predefined threshold. This reduction in the number of cloud tests is achieved at the cost of requiring a large amount of *a priori* information, including a Gaussian multivariate statistical distribution of NWP variables and an empirical probability density function (PDF) of BTs over cloudy areas, Petrenko et al. (2010).

To perform the cloud screening, most methods (either quasi-deterministic or stochastic) define thresholds in the analyzed spectral channels or channel combinations, or probability distribution, Petrenko et al. (2010). Quasi-deterministic thresholds may be static (empirically or climatologically derived) but most methods pre-calculate them stochastically by use of RTM with calculations initialized with various ancillary data (*e.g.*, satellite viewing and solar geometry information and prescribed surface temperatures and atmospheric profile data from NWP models). The stochastic method depends highly on the *a priori* data. Both the methods, either quasi-deterministic or stochastic, are constrained by their own assumptions in trying to capture the real time dynamic state of the atmosphere and surfaces. Thus, we have implemented multispectral dynamic threshold conditions based on the near real time (NRT) atmospheric condition using global forecast simulation (GFS) data to identify confirmed cloudy pixels and mask other potential errors, specific to our applications. This could potentially separate the major obvious cloud, but the difficult task is to separate out the fractional cloud and overlapping of two classes (cloud free and cloudy) due to the limitation of the number of channels of any imager as compared to the variability of the number of states. To overcome such problems, we will introduce the double differences of model minus observation criteria and RTM based single channel total column water vapor (TCWV) retrievals for cloud detection, where the effect of ambiguity in *a priori* or model knowledge are substantially reduced.

The most challenging task is the quantitative performance analysis of a cloud detection algorithm in the operational environment on a daily basis. To detect the presence or absence of cloud for a particular pixel in an image is a difficult task, especially where the number channels are limited (*i.e.* GOES-13). Hence all comparisons are relative, with the assumption that one of the cloud information sources is more accurate, and many validations are based on visually estimated cloud amounts reported by observers (*e.g.* Barnes & Hu, 2013; Kotarba, 2009). Thus, we propose in this paper an independent quantitative validation procedure to judge the quality of the cloud detection scheme based on the deterministic single channel SST retrieval collocated with buoy temperature. The number and global distribution of good buoy measurements at present is sufficient to provide a reasonable assessment of any cloud detection algorithm, at least one designed for the purposes of SST retrieval.

The main aim of this paper is to improve the quality of operational SST product using a better cloud detection scheme, which can be assessed by overall validation. The operational satellite retrieved SST includes two distinct different errors: 1) cloud detection error and 2) retrieval method related error. Sometimes the cloud detection scheme is mutually consistent with the applied SST retrieval method. In IR remote sensing applications, cloud (especially fractional cloud) cannot easily be separated as an individual entity. We cannot yet model cloud occurrence (using a separate variable) accurately with inverse theory for the limited number of imager measurements. Here, we argue that assessing the accuracy of SST retrieval with respect to buoy measurement is valid and a sound approach to quantify the quality of cloud detection algorithms. The quality of SST with respect to collocated buoy measurement has been already used to understand the performance of cloud detection algorithms (Barnes & Hu, 2013). To understand the quality of cloud detection in depth, we will discuss the quality of SST retrieval using deterministic modified total least squares (MTLS) method

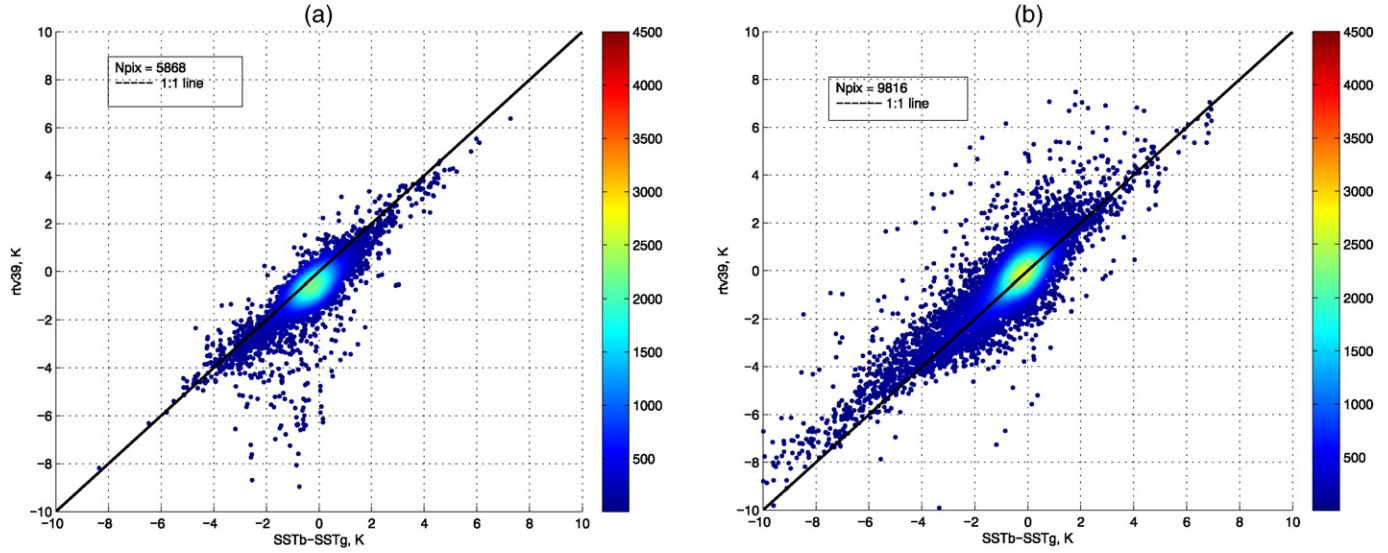

Fig. 1. Plot for $rtv_{3.9}$ where pClr > 0.98 against the differences of SSTb and SSTg, which are the buoy temperature and the given initial guess for CRTM calculation. (a) For night and (b) for day. The number of color bar stands for per K^2 and Npix stands for number of pixels.

Koner, Harris, and Maturi (2015a), and stochastic methods of optimal estimation (OE) and regression under different cloud detection schemes in this paper.

We will propose the experimental filter to quantify the performance of the cloud detection algorithm in Section 3, then describe the new cloud and error masking(CEM) algorithm in Section 4. Our proposed quantitative validation scheme for the performance will be discussed in Section 5. Two cloud detection schemes, namely BCD and CEM will be compared using different retrieval methods beyond Section 7.

2. Description of data and forward model

The satellite SST and buoy match-ups operationally generated at the Office of Satellite Products and Operations (OSPO) at NOAA are used here. The match-up window of this monthly matchup database (MMDB) was set to ±30 min for buoys (drifters, coastal and tropical moorings) coincident with satellite pixel point (unlike some conventional matchup criteria of ±3 h and a spatial resolution of 10-km). This MMDB provides BTs from the GOES-13 imager and National Centers for Environmental Prediction (NCEP) surface and upper-air forecast fields using the global forecast system (GFS) model (4× daily model runs, 3 & 6 h forecast). We use GFS data instead of reanalysis profiles data to ensure the methodology will be easily adaptable to the operational environment. The GFS data are used as the initial guess (IG) of SST and TCWV, as well as providing input for the radiative transfer modeling. The *in situ* data in our MMDB are retrospectively quality controlled using corresponding quality flags from NOAA *iQUAM* (Xu & Ignatov, 2014). We use a fast forward RTM (FFRTM), the Community Radiative Transfer Model (CRTM), in the operational environment to reduce computational cost. Simulated BT's were calculated employing the CRTM v2.1 (http://ftp.emc.ncep.noaa.gov/jcsda/CRTM/REL-2.1/) and GFS profile data for the development of our experimental filter, and input for the SST retrieval using MTLS and OE methods. We have also used the CRTM-derived partial derivatives (Jacobians) of the channel BTs with respect to surface temperature ($\delta y_\lambda/\delta s$) and logarithm of TCWV ($\delta y_\lambda/\delta \log(w)$ where w = TCWV). However, the detection of nighttime aerosol contaminated pixels is not really feasible by employing any spectral differences available in this study. Thus, we use the CMIP5 climatological aerosol profiles [©RCP Database (Version 2.0.5)] (Van-Vuuren et al., 2007) for forward model simulation to reduce nighttime retrieval errors only, which will serve to partly alleviate the issue of aerosol contamination. A constant offset of −0.17 K to account for the skin to bulk SST differences of buoy was used as a first-order approximation for night cases only (*e.g.* Donlon et al., 2002).

3. Experimental filter

In this section, we propose an experimental filter (EXF) for examining the quantitative performance of prevalent BCD schemes. It is straightforward to develop a test to understand cloud contamination when two exactly collocated measurements at the top of the atmosphere and sea surface/ground are available. The experimental filter is designed using a simple physical understanding of the noise in measurements. We assume the buoy temperature is 'true' (after discarding bad buoys using iQuam) and the 3.9 μm measurement is not affected by residual variability of water vapor, Koner and Harris (2015b). The model minus observation differences of the 3.9 μm channel are divided by the partial derivative with respect to SST values ($rtv_{3.9}$) to transform them from measurement space to state space and allow a direct comparison with "Buoy (SST_b) minus IG of SST (SST_g)". Provided the pixels are cloud-free then we expect the retrieved and true SST innovations to closely follow a 1:1 ratio. The noise in state-space due to error in the RTM (fast forward RTM and GFS errors) and measurement noise is approximated with some threshold. We have used a 1 K noise threshold for experimental purposes, to approximately separate the cloud-free and the cloudy pixels. The gradient function is typically between 0.7 and 0.9, so channel noise (typically ~0.15 K) is only modestly magnified. The total magnitude of the atmospheric deficit in the 3.9 μm channel is typically 2–4 K so a water vapor profile error of as much as ~25% should still be accommodated within this threshold, more discussions will be made in Section 7.

A Bayesian cloud-mask, Merchant et al. (2005) was already built-in as part of the operational GOES-13 SST retrieval system at OSPO. As discussed our earlier paper, Koner et al. (2015a), the quality of SST retrieval improved drastically using additional quality constraints at the solution time of MTLS, which indicated a significant cloud leakage problem in the BCD scheme. To further illustrate this issue, Fig. 1 plots $rtv_{3.9}$ against "SST_b minus SST_g", for pixels with probability of clear sky (pClr) greater than 0.98, which is the operational threshold used for the BCD. For the month of March, 2012, the number (5868 for night and 9816 for day) of observations from our MMDB is flagged as cloud-free by using a probability threshold (pClr > 0.98). However, for some points where "SST_b minus SST_{IG}" values are close to zero, the values of $rtv_{3.9}$ are less than −5 K. There are two possible causes for these large errors: (a) cloud contamination in supposedly cloud-free measurements, or (b) error in the forward model calculations. Since the spectral transmittance coefficients used in the CRTM are derived using regression, it is possible that the errors come partly from the tail end of the distribution in the training set. However, considering the high and varying values of BT differences, it is more likely that they arise from fractional clouds in the specified cloud-free observed data.

Fig. 2 plots the same parameters as in Fig. 1 but for measurements where pClr ≤ 0.98 (considered as cloudy according to the operational threshold set on the BCD) and those are classified as good by EXF. It is observed in Fig. 2 that a large number of pixels discarded as cloudy by the BCD are in reality good measurements (*i.e.*, closely follow the 1:1 line), assuming that the forward model calculation is sufficiently accurate. This number (7516 for night and 6377 for day) is statistical significant, indeed it is comparable to the number of pixels (5868 for night and 9816 for day) flagged as cloud-free by BCD in Fig. 1. On the other hand, more than 13% for night and 18% for day of cloud-free pixels in Fig. 1 are found to be cloud contaminated (*i.e.* lie outside the ±1 K range) by the EXF. One explanation for the small bias observed in Fig. 2a for nighttime is due to 'delta' residual cloud in the 3.9 μm channel and incorrect aerosol input from CMIP5 data. In the daytime, the absorption and solar scattering effect due to water vapor in the measurement of 3.9 μm channel compensate and more discussions will be made later. It should be noted that EXF is not applicable for an actual cloud detection algorithm since it requires an *in situ* ("truth") measurement, it is just a diagnostic test for the effectiveness of the cloud detection algorithm. Overall, it shows that ~32% (night) and 37% (day) of total matches are effectively cloud-free (*i.e.* a retrieval of SST is viable according to the EXF) as opposed to 15% for night and 25% for day reported by the Bayesian technique (Fig. 1). These results inspire us to develop a new cloud detection algorithm.

4. New algorithm for cloud and error mask

Cloud and error mask algorithm is described using four distinct different filters: 1) functional spectral differences thresholds; 2) double difference filter; 3) retrieved TCWV threshold; and 4) spatial filter.

4.1. Functional spectral differences

The detection of major opaque clouds is certainly possible by using only the combination of spectral differences between various channel 6.7, 11, 13.4 μm, *cf.*, Ackerman et al. (1998). Thus, detection of these clouds does not require forward simulations. This reduces the need of forward simulation, since only cloud-free regions that pass these tests require the model to be run. Besides the need of reduced number of forward simulations, the main advantage of spectral difference methods is that they depend only on measurements that pass through the same

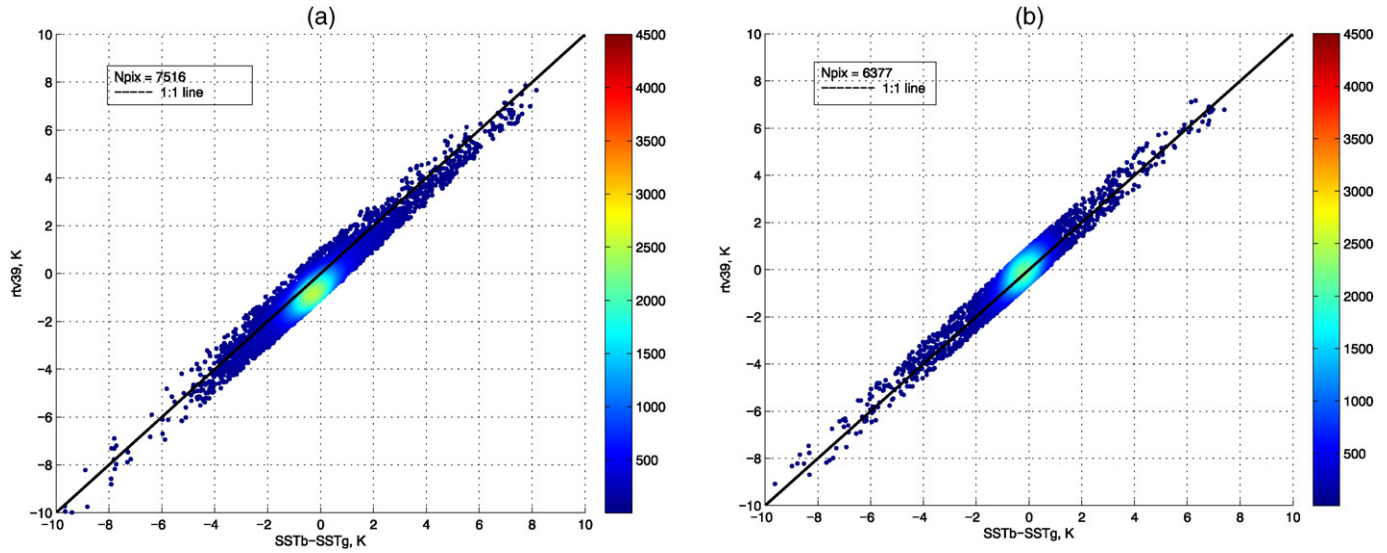

Fig. 2. The differences of temperature plot same as Fig. 1 using EXF passed pixels but pClr ≤ 0.98 (*i.e.* deemed cloud-contaminated by Bayesian methodology).

unique atmosphere, which greatly reduces the impact of absolute modeling error. This, along with the available major influencing atmospheric parameters (*e.g.*, TCWV), allows us to model dynamic thresholds without requiring explicit case-by-case forward model calculation.

To perform cloud screening, most methods define thresholds in the analyzed spectral channels or channel combinations. For example, there is a mean difference of 30 K between the 6.7 and 11 μm BTs for a standard cloud free atmosphere over the tropical ocean, according to the study of MODIS cloud clustering classification, Zhenglong et al. (2007). This difference will vary with the TCWV amount, and case studies have been made using 'sonde WV data, *e.g.*, Hutchison, Hardy, and Gao (1995). To understand this for the GOES-13 imager, we have plotted the difference of 6.7 and 11 μm BTs with respect to TCWV for all data points for the same month in 'blue' and same for the cloud free model calculated BTs top of this plot as 'red' as shown in Fig. 3.

The difference BTs of 6.7 and 11 μm with TCWV is able to detect the obvious cloud measurement using the functional threshold of TCWV as shown in Fig. 3 by the 'green' line. The absorption at ~6.7 μm is much higher than ~11 μm, so that when the atmosphere is very moist (TCWV ≥ 50 kg/m^2) the difference can reach up to 50 K. Similarly, the difference will be less for a drier atmosphere, and can drop down to ~30 K for the GOES-13 imager. This is the main problem associated with fixed threshold cloud detection algorithms and most of the operational cloud detection algorithms that are based on fixed lookup thresholds (*e.g.* Saunders & Kriebel, 1988; Stowe et al., 1991; 1999), may sometimes include adjustment based on observed BTs using offline calculation, rather than using NRT TCWV. Generally, high latitude regions are dry while tropical atmospheres are moist, thus fixed threshold techniques typically discard many good measurements in the high latitudes and fail to detect clouds in the tropics. As a result, the general perception is that the high latitude areas are heavily cloud covered, which may not actually be the correct scenario. Sometimes a latitude-dependence is used in the threshold calculation; however, a functional TCWV-dependent threshold should be better choice. To the best of our knowledge, no published spectral difference cloud detection operational scheme has included TCWV (NRT) as a functional parameter as yet. We take this opportunity to include TCWV (using operationally available GFS data) as a parameter to develop the threshold conditions for various spectral difference spaces. Our implementation is conservative in such a way that only confirmed cloud will be discarded if we develop the functional relation below the green line of Fig. 3. We also consider the normalized spectral differences (divided by average BT value of selected two channels), which gives an opportunity to find a "unitless" unique functional threshold value and may be used for different instruments. For the GOES-13 "11 and 6.7 μm" channel pair, we develop the following relation of threshold condition for normalized spectral difference:

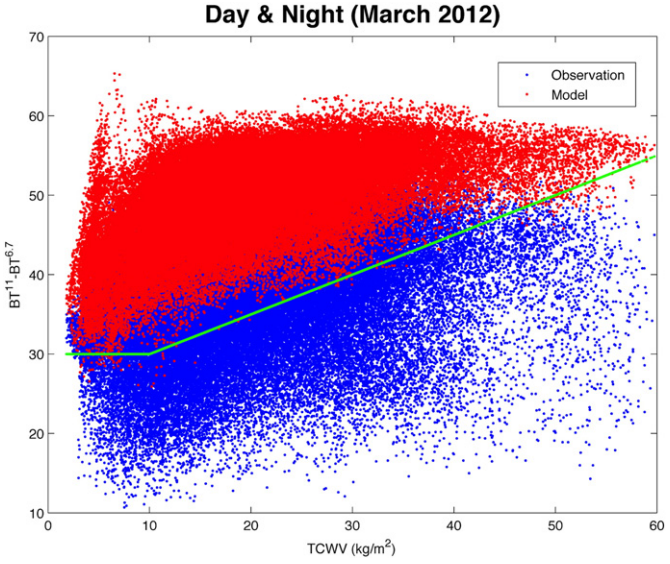

Fig. 3. Plot of modeled (red) and measured (blue) BT differences of '6.7 & 11 μm' with respect to TCWV and tentative threshold (green line) for discarding cloudy pixels.

$$\frac{2\left(T_{11}^{m}-T_{6.7}^{m}\right)}{T_{11}^{m}+T_{6.7}^{m}} > a_1 + \max\left(\frac{TCWV-b_1}{c_1}, 0\right) \quad (1)$$

where, T^m stands for the BT of measurement and a_1, b_1 *and* c_1 are coefficients of thresholds for GOES-13 and are given in Table 1.

The above-mentioned test eliminates ~10% of total pixels as obvious cloudy for matchups from, *e.g.*, March 2012. Note that the statistics vary for different months. For example, around 20% obvious cloud pixels are removed using the same constraint for the month of October 2011. It can be argued that the spectral difference of two measurements under cloud-free conditions is dependent on the shape of the profile of water vapor. However, we have approximated the radiative transfer

Table 1
Coefficients of the newer cloud-detection thresholds in Eqs. (1) through (4) with TCWV as a varying parameter.

Sensor: GOES-13 Imager				
Indices (i) of coefficients for Eqs. (1)–(4)	1	2	3 (Threshold envelop)	4
a	0.1	0.05	−0.006	0.004
b	20	10	30	15
c	600	2000	3000	1500

equation of spectral differences with a function of TCWV to determine the threshold for eliminating absolutely sure cloudy pixels. Thus, this test by itself is not capable of detecting all cloudy pixels and other tests are required to increase cloudy/clear discrimination. Similar to Eq. (1), we develop a relation for the threshold condition for normalized spectral difference of GOES-13 "11 and 13.4 μm" channel pair from:

$$\frac{2\left(T_{11}^{m}-T_{13.4}^{m}\right)}{T_{11}^{m}+T_{13.4}^{m}} > a_2 + \max\left(\frac{TCWV-b_2}{c_2}, 0\right) \tag{2}$$

It is observed from this study that 15% of total pixels can be eliminated by combining these two tests for "obvious cloud" for March 2012. Again, this statistic will change for different months or the same month of a different year for the presence of particular cloud. For example, the 13.4 μm channel is efficient for detecting high thin clouds over ocean, Ishida and Nakajima (2009). Use of the 13.4 μm channel provides an opportunity to detect low-altitude clouds, which is otherwise difficult using only window channels.

Similar to Eq. (1), we develop the relation of a threshold condition for normalized spectral difference of GOES-13 "3.9 and 11 μm" channel pair. The bi-spectral difference of "3.9 & 11 μm" is a good technique to detect fractional and thin clouds, *e.g.* Jedlovec, Haines, and LaFontaine (2008). The emissivity and reflected component of 3.9 μm BT varies with the amount and type of clouds. For example, as compared to 11 μm BT, 3.9 μm BT is higher in the daytime and lower at night in the presence of ice cloud or fog. This test is very important for the application of SST retrieval by eliminating pixels where 3.9 μm observed BT is affected by large solar scattering/reflection in the daytime, because it is still desirable to use the information-rich 3.9 μm channel for daytime SST retrieval. We implement spectral difference of "3.9 & 11 μm" using the "min–max" principle in conjunction with TCWV for calculating threshold conditions of the normalized spectral difference of "3.9 & 11 μm". The use of "min–max" principle allows us to use an envelope of threshold conditions rather than a fixed number:

$$\frac{2\left(T_{3.9}^{m}-T_{11}^{m}\right)}{T_{3.9}^{m}+T_{11}^{m}} > a_3 + \max\left(\frac{TCWV-b_3}{c_3}, 0\right) \tag{3}$$

$$\frac{2\left(T_{3.9}^{m}-T_{11}^{m}\right)}{T_{3.9}^{m}+T_{11}^{m}} < a_4 + \max\left(\frac{TCWV-b_4}{c_4}, 0\right). \tag{4}$$

After using the above four conditions, it is possible to eliminate 42% of the total pixels as cloudy or highly erroneous as shown in Fig. 4 for the month of March 2012. (Again, note that the percentage value may be expected to be time-variant.) Fig. 4 shows the density distribution of pixels with a composite variable in x-axis (arbitrary units) for rejected and selected classes as determined by tests 1) through 4). Rejected pixels have negative x-values, while selected ones are positive. For each of these categories, the cloudy and clear density distributions as determined by the EXF with same variable of x are also shown (red and blue respectively). It can be observed from Fig. 4 that the aim of the above four filters to discard the pixels those are confirmed cloud, has been satisfactorily executed with only a modest false alarm rate.

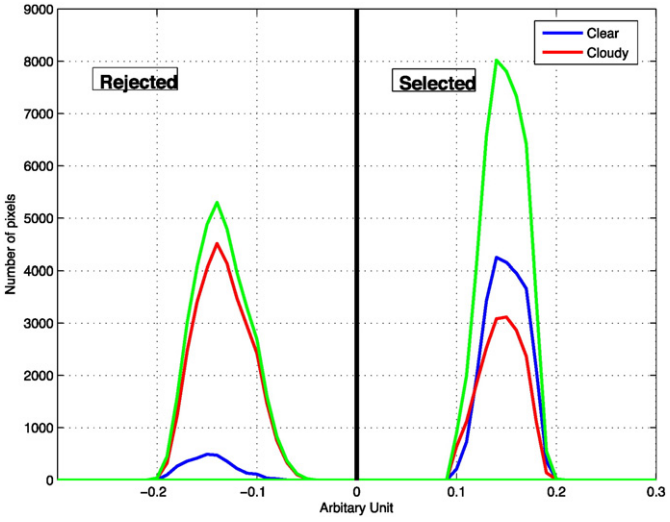

Fig. 4. Density plots the group of rejected and selected pixels with a composite of the cloudy (red) and clear (blue) pixels according to the EXF.

4.2. Double differences filter

Fig. 4 also demonstrates that spectral differences techniques alone cannot completely separate out the cloudy pixels from a set of measurements, because the group of selected pixels still contains a significant number of cloudy pixels according to EXF. There are also many additional sources, other than the obvious clouds, contributing to the SST retrieval error. The main difficulties associated with good SST retrievals are the detection of (a) fractional clouds, (b) clouds over cold water (Barnes & Hu, 2013; Hu et al., 2009) and (c) error in the fast forward modeling including those due to ancillary data and unaccounted parameters (*e.g.*, aerosols). It is difficult to retrieve an accurate SST when an atmospheric temperature inversion occurs near the surface, and it is also difficult to detect such pixels using algorithms based on radiance thresholds and/or cloudy PDF without help of a RTM. For the lack of a better alternative within the scope of this study, we use an empirical filter to reduce the erroneous retrievals as follows. We plot the double difference values of simulation ($e.g. T_{3.9}^{s}$) and measurement for the spectral differences of "3.9 & 11 μm" channels $(T_{3.9}^{s}\text{-}T_{11}^{s})\text{-}(T_{3.9}^{m}\text{-}T_{11}^{m})$ in Fig. 5 against TCWV for all matchups of March 2012 as black dots. A density plot of these differences is overlaid on top of the previous plot in Fig. 5 for the pixels that are determined cloud-free by EXF. Then we employ a conservative constraint of ±2 K to retain a major portion of the cloud-free pixels. It can be seen from Fig. 5 that some cloud-free pixels are discarded, but a large majority of cloudy pixels are filtered at the cost of relatively few cloud-free pixels.

This double differences filter (DDF) requires model calculations but ensures the dependency on *a priori* data is low. CEM, similar to BCD, is always dependent on the accuracy of RTM in terms of approximation of the radiative transfer physics and quality of GFS data.

Behavior of DDF filter:

- During daytime, under sun-glint condition and in the presence of fractional cloud and aerosol, 3.9 μm BT is much higher than 11 μm BT (due to solar scattering). This results in a measurable difference between the two single-channel retrievals and it is easy to detect affected pixels using thresholds.
- The second parameter in our SST retrieval scheme is TCWV, for which the underlying assumption is that, if the true shape of the WV profile is known and the equation of state is linear, it may be retrieved by a single iteration. These two assumptions do not always hold for the presently employed forward model. There may be instances where for a given pixel the GFS WV profile shape and the total amount are

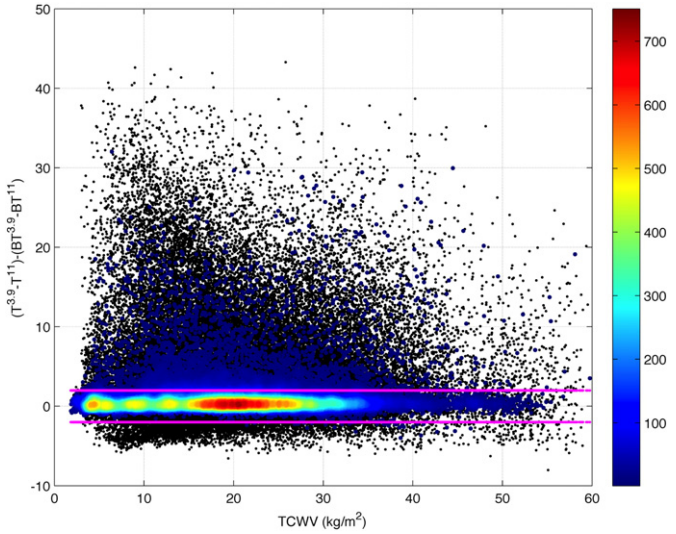

Fig. 5. Double differences of model and measured of 3.9 and 11 μm channels of GOES-13 imager for all matchups (black) and density plot of the same for pixels determined as cloud-free using experimental filter.

quite far from the truth, propagating error into the SST retrieval. This error generally affects the 11 μm BT more than those for the much more transparent 3.9 μm channel. By using appropriate thresholds in our filter, such pixels are also detected and removed.

4.3. Retrieved TCWV thresholds

It may be possible to achieve near-perfect SST retrievals by supplying accurate atmospheric data into a fully physical forward model. However, we are currently using a regression-based fast forward model due to operational constraints, and statistical error is unavoidably embedded in single sensor level calculations. There will also be some error in the input atmospheric data (GFS). As we reported in our previous publication, the physical deterministic MTLS method can retrieve good SST, Koner et al. (2015a), even for a significant departure in TCWV from truth if the shape of the profile of WV is reasonable, because the TCWV is in the retrieval vector. However, if the profile shape is not representative, excessively large adjustments in TCWV may be retrieved to compensate for the concomitant change in modeled brightness temperatures. To resolve such issues, we implement a conservative filter to discard such pixels, which are likely to result from large differences between the shapes of GFS and true WV profiles, as well as some residual clouds which were not detected by spectral differences. For this filter, first we calculate the TCWV using a linear relation of the radiative transfer model calculation and measurement in the 11 μm channel:

$$rtv_{TCWV} = \frac{T_{11}^{m} - T_{11}^{s} - K(2,1)rtv_{3.9}}{K(2,2)} \tag{5}$$

where, $K(2,1)$ and $K(2,2)$ are the SST and TCWV Jacobian and T_{11}^{s} is the simulated BT of the 11 μm channel using CRTM v.2.1. The single channel (3.9 μm) deterministic SST retrieval ($rtv_{3.9}$) is calculated as

$$rtv_{3.9} = \left(T_{3.9}^{m} - T_{3.9}^{s}\right)/K(1,1) \tag{6}$$

where $K(1,1)$ is the SST Jacobian and $T_{3.9}^{s}$ is the simulated brightness temperature of the 3.9 μm channel using CRTM v.2.1. The values of rtv_{TCWV} are plotted against TCWV for all matchups in black and the density plot of the rtv_{TCWV} values for the pixels are determined as cloud-free by EXF in Fig. 6. As we mentioned earlier, the TCWV is replaced by log (TCWV) in our retrieval scheme, Koner et al. (2015a), thus the innovation values shows in Fig. 6 correspond to the exponential value minus one times the original value of TCWV. For example, an innovation of one gives an adjustment of 1.71 × TCWV. We are fully aware that we have made a reduced state vector retrieval model due to a small number of measurements and it is not possible to solve for the profile shape of the water vapor using only the four channels available for the GOES-13 imager. Thus, we choose to discard such measurements, which are not adequately solvable using any community accepted RTM-based SST retrieval method. The threshold value of ±1 for log(TCWV) is considered as a valid range as shown in Fig. 6 even though some good measurements (according to EXF) are discarded. It can be seen from the density plot in Fig. 6 that the major portion of the cloud-free pixels according to EXF are retained in the CEM algorithm with this TCWV retrieval threshold filter (WVTF).

4.4. Spatial filter

We have employed a three-by-three spatial homogeneity filter to remove pixels affected by cold fractional clouds both during day and night. The values of the eight pixels surrounding the target pixel are compared. We have employed two different threshold conditions for the adjacent nine measurements: a) the "maximum *minus* minimum" of these measurements should be less than 5 K; b) the value of the target pixel must be within 0.6 K of the maximum value. These values represent tuning parameters which we have set to preserve oceanic fronts by increasing the threshold value, which may result in increasing cloud leakage, or choosing a lower threshold value to improve the quality of SST retrievals at the cost of increased false alarms. Finally, we also implemented a simple filter to remove the snow/ice surface and wrong measurements by restricting the measurement of 3.9 μm BT to be greater than 271.16 K.

5. Quantitative analysis of cloud detection

The coefficients as described in Eqs. (1)–(4) are calculated using cloud-free pixels according to the abovementioned EXF for the month of match-ups data of October 2011. We chose a relaxed constraint to determine these coefficients to retain the maximum number of cloud-free pixels. The main aim for these tests is to develop an equation using a relational operator for elimination of confirmed cloud. The calculated coefficients for CEM are shown in Table 1. These coefficients are used in our time series and all other validation purposes. Although, these

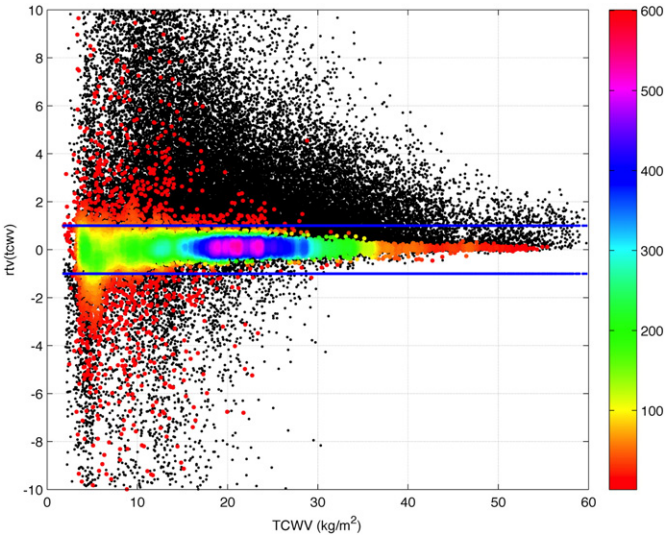

Fig. 6. Single channel TCWV retrieval from 11 μm using single channel SST retrieval from 3.9 μm channel for all matchups (black) and density plot of same for the pixels determined as cloud-free using experimental filter.

coefficients are calculated for GOES-13, the offline study shows that these coefficients can be used for other sensors, *e.g.* MTSAT or MODIS, depending on available channels. Note that the threshold value for Eq. (2) independently calculated in this study are consistent with those reported for GOES-13 by Walker, Mackenzie, Mecikalski, and Jewett (2012).

The performance of CEM is shown in Fig. 7. The number of false alarms (identified by pixels which pass the experimental filter shown in Fig. 2a that are still flagged as cloud using BCD) can be reduced from 7516 to 2547 pixels for night cases using CEM (figure not shown). The number of false alarms cannot really be zero due to the constraints of DDF and WVTF as discussed before. The comparative statistics of CEM and BCD for the month of March 2012 are given in Table 2. Nevertheless, the cloud-free class increases from 5868 to 11,871 pixels as compared to the presently implemented BCD in OSPO (see Fig 7a), which is a coverage gain of more than 100%. The calculated cloud leakage (based on the EXF) for night cases is 10% with respect to the number cloud-free pixels, comparable to the number for BCD, but it can be observed from figures (Fig. 1a & Fig. 7a) that the more severe cloud leakage is reduced. These statistics vary depending on the selected month of the matchups under analysis, but the overall trend is the same.

Fig. 7b shows CEM daytime performance with respect to the EXF. Also, the false alarm rate for the day cases reduces somewhat from 6377 to 5093 (20% reduction, figure not shown), which is not as significant as is observed for night, and the total data coverage of both cloud schemes are comparable. The improved performance of the daytime OSPO cloud masking is primarily because of use of the additional instantaneous visible measurements. CEM can be extended using a visible channel to improve the daytime cloud detection in the future when a reasonable fast forward model for the channel is available. However, the cloud leakage according to the EXF is still reduced significantly (from 20% to 6% with respect to the total number of cloud-free pixels) using CEM algorithm as compared to implemented BCD. While the statistics vary for different months, generally the CEM outperforms over BCD regarding the following criteria:

Table 2
The number pixels (Npix) of cloud-free, and leakage and false alarm which are based on EXF for CEM and BCD.

	Npix	Cloud-free	Leakage	False alarms
CEM	Night	11,871	1305	2547
	Day	9752	614	5093
BCD	Night	5868	581	7.516
	Day	9816	2195	6377

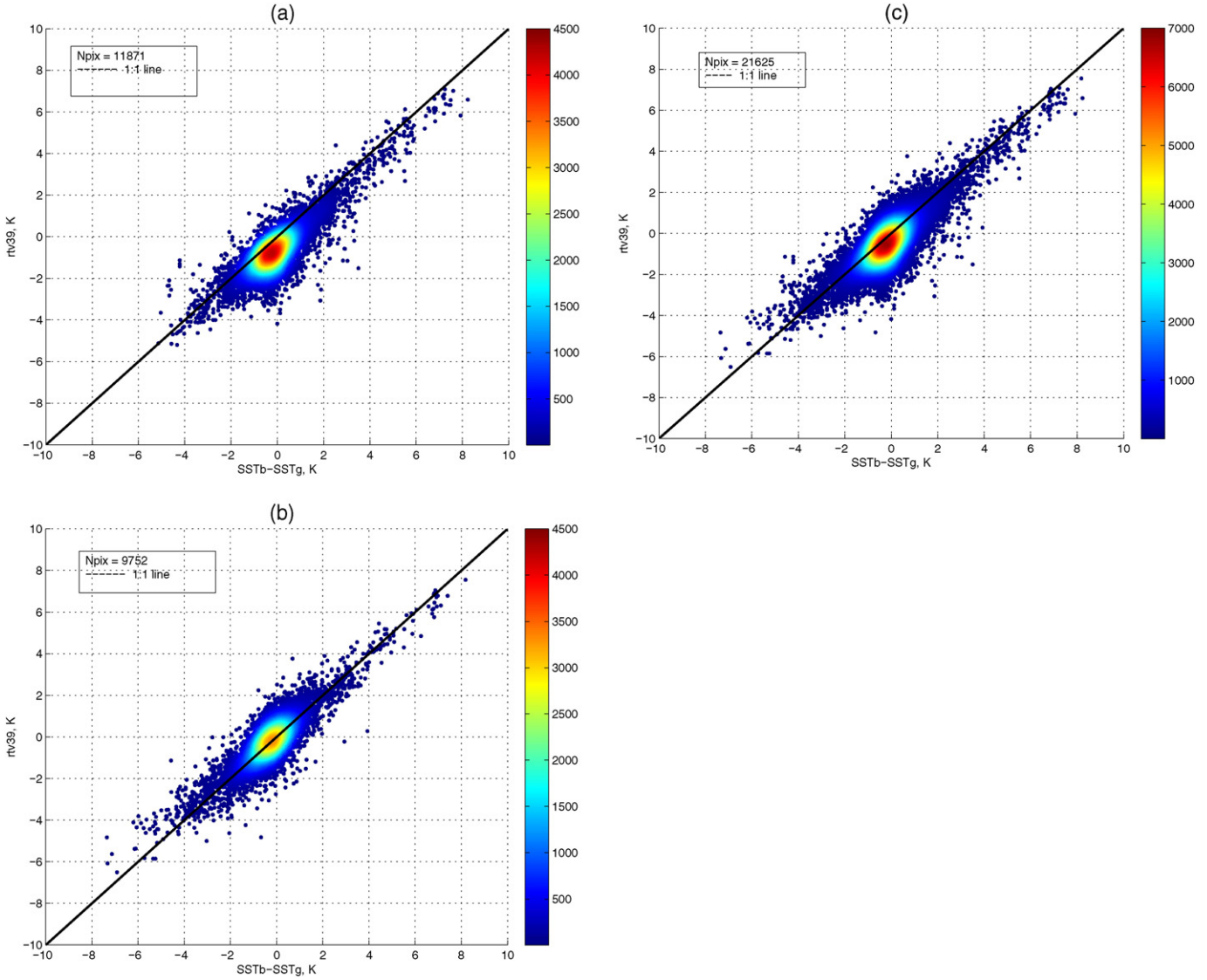

Fig. 7. Plot for $rtv_{3.9}$ against the differences of SSTb and SSTg under CEM determined cloud-free pixels: (a) for Night, (b) for Day and (c) for composite Day and Night.

a) Data coverage for daytime is slightly improved or comparable, but there is significant improvement of data coverage for nighttime cases.
b) Cloud leakage (failure to detect) decreases significantly for daytime and a little for nighttime, but severe cloud leakage is reduced.
c) False alarms decrease for both day and night, but most significantly for nighttime.

We have used a consistent cloud and error mask (CEM) for both day and night, however, the performance varies slightly between day and night because of different physics (primarily the scattering of solar radiation in the 3.9 μm channel). Fig. 7c shows the day and night composite performance under CEM. The false alarms can be reduced from 15,061 to 8590 (~43% reduction with a base count of BCD cloud-free) according to the EXF and data coverage increases from 15,684 to 21,625 (~38% improvement with respect to BCD). Note that the previous numbers were obtained after discarding the pixels in the last bin of MTLS solution (additional feature of MTLS for some cloud elimination as discussed in Koner et al., (2015a). We cannot yet model cloud occurrence accurately using any inverse theory, thus it is almost impossible to remove all cloud leakage and preserve all cloud free pixels using only the few channels available from GOES-13 imager measurements. Additionally, forward model error due to erroneous ancillary profiles data from GFS and stochastic approximations of radiative transfer physics restrict the performance of the quasi-deterministic cloud algorithm. The performance of the CEM will be improved with advanced sensors where more channels are available, and with reduced forward model error.

In addition to the deficiency in the BCD scheme as shown above, another potential demerit of such an approach is the requirement of forward simulation for all the pixels. In practice, the simulations are done for available NCEP grids and then interpolated (which introduce further approximations to the "true" radiative transfer function) to each pixel at the BT level, rather than the input level, introduces additional ambiguity.

6. Various SST retrieval methods

Four methods of sea surface temperature retrieval are considered as described in our recent publication (Koner et al., 2015a). Four methods are:

a) Modified total least squares (MTLS)
b) Optimal Estimation method (OEM)
c) Regression method with cal-val with buoy (REGB)
d) Operational SST retrieval (OSPO)

All coefficients and parameterizations for all methods are kept identical as described in the abovementioned reference. The basic form and some modification in MTLS method are described below.

$$\boldsymbol{x}_{mtls} = \boldsymbol{x}_{ig} + \left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}^2}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T\Delta\boldsymbol{T}_\delta \tag{7}$$

The analytical error calculation using (Koner & Drummond, 2008a; Koner et al., 2015a)

$$\|e\| = \left\|(\boldsymbol{M_{rm}}-\boldsymbol{I})(\boldsymbol{x}_{true}-\boldsymbol{x}_{ig})\right\| + \left\|\left(\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}}\sigma_{end}^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}^T\right\|\left\langle\left\|(\Delta\boldsymbol{T}_\delta - \boldsymbol{K}(\boldsymbol{x}_{true}-\boldsymbol{x}_{ig}))\right\|\right\rangle \tag{8}$$

where, $\Delta\boldsymbol{T}_\delta = \boldsymbol{T}_\delta - f(\boldsymbol{x}_a)$ is the model minus observation; $\boldsymbol{x}_{ig} = \boldsymbol{x}_a$ is initial guess or *a priori* and $\boldsymbol{T}$ is a measurement vector of GOES-13 channels with elements of $[y_{3.9}\ y_{11}\ y_{13.4}]$ for brightness temperature of 3.9, 11 and 13.4 μm, $\boldsymbol{M_{rm}} = (\boldsymbol{K}^T\boldsymbol{K} + \frac{2\log(\kappa)}{\gamma_{snr}^2}\sigma_{end}^2\boldsymbol{I})^{-1}\boldsymbol{K}^T\boldsymbol{K}$ is the model resolution matrix, $\boldsymbol{I}$ is the identity matrix. $\boldsymbol{x}_{true}$ is assumed to be true parameter, $\boldsymbol{K}$ is the Jacobian, σ_{end} is lowest singular value of $[\boldsymbol{K}\ \Delta\boldsymbol{T}_\delta]$, κ is the condition number of the Jacobian, and $\boldsymbol{x} = \begin{bmatrix} s \\ w \end{bmatrix}$ is used, where, as before, s is SST and w is log(TCWV).

The quality indexing (QI) of retrievals is calculated using the value of total analytical error (Eq. (8)) in MTLS. The binning based on the fixed value of analytical error ($0 < \sqrt{||e||} < 1$) into 10 bins evenly spaced in a linear scale. If a bin does not get at least 10% 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. MTLS has additional advantage to permit detection of bad retrievals, due to fractional cloud or error in model simulation, and place in the last bin as discussed elaborately in the abovementioned reference.

The brief description of OEM is:

$$\boldsymbol{x}_{oe} = \boldsymbol{x}_a + \left(\boldsymbol{K}^T\boldsymbol{S}_e^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1}\boldsymbol{K}^T\boldsymbol{S}_e^{-1}\Delta\boldsymbol{T}_\delta \tag{9}$$

where, $\boldsymbol{S_e}$ and $\boldsymbol{S_a}$ are the measurement and *a priori* error covariance matrices. The major ambiguity in OEM is that treating errors as definite information and using these as the input parameters.

The REGB is defined as a linear combination of regression formulation, similar to the equations in references, *e.g.* Koner et al. (2015a) for night and we use same equation for day in this study.

$$s_{regb} = a_1T_{3.9} + a_2T_{11} + a_3T_{13.4} + (\sec(sza) - 1) \times (a_4T_{3.9} + a_5T_{11} + a_6T_{13.4}) + C \tag{10}$$

The coefficients are calculated separately for day and night using the matchups of the month of June 2010 and studied for whole time series.

The OSPO is defined for night, Merchant et al. (2009), as

$$s_{ospo} = a_{1n} + a_{2n}T_{3.9} + a_{3n}T_{11} + (sec(sza) - 1) \times (a_{4n} + a_{5n}T_{3.9} + a_{6n}T_{11}) \tag{11}$$

where, *sza* is the satellite zenith angle and a_{ij} are regression coefficients. Eq. (11) was also used for daytime, but with a simple parameterization for the solar component of the 3.9 μm channel, Merchant et al. (2009).

Note that the values of the OSPO SST used in this paper are the one available in the operational database, which will not be recalculated. Our aim is to use OSPO as an independent reference for the test of cloud detection algorithms and studies of performance of other methods.

OSPO was the operational SST retrieval method at the NOAA Office of Satellite Product Operations for the GOES-Imager until August 2013. After a rigorous studies and longtime validation, we have implemented an initial version of the MTLS method.

7. Comparative SST retrievals

To validate the usefulness of our proposed experimental filter, which is intended to be completely cloud free by assumption, root mean squares errors (RMSE) of SST retrievals using the different methods are shown in Fig. 8 that pass the EXF. The results for BCD are also shown in Fig. 8. Note that the 'percentage of total matches' obtained for regimes of the Bayesian and other (EXF and CEM) in Figs. 8–9 are shown in the abscissa but do not represent identical subsets. As we mentioned in our previous paper, the 'percentage of total matches', which is the x-axis of Fig. 8 and Fig. 9, calculation is based on the QI obtained from the analytical error calculation of the MTLS method, which will be used for EXF and CEM. The number of retrievals in the Bayesian domain is shown for pClr values of 0.8 through 0.98 at regular intervals.

Consistent with this, retrieval errors for all the methods in the Bayesian domain are noticeably higher than the same in the domain using the experimental filter (Fig. 8). This analysis was performed primarily to understand the cloud detection problem in terms of the quality of SST

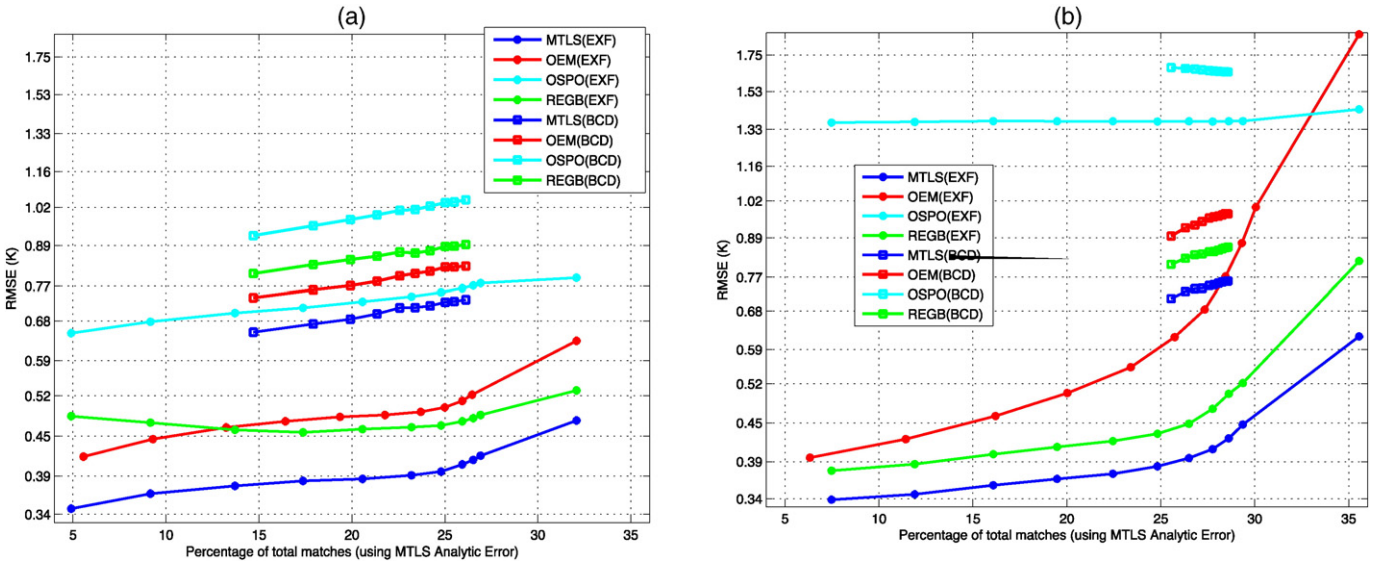

Fig. 8. Root mean square error (RMSE) in SST retrievals using four different methods for March 2012 for two different cloud detections techniques: 1) Bayesian (solid square) and 2) using experimental filter (solid *). (a) for night and (b) for day.

retrieval, and it confirms that EXF is superior to the current implementation of BCD. The EXF is not perfect either because all retrieval errors are high in the last bin. The improvement in SST quality for all methods is not the same degree under EXF as compared to BCD.

There are some interesting observations regarding Fig. 8, as follows.

- It is seen that MTLS retrieval errors for night using EXF excluding last bin (according to the QI of MTLS) is 0.43 K with data coverage of 27%, while the Bayesian subset (pClr > 0.98) is 0.65 K with a data coverage ~15%. Thus, the error reduction is ~34% (cloud related error is considered as bias) with an increase of data coverage of 80%. On the other hand, the OSPO retrieval error for night reduced from 0.92 K to 0.78 K (~15% improvement). This implies that the improved cloud detection has different impact on error reduction depending on the SST retrieval scheme.
- The improvement of REGB for night using EXF including last bin is ~32% (from 0.8 K to 0.54 K) confirms the usefulness of the experimental filter to identify cloud free measurements, which are otherwise very difficult to discriminate. EXF is developed using RTM and REGB

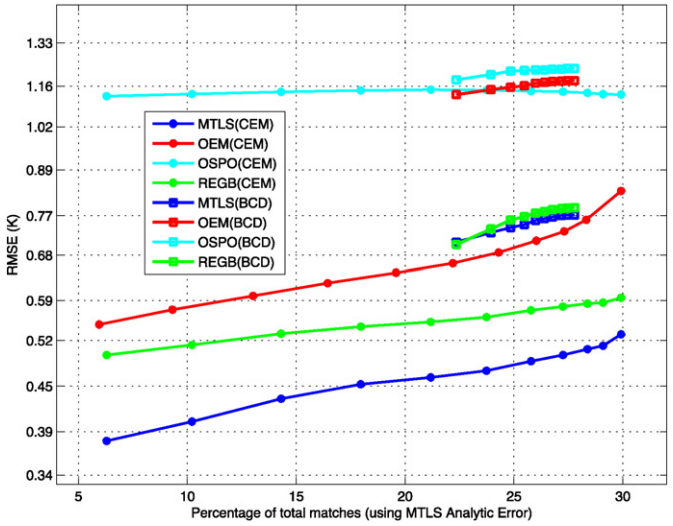

Fig. 9. Comparative RMSE under CEM and BCD algorithms and quality indexing of two different groups based on total error and Bayesian pClr values: 1) Bayesian (squares) and 2) CEM (*).

retrieval is independent of the radiative transfer model, which confirms that RTM error is insignificant for the proposed quantitative validation of cloud detection using EXF.

- The comparatively high rise of cumulative error of MTLS for both day and night in the last bin confirms that some pixels in the subset of EXF are contaminated by fractional cloud or RTM error. This may indicate that the assumed threshold of EXF of 1 K may not be optimal. The threshold was selected empirically to increase data coverage, while reducing cloud leakage and RTM error.
- The reduction of MTLS error for daytime is approximately the same as night excluding the "last bin". However, the error reduction is different for day (~8%) and night (~26%) including the last bin. This is mainly due to the fact that the daytime solar scattering effect of 3.9 μm is not fully functional in CRTM v2.1. The BT/radiance of 3.9 μm channel is affected during the daytime by the reflected solar radiance, Non-local thermodynamic equilibrium (NLTE) emission occurs at ~40-km altitude, and there is also the issue of the sea surface reflectance approximation in the model. CRTM 2.1 has already implemented such corrections by approximation of solar transmittance in terms of the satellite zenith angle and surface temperature, a fast model of NLTE and includes a BRDF surface reflection model using regression (Chen, Weng, Han, & Liu, 2012; Chen, Han, Van-Delst, & Weng, 2013). These have improved significantly the capabilities of the fast forward model during the daytime, but it still produces higher error than the nighttime. The EXF has been developed using only 3.9 μm measurement and obviously some fractional cloud leakage remains. After discarding the pixels of the last bin (extra cloud removal using MTLS), the MTLS error is much lower than the same using BCD (pClr > 0.98). Note that it is not the case of our new CEM package, where all thermal IR measurements are used.
- OE error for daytime for all cloud-free pixels (according to EXF) is much higher than in the BCD regime (pClr > 0.98) where the error reduction is observed from all other method. Even after discarding the pixels of the last bin, the OEM error contradict all other methods using EXF is higher than BCD (pClr > 0.98). This indicates that the OE algorithm is questionable, rather than the cloud algorithm. The main difficulty of the OE method is that it requires representative measurement error as an input, which is difficult to get in an operational system, Koner and Drummond (2008b). The problem is solved in residual space, thus forward model error is treated as a measurement error. To estimate forward model error, a perfect forward model is required, which is impossible for any operational problem

(even in any science problem). Furthermore, mutual consistency between the equations for Bayesian cloud and OE retrievals produce lower error compared to the EXF regime.

- As indicated above, the error covariances employed in the Bayesian cloud detection are essentially the same as those used in the OE solution, so it is not too surprising that the results are quite well matched. Also, the basic drawback of such approaches is that errors are treated as definite information and used as input parameters. Furthermore, since BCD still has significant cloud leakage (Fig. 1b), all methods (except OE) cannot obtain a good solution compared to EXF, thus the comparative results of OE (EXF and BCD) in this case are not essentially limited by the cloud detection (EXF).
- Additional information: Despite the fact that SST community is reluctant to use 3.9 μm channel for daytime regression based SST retrieval, due to the contamination of solar scattering and reflection, this study shows that three channels regression retrieval (REGB) produces good solutions for both cloud detection schemes. It can be argued that 3.9 μm channel is information rich for SST retrieval and the effects of solar scattering and reflection are sufficiently stable to permit them to be accounted for to first order in the daytime regression coefficients, *i.e.* the error that their variability introduces is less than the benefit gained from information added using the 3.9 μm channel.

Clearly, there are some intricacies in the retrieval process itself for daytime, which are beyond the scope of this paper, and a detailed study on the various issues of daytime SST retrieval using different methods will be reported in a future publication.

As the primary focus of this paper is to evaluate the performance of the CEM, from this point on, only composite results from day and night will be discussed, because the CEM uses identical tests for day and night. Fig. 9 shows comparison of retrievals using BCD and CEM techniques for a different month (July 2013) to gain confidence in the stability of our results. SST retrieval errors (RMSE) for different methods (MTLS, OEM, OSPO and REGB) are compared for different cloud algorithms. In order to interpret Fig. 9, it is necessary to (a) compare various retrievals in the BCD domain, (b) compare retrievals in the CEM domain, and (c) cross-compare various retrievals for both domains. Interpretation of (a) and (b) is straightforward and not focus for this paper, and will not discussed further, beyond noting the fact that they invariably show the superior performance of MTLS. Point (c) provides insight into the two cloud masking algorithms and is discussed below (in both vertical and horizontal dimensions). The errors for all methods (before last bin filtering using MTLS) using CEM is much lower than for BCD (pClr > 0.98), including OEM. As we observed earlier that the OEM error using EXF is higher than for BCD for daytime, we have conducted further studies offline (figure not shown) and found that the OEM error in CEM is lower than for BCD (pClr > 0.98) for daytime.

Considering the point where pClr > 0.98 in Bayesian regime (~22% cloud-free), one can see that the performance of all the retrieval methods are significantly worse compared to the approximately corresponding point of 22% cloud-free in the CEM regime. This is a further indication that even above pClr > 0.98, the "cloud-free" pixels according to the Bayesian technique are not fully free of clouds. For example, even for pClr > 0.98, the RMSE is about 0.71 K for MTLS with BCD, which reduces to 0.46 K for CEM at 22% clear-sky match-ups, which corresponds to the same fraction of pixels for pClr > 0.98. This conclusion is further supported by the fact that the regression retrievals, which are free of any influence from forward simulations, also show worse performance in the Bayesian regime. Fig. 9 further confirms the findings shown in Figs. 1 and 2 (demonstrated using EXF) that the current implementation of Bayesian cloud detection has significant false alarms and cloud leakage than our hybrid approach (both for EXF and CEM algorithms). One of the reasons is that GFS error feeds directly into error in a Bayesian prior and will affect the accuracy of the cloud mask. However, our algorithm (EXF and CEM) operates in physical deterministic inverse retrieval space where the Jacobian and residual act as the interpolation operator, and variations in TCWV and/or SST are accounted for to 1st order. Another interesting observation comes from the comparison of performances in the two different regimes in terms of the percentage of cloud-free pixels (horizontal). For example, performances of MTLS and REGB retrievals corresponding to ~30% in the newer cloud algorithm regime are better than the performances corresponding to ~22% in the Bayesian regime. This confirms that BCD rejects significant number of cloud-free measurements by falsely detecting them as cloudy pixels, which can be preserved using CEM.

The operational SST retrieval in OSPO, which is based on regression coefficients derived from radiative transfer modeled BTs, always shows worse performances in terms of RMSE for any cloud detection scheme, including CEM. It is also surprising that there is a large difference in errors between the RTM-based regression SST (OSPO) and buoy-based regression SST (REGB). This may be coming from the formulation of the regression algorithm itself and is not at all related to the choice of cloud algorithm. [Note that, OSPO uses only two channels, whereas REGB uses three channels including 3.9 μm for daytime]. We note in passing that the RMSE of REGB is much lower (~50%) than for OSPO.

From the observations in our monthly analyses (Fig. 9), we conclude that employing the proposed MTLS method in conjunction with the CEM algorithm provided significant improvement in the SST retrievals, based on the following reasons:

a) Validation of MTLS solutions against buoys is an independent assessment of its performance because it does not require any *a priori* data (just initial guess) or *a priori* and observational error covariance matrices.
b) The analytical error calculation, within the scope of the MTLS package, provides an excellent framework to develop quality indexing of the retrievals. Fig. 9 shows that RMSE of MTLS is 0.54 K at 30% of matches and reduces only to 0.4 K at 10% of matches under CEM.
c) Generally, it is perceived that data coverage is inversely proportional to the quality of retrievals. In this aspect, the CEM algorithm improves both the data coverage and the quality of retrievals. For example, the operational regression retrievals at OSPO show RMSE of ~1.2 with data coverage of ~22% (Fig. 9) whereas the proposed method (MTLS plus CEM) shows RMSE of 0.54 with a data coverage of ~30%. Moreover, the RMSE of MTLS is still ~0.47 at ~22% matchups, which 2.5 times error reduction compared to OSPO (~1.2).

8. Time series results

The above conclusions that are drawn on the basis of one month's analysis also hold for longer time series, implying that the improvements shown by the CEM methodology is temporally robust. Fig. 10 shows time series RMSE for MTLS, OSPO REGB retrievals, and fraction of cloud free (FCF) pixels for CEM and BCD, during the period of June 2010 through November 2014. We consider the REGB as a useful reference (since the SST retrieval scheme itself is independent of forward model error) to understand the performance of both cloud detections, which are based on forward modeling. The following conclusion can be made from this study (Fig. 10) as:

a) Large variations of the RMSE for both retrievals (MTLS and REGB) under BCD are observed. For example, the monthly RMSE of MTLS ranges from 0.52 K to 0.85 K. In contrast, the MTLS solution under CEM is relatively consistent. There is a tendency to low error in more recent months as compared to the older data. One possible reason is that there has been some improvement in accuracy of the *a priori* data. On the contrary, this trend is not found in the MTLS solution using BCD. It may be because the BCD uses fixed *a priori*

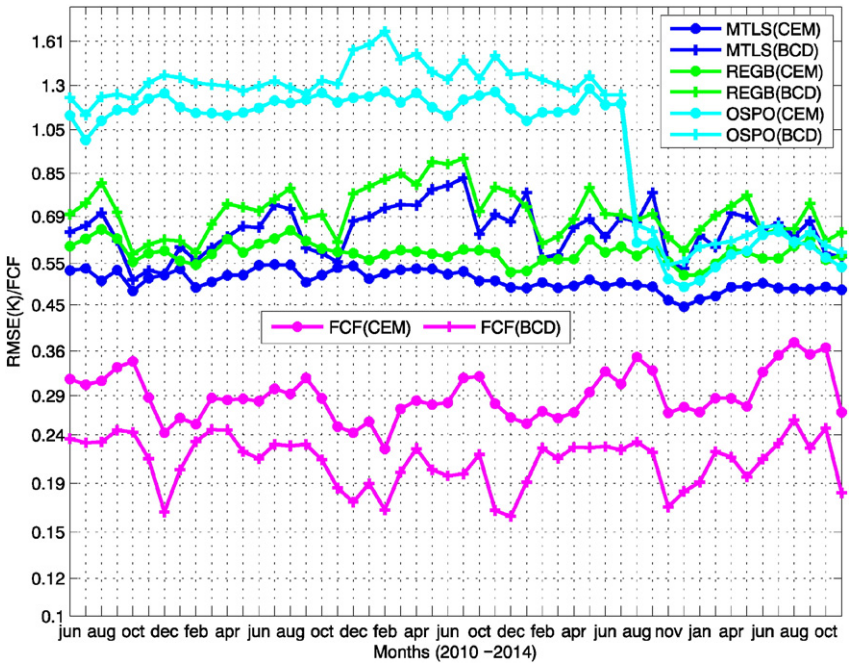

Fig. 10. Time series plots for the RMSE of MTLS (blue), OSPO (cyan) and REGB (green) under CEM (*) and under BCD (+). The fraction of cloud free (FCF) of the total matches: BCD (red +) and CEM (red *).

uncertainties and produce high and variable error in cloud classification.

b) Under the BCD, the errors of MTLS and REGB are somewhat comparable and REGB errors are occasionally lower than MTLS errors. In contrast, the MTLS retrievals have been significantly improved under the CEM scheme and there is a clear distinction between the errors of MTLS and REGB in Fig. 10. As we discussed before, CEM discards pixels, which may be contaminated by high forward model error, which gives an additional advantage to its application in conjunction with MTLS.

c) It may be regarded as an achievement to get the RMSE of MTLS for day and night composite as low as 0.43 K, with a fifty month average of 0.5 K for GOES-13 imager, which is not deemed as a very good sensor for the climate quality task and its observations are made over a highly dynamic part of the global ocean and atmosphere. The fifty months' average MTLS error reduces from 0.67 K to 0.5 K due only to change to the CEM scheme.

d) The reduction of REGB error for all months with CEM compared to BCD underscores the performance advantage of CEM over BCD in terms of cloud leakage holds for a long time series.

e) Data coverage of CEM is always higher than for BCD, and the improvement of this aspect is more than double for some months. Over the fifty months, the average data coverage improvement is 38%. Note that this is even after discarding pixels which are unsuitable for retrievals due to errors in the RTM, potential aerosol contamination and ancillary data.

f) Both cloud detection schemes show that data coverage is higher in summer months and low in winter months. This is quite likely to be due to the geographic distribution of matchup data (primarily in the Northern hemisphere), where cloud cover is greater in the winter.

g) As previously mentioned, an early version of the MTLS method was implemented in OSPO in August 2013. As a result the OSPO SST performance dramatically improves from this date onwards for both cloud detection schemes. The initial implementation included some additional processing steps to correct for radiance bias. The intent of the additional processing was to reduce the errors, but further study indicates that these algorithms may not be necessary and add ambiguity to the OSPO retrieval. The offline MTLS shown here does not make use of radiance bias correction and clearly outperforms the OSPO retrieval with the improved cloud detection of the CEM.

h) Most interesting results are observed in OSPO retrieval error after July 2013. Since then the OSPO retrieval error for both clouds are comparable where other two methods (MTLS and REGB) results show a distinct different gap between the implemented cloud detection scheme of CEM and BCD. Even though we have already extensively demonstrated that CEM provides better cloud detection, it does not work well with the OSPO retrieval. As we discussed in our earlier paper (Koner et al., 2015a), radiance bias correction (RBC) may have several ambiguities, however a form of RBC is currently implemented in OSPO, as it is quite popular in operational realm. RBC adjusts the measurement based on an assumption of cloud-free conditions, as currently determined by BCD. In such a situation, unscreened cloud contamination is included in the calculation of the radiance bias adjustment and overall validation statistics are improved due to a good quality SST obtained from cloudy pixels. As a result, statistical OSPO error is lower than the offline MTLS error under BCD. On the other hand, radiance bias adjustment is applied to whole dataset, and cloud free measurements are altered by this bias adjustment. This increases the SST error for OSPO under CEM, since the cloud contamination has been removed but the incorrect radiance bias still affects the OSPO retrieval. A detailed study will be reported in a future publication.

As we observed in Fig. 10 that MTLS error reduced significantly under CEM because CEM was developed in part to discard pixels which cannot be solved uniquely due to RTM error. To investigate the mutual consistency between CEM and physical retrieval method, the performance of MTLS is compared with two other RTM-based methods (namely, LS and OEM) for both (CEM and BCD) cloud detection as shown in Fig. 11. As we discussed in our last publication, Koner et al. (2015a), the condition number of the Jacobian for two parameters SST retrieval from GOES-13 imager measurement is typically around 5. Thus, the error propagated from measurement space to state space,

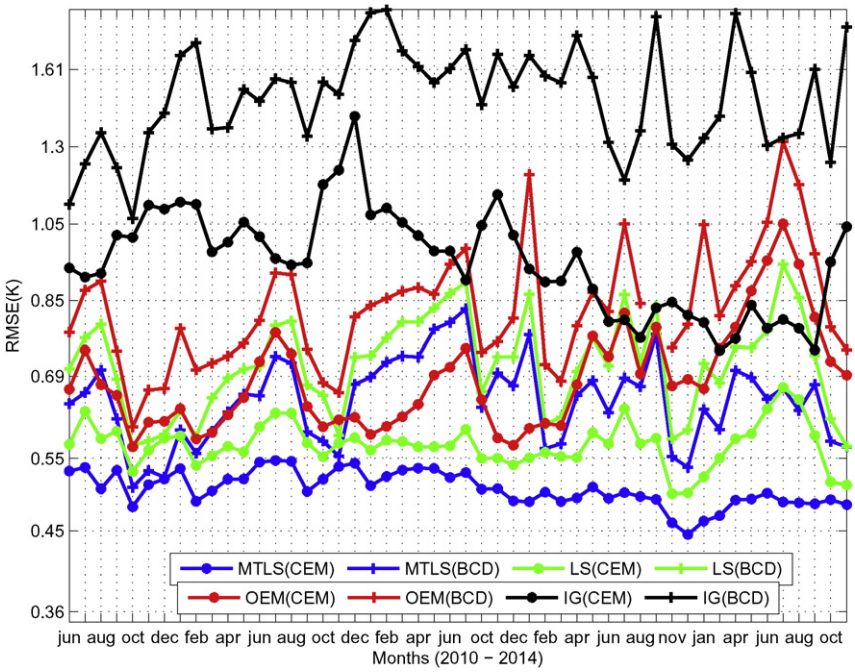

Fig. 11. Time series plot for RMSE values of MTLS, OEM, LS and IG under comparative CEM and BCD.

which is multiplied by the condition number of the matrix (Koner, Battaglia, & Simmer, 2010; Ralston & Rabinowitz, 1978), is low. Although technically ill-conditioned, it is quite close to being a well-conditioned problem and thus we add the least squares (LS) solution as a reference. The variation and magnitude of MTLS errors are low as compared to the other two methods for all the months under CEM. The reduction of LS error under CEM is more or less similar to MTLS. The highest variation is observed in the OEM solutions, and the OEM error is higher than *a priori* error for some months (April–September 2014) under CEM. One possible reason for the bad performance during this period is that the error of the retrieval system is reduced in reality, as evidence of the MTLS error is relative low (<0.5 K) for this period, and input OEM error is not in agreement with real errors. This is one of drawbacks of the OEM implementation, where errors are input parameters and these are difficult to specify in a dynamic operational environment. The overall error can be significantly high even in good cloud detection regime if the algorithm error is high. On the other hand, MTLS employs data driven regularization and inherently produces the best possible results according to total errors that exist in the system for any individual realization. There is no evidence of a similarly dramatic degradation of the REGB retrievals for this period (see Fig. 10), thus the reduction of instrument noise is unlikely to be the cause. It may be concluded that reduced *a priori* error is one possible cause for reduced MTLS and increased OEM errors for this period.

Another interesting result observed from this study is that the OEM errors are always higher than those for LS error for all months and both cloud detection schemes, because this is a low conditioned inverse problem. This result should alert the satellite inverse community that implementation of OEM without a proper understanding of the mathematical behavior of the particular inverse problem may cause degradation of information relative to simple least squares. Treating *a priori* and measurement errors as information for input is a potential risk for any science problem. Using LS solution as a reference, MTLS always improves its retrievals for all months and both cloud detection schemes from that reference. Additionally, the improvement of the LS solution under CEM is further strong evidence that our CEM is a better cloud detection scheme than the OSPO implementation of BCD.

9. Conclusions

We have used a simple experimental filter based on matchup data, radiative transfer and a single-channel SST retrieval to assess the degree of cloud contamination in satellite-observed brightness temperatures. This has demonstrated that there is significant cloud leakage in the current OSPO Bayesian Cloud Detection, and that there are many observations that are deemed cloudy for which good SSTs may still be retrieved. We have developed a cloud and error mask for the purposes of SST retrieval, which achieves the highly desirable combination of improved coverage and accuracy by utilizing a pragmatic combination of tests based on physics and radiative transfer calculations. The methodology is shown to be well-suited to an operational environment. An interesting aspect of this algorithm is that it not only improves the quality of physical SST retrievals but it also improves the quality of two different regression based SST retrievals. This confirms that this algorithm is not compromised by potential tuning of the FFRTM, although two tests utilize FFRTM.

The most unique aspects of combined MTLS and CEM package are the use of the double differences filter, extra cloud/error filter using the “last bin” assignment afforded by the MTLS solution and analytical total error calculation. A crucial aspect of the accuracy and stability of the final result remains the flexibility of the MTLS retrieval, which has the ability to cope with varying quality of input *via* its data-driven regularization capability.

Acknowledgments

This study was supported by the NESDIS Product Systems Development and Implementation program through NOAA grant NA09ES4400006 (Cooperative Institute for Climate and Satellites—CICS) at the University of Maryland/ESSIC. We thank all the reviewers for their suggestions to improve the readability of the manuscript. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or US Government position, policy, or decision.

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, A., Moeller, C. C., & Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research*, *103* (D24), 32141–32157.

Barnes, B. B., & Hu, C. (2013). A hybrid cloud detection algorithm to improve MODIS sea surface temperature data quality and coverage over the eastern Gulf of Mexico. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 3273–3285. http://dx.doi.org/10.1109/TGRS.2012.2223217.

Chen, Y., Han, Y., Van-Delst, P., & Weng, F. (2013). Assessment of shortwave infrared sea surface reflection and nonlocal thermodynamic equilibrium effects in the community radiative transfer model using IASI data. *Journal of Atmospheric and Oceanic Technology*, *30*, 2152–2160.

Chen, Y., Weng, F., Han, Y., & Liu, Q. (2012). Planck-weighted transmittance and correction of solar reflection for broadband infrared satellite channels. *Journal of Atmospheric and Oceanic Technology*, *29*, 382–396.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., & Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*, 353–369.

Dybbroe, A., Karlsson, K., & Thoss, A. (2005). NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling, part I: Algorithm description. *Journal of Applied Meteorology and Climatology*, *44*, 39–53.

Heidinger, A. K., Evan, A. T., Foster, M. J., & Walther, A. (2012). A Naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *Journal of Applied Meteorology and Climatology*, *51*, 1129–1144.

Hu, C., Muller-Karger, F. E., Murch, B., Myhre, D., Taylor, J., Luerssen, R., ... Hendee, J. (2009). Building an automated integrated observing system to detect sea surface temperature anomaly events in the Florida keys. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 1607–1620.

Hutchison, K. D., Hardy, K. R., & Gao, B. C. (1995). Improved detection of optically thin cirrus clouds in nighttime multispectral meteorological satellite imagery using total integrated water vapor information. *Journal of Applied Meteorology*, *34*, 1161–1168.

Ishida, H., & Nakajima, T. Y. (2009). Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager. *Journal of Geophysical Research*, *114*, D07206. http://dx.doi.org/10.1029/2008JD010710.

Jedlovec, G. (2009). Automated detection of clouds in satellite imagery. Book chapter (ISBN: 978-953-307-005-6): Advances in geoscience and remote sensing.

Jedlovec, G. J., Haines, S. L., & LaFontaine, F. J. (2008). Spatial and temporal varying thresholds for cloud detection in GOES imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *46*, 1705–1717.

Karlsson, K. G., Johansson, E., & Devasthale, A. (2015). Advancing the uncertainty characterisation of cloud masking in passive satellite imagery: Probabilistic formulations for NOAA AVHRR data. *Remote Sensing of Environment*, *158*, 126–139.

Koner, P. K., & Drummond, J. R. (2008a). A comparison of regularization techniques for atmospheric trace gases retrievals. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *109*, 514–526.

Koner, P. K., & Drummond, J. R. (2008b). Atmospheric trace gases profile retrievals using nonlinear regularized total least squares method. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *119*, 2045–2059.

Koner, P. K., Battaglia, A., & Simmer, C. (2010). A rain rate retrieval algorithm for attenuating radar measurement. *Journal of Applied Meteorology and Climatology*, *49*, 381–393.

Koner, P. K., Harris, A., & Maturi, E. (2015a). A physical deterministic inverse method for operational satellite remote sensing: An application for SST retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, *53*, 5872–5888.

Koner, P. K., & Harris, A. R. (2015b). A deterministic inversion technique for sea surface temperature retrieval from MODIS radiances. *Proc. SPIE 9459, Ocean Sensing and Monitoring VII, 94590Y (19 May 2015)*. http://dx.doi.org/10.1117/12.2179868.

Kotarba, A. Z. (2009). A comparison of MODIS-derived cloud amount with visual surface observations. *Atmospheric Research*, *92*, 522–530.

Kriebel, K. T., Gesell, G., Kästner, M., & Mannstein, H. (2003). The cloud analysis tool APOLLO: Improvements and validations. *International Journal of Remote Sensing*, *24*, 2389–2408.

Merchant, C. J., Harris, A. R., Maturi, E., & MacCallum, S. (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society*, *131*, 2735–2755.

Merchant, C. J., Harris, A. R., Maturi, E., Embury, O., Macallum, S. N., Mitaz, J., & Old, C. P. (2009). Sea surface temperature estimation from the Geostationary Operational Environmental Satellite-12 (GOES-12). *Journal of Oceanic and Atmospheric Technology*, *26*, 570–581.

Murtagh, F., Barreto, D., & Marcello, J. (2003). Decision boundaries using Bayes factors: The case of cloud masks. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 2952–2958.

Petrenko, B., Ignatov, A., Kihai,Y. & Heidinger, A. (2010). Clear-sky mask for the advanced clear-sky processor for oceans. Journal of Atmospheric and Oceanic Technology, 27, 1609–1623.

Ralston, A., & Rabinowitz, P. (1978). *A first course in numerical analysis.* New York: McGraw-Hill.

Saunders, R. W., & Kriebel, K. T. (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, *9*, 123–150.

Stowe, L. L., Davis, P. A., & McClain, E. P. (1999). Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer. *Journal of Atmospheric and Oceanic Technology*, *16*, 656–681.

Stowe, L. L., McClain, E. P., Carey, R., Pellegrino, P., Gutman, G., Davis, P., ... Hart, S. (1991). Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Advances in Space Research*, *11*, 51–54.

Uddstrom, M. J., Gray, W. R., Murphy, R., Niles, A. O., & Murray, T. (1999). A Bayesian cloud mask for sea surface temperature retrieval. *Journal of Atmospheric and Oceanic Technology*, *16*, 117–132.

Van-Vuuren, D., den-Elzen, M., Lucas, P., Eickhout, B., Strengers, B., van Ruijven, B., ... van Houdtet, R. (2007). Stabilizing greenhouse gas concentrations at low levels: An assessment of reduction strategies and costs. *Climatic Change*, *81*, 119–159.

Walker, J. R., Mackenzie, W. M., Mecikalski, J. R., & Jewett, C. P. (2012). An enhanced geostationary satellite-based convective initiation algorithm for 0–2-h nowcasting with object tracking. *Journal of Applied Meteorology and Climatology*, *51*, 1931–1949.

Xu, F., & Ignatov, A. (2014). In situ SST Quality Monitor (iQuam). *Journal of Atmospheric and Oceanic Technology*, *31*, 164–180.

Zhenglong, L., Li, J., Menzel, W. P., Schmit, T. J., & Ackerman, S. A. (2007). Comparison between current and future environmental satellite imager on cloud classification using MODIS. *Remote Sensing of Environment*, *108*, 311–326.