

Retrieval of Sea Surface Temperature from Space, Based on Modeling of Infrared Radiative Transfer: Capabilities and Limitations

CHRISTOPHER J. MERCHANT

Atmospheric and Environmental Science, The University of Edinburgh, Edinburgh, United Kingdom

PIERRE LE BORGNE

Centre de Meteorologie Spatiale, Meteo-France, Lannion, France

(Manuscript received 3 February 2003, in final form 22 April 2004)

ABSTRACT

The retrieval (estimation) of sea surface temperatures (SSTs) from space-based infrared observations is increasingly performed using retrieval coefficients derived from radiative transfer simulations of top-of-atmosphere brightness temperatures (BTs). Typically, an estimate of SST is formed from a weighted combination of BTs at a few wavelengths, plus an offset. This paper addresses two questions about the radiative transfer modeling approach to deriving these weighting and offset coefficients. How precisely specified do the coefficients need to be in order to obtain the required SST accuracy (e.g., scatter <0.3 K in week-average SST, bias <0.1 K)? And how precisely is it actually possible to specify them using current forward models? The conclusions are that weighting coefficients can be obtained with adequate precision, while the offset coefficient will often require an empirical adjustment of the order of a few tenths of a kelvin against validation data. Thus, a rational approach to defining retrieval coefficients is one of radiative transfer modeling followed by offset adjustment. The need for this approach is illustrated from experience in defining SST retrieval schemes for operational meteorological satellites. A strategy is described for obtaining the required offset adjustment, and the paper highlights some of the subtler aspects involved with reference to the example of SST retrievals from the imager on the geostationary satellite *GOES-8*.

1. Introduction

The retrieval of sea surface temperature (SST) from space-based multichannel observations of infrared radiances (Deschamps and Phulpin 1980) has been performed routinely since 1981 and is now an important element of the global observing system for weather prediction and climate monitoring. Early retrieval schemes for the Advanced Very High Resolution Radiometer (AVHRR) were empirically determined by regression of observations matched to in situ SSTs (McClain et al. 1985). Llewellyn-Jones et al. (1984) applied a contrasting approach based on results of radiative transfer (RT) simulations to AVHRR, and this strategy was adopted to define an SST retrieval scheme prior to launch (Zavody et al. 1995) for the Along-Track Scanning Radiometer (ATSR). More recent work on ATSR (Merchant et al. 1999) has demonstrated that subsequent developments of the spectroscopy of water vapor have con-

tributed to improvements in RT-based coefficients, such that ATSR SSTs were validated in tropical regions to have a bias of ~ 0.1 K and standard deviation of ~ 0.25 K (Merchant and Harris 1999). The retrieval scheme for the Advanced ATSR (AATSR) flying on the European Space Agency's platform, ENVISAT, is based on RT modeling. Radiative transfer is also the preferred approach to defining retrieval coefficients for operational meteorological satellites in the work of the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) (Francois et al. 2002).

The purpose of this article is to set out what can (and cannot) be reasonably expected of RT-based coefficients in terms of SST accuracy. Our conclusions will be as follows. First, RT-based coefficients are able to retrieve SSTs with a precision that appears nearly optimum in validation studies—that is, the standard deviation of the SSTs is good. Second, SSTs from RT-based coefficients are likely to be biased by up to several tenths of a kelvin; thus, in contexts where this is not acceptable, an additional step of empirical bias correction is necessary. (In formulating these conclusions, we take for granted that the important cloud-screening step in SST determination has been adequately achieved, since retrievals can only be made for ocean under clear skies.)

Corresponding author address: Dr. Christopher J. Merchant, Institute for Atmospheric and Environmental Sciences, The University of Edinburgh, Room 8207, JCMB, King's Bldgs., Mayfield Rd., EH9 3JZ Edinburgh, United Kingdom.
E-mail: chris.merchant@ed.ac.uk

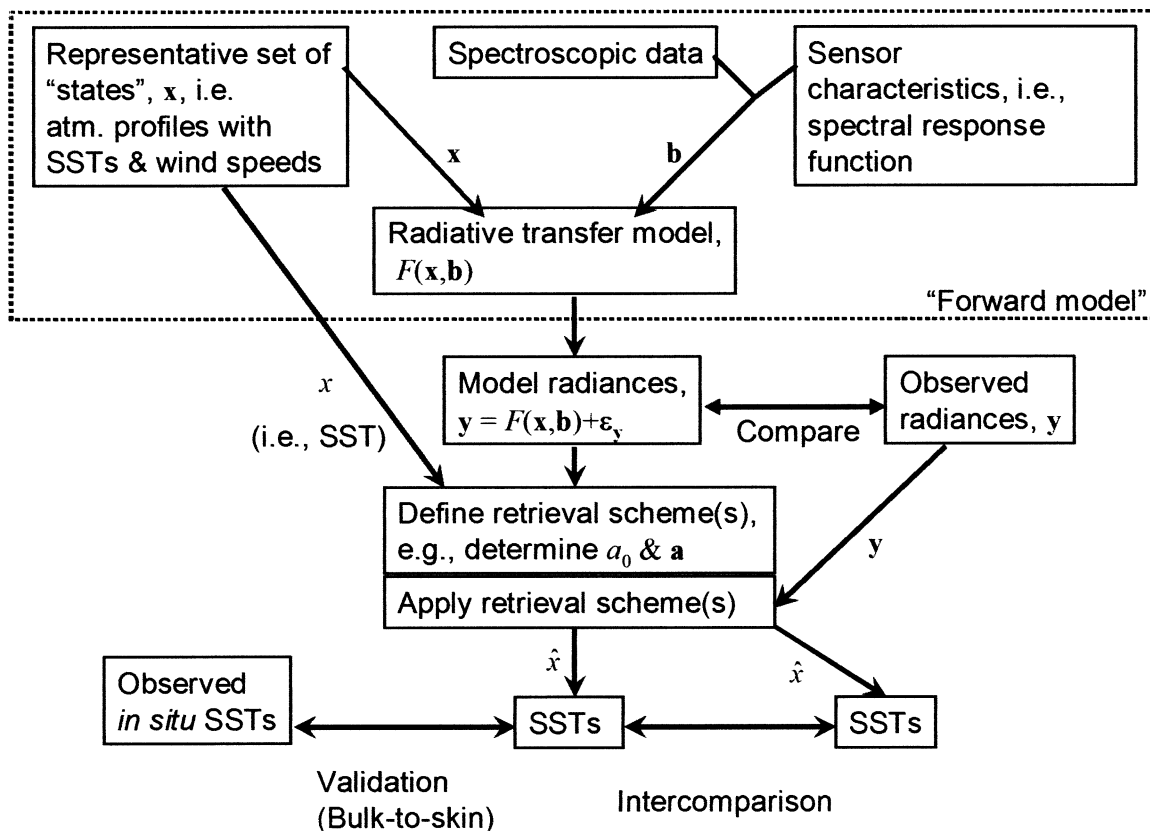


FIG. 1. Schematic of the process of defining and validating coefficients for sea surface temperature retrieval using radiative transfer modeling. (After a slide presented by A. R. Birks.)

These conclusions are justified on theoretical grounds in section 2 and are borne out by the experience of the OSI-SAF reported in section 3. A strategy for performing the bias-correction (or “offset adjustment”) step is discussed in section 4, taking into account issues of skin-bulk difference and diurnal SST variability (Murray et al. 2000; Gentemann et al. 2003). This is illustrated by a particular example in section 5.

2. Theory of physically based SST retrieval

a. Defining retrieval coefficients

The process of defining coefficients for SST retrieval using RT-modeled radiances is illustrated schematically in Fig. 1. To define the coefficients, we need to simulate radiances observed by the infrared sensor as it views the oceans under realistic clear-sky conditions. We will define the complete setup for performing this simulation as the “forward model,” which has a number of elements.

A central component of the forward model is, of course, software to simulate the radiances—the radiative transfer model (RTM). As indicated in Fig. 1, the inputs to the RTM constitute the other two components. Radiances for a range of SSTs and corresponding atmo-

spheric conditions are required, in order to capture the distribution of relationships between satellite observations and the SST to be estimated. Thus, a component of the forward model is a set of atmospheric profiles and associated surface variables (SST and wind speed) that is representative of what the sensor will be required to observe. One of the authors, Merchant (Merchant et al. 1999), has approached this using fields of variables from numerical weather prediction (NWP), while the other, Le Borgne (Francois et al. 2002), has preferred to base the set on radiosoundings. Following usage in the retrieval papers of Rodgers (1990, 1976), we represent these input parameters to the forward model as a “state vector,” x . One of the elements of x is the SST, to which we refer simply as x .

As well as the state vectors, the RTM must also be given the spectral response function that characterizes the sensor. Moreover, RT calculations are based on spectroscopic parameters for the relevant atmospheric species (gases and aerosols). These spectroscopic parameters embody measurements of the many weak absorption features in the relevant regions of the infrared spectrum (typically, 3.5–4.1 and 10.5–12.5 μm). Depending on the RTM, these spectroscopic parameters may be explicit, separated from the software, or implicit, em-

bedded in it. In either case, from a formal viewpoint, the spectroscopic data and sensor characterization used in the RTM are the third element of the forward model: the “model parameters,” \mathbf{b} .

The forward model generates simulated radiances, \mathbf{y} , where $\mathbf{y} = F(\mathbf{x}, \mathbf{b}) + \varepsilon_F$; F here represents the function of the RTM, and ε_F the radiative transfer model error—that is, the departure of the simulation from what would really be observed by a sensor observing the situation described by \mathbf{x} and \mathbf{b} . This is not the full forward-model error, because there may be systematic differences between the state vectors used and reality, and there are errors in the model parameters. The full forward-model error is (approximately)

$$\varepsilon_y = \varepsilon_F + \frac{\partial F}{\partial \mathbf{x}} \varepsilon_x + \frac{\partial F}{\partial \mathbf{b}} \varepsilon_b, \quad (1)$$

where the subscripts of ε define the parameter in error. As we illustrate later, any or all of the terms in this expression for the full forward-model error can be significant in the context of RT-based retrieval of SST.

The radiances obtained from the forward model are then used to define one or more retrieval schemes for estimating SST from radiances observed by the sensor. The retrieval schemes can be tested by looking for consistency between SST estimates of different schemes, and by validation exercises in which satellite SSTs are compared with measurements matched in time and space and made by drifting buoys, moorings, or shipborne radiometers. In validation, it is sometimes necessary to account for the possibility of difference between the temperature of the ocean skin (to which radiometers are sensitive) and the bulk water a few centimeters to a few meters below the surface. Fuller accounts of the RT modeling and validation processes are given elsewhere (Zavody et al. 1995; Merchant et al. 1999; Merchant and Harris 1999; Llewellyn-Jones et al. 1984; Johnson and Weinreb 1996).

Linear and near-linear retrieval schemes are nearly optimal for SST estimation if the infrared radiances are expressed as brightness temperatures (BTs). In the case of linear retrieval, the SST estimate, \hat{x} , is formed from a weighted combination of BTs. In this paper, we use matrix–vector notation as follows (consistent with Merchant et al. 1999): all vectors will be column vectors, and appear as lowercase; all matrices appear in uppercase; and the transpose operator is superscript T. Thus,

$$\hat{x} = a_0 + \mathbf{a}^T \mathbf{y}, \quad (2)$$

where a_0 is the offset coefficient, and $\mathbf{a}^T = [a_1, \dots, a_n]$ is a vector of n weighting coefficients that each multiply one of the n BTs in the observation vector \mathbf{y} . These infrared observations are at different wavelengths and/or view angles, and $\mathbf{a}^T \mathbf{y}$ is the inner product of the weighting and observation vector, equivalent to the summation $\sum_i^n a_i y_i$.

The coefficients are found by minimizing the mean square difference between the “true” SST input to the

RTM and the “retrieved” SST given by Eq. (2), for the population of atmospheric and surface states and associated RTM BTs. The least squares formulas for the offset and weights are

$$\mathbf{a} = \mathbf{S}_{yy}^{-1} \mathbf{s}_{xy}, \quad a_0 = \bar{x} - \mathbf{a}^T \bar{\mathbf{y}}, \quad (3)$$

where a bar above a quantity indicates the mean value, x is the “true” SST associated with a given set of simulated BTs, \mathbf{S}_{yy} is the covariance matrix of observations, and \mathbf{s}_{xy} is the covariance vector of SST and observations. [The standard least squares equations of Eq. (3) can be modified to account for noise (Zavody et al. 1995) or to introduce additional constraints (Merchant et al. 1999); similarly, slightly nonlinear formulations for SST retrieval are in use (Walton et al. 1998), where the coefficients are slowly varying with respect to a first-guess or estimated SST. None of these variations on Eq. (3) significantly changes the development that follows.]

The accuracy and precision required for retrieved SSTs depends, of course, on the application. Most weather and climate users’ needs are met by a precision (standard deviation) of order 0.3 K on a weekly time scale. [During a period of order 1 week, the number of useful observations ranges from a few (for a polar-orbiting satellite) to many (geostationary), and the precision this implies for each independent SST estimate is $0.3 \times (\text{number of estimates})^{1/2}$ K.] In the contexts of climate-trend detection, of the merging step in the creation of multisensor products, and of the assimilation of SSTs in ocean models, accuracy (mean bias) of order 0.1 K is a useful target. This prompts two questions: How accurately do the retrieval coefficients need to be specified in order to meet this accuracy and precision? And how accurately is it actually possible to specify them using the approach just described?

b. Requirements for coefficients

Consider first the weighting coefficients. Each of these is multiplied by a BT, $y \sim 300$ K, and a misspecification of the i th weight of order $\varepsilon(a_i)$ causes an error in the SST estimate of $\sim y \varepsilon(a_i)$. If each of the n weights is subject to a random zero-mean mis-specification with a standard deviation σ , the cumulative effect on the SST estimate is an error $\sim n^{1/2} y \sigma$. SST retrievals are commonly made using two or three channels (four or six for ATSR-series sensors). To have the contribution to errors from this source be $\ll 0.3$ K implies (using $n = 3$) that $\sigma \ll 5 \times 10^{-4}$. Thus, for BTs and SSTs expressed in kelvins, weights should be given to five decimal places with uncertainty permissible in the last digit. (As an aside, if BTs and SST estimates are expressed in degrees Celsius, the weighting coefficients need be expressed to only four decimal places to achieve the same SST precision, since the relevant temperatures expressed this way are an order of magnitude smaller than corresponding temperatures in kelvins. Less numerical precision in computation is required for degree Celsius calcula-

TABLE 1. Example of differences in simulated BTs for SST. Simulated BTs were obtained for SEVIRI spectral response functions (see <http://www.eumetsat.de>) for the channels and view angles indicated, using 58 NWP profiles chosen to encompass the full global range of variability of maritime atmospheres and SST. See main text for details of the models compared.

	3.9 μm , 0°	3.9 μm , 60°	11 μm , 0°	11 μm , 60°
MODTRAN 4 minus RFM	0.14 K	-0.03 K	-0.03 K	-0.14 K
RTTOV-7 minus RFM	-0.09 K	-0.07 K	-0.05 K	0.04 K
RAL model minus RFM	0.02 K	0.10 K	0.00 K	0.04 K

tions. A consequence is that additional retrieval errors will be introduced if one attempts to re-express a retrieval equation designed for degree Celsius with four decimal places as a retrieval equation for temperatures in kelvins. However, the conversion between temperature scales can proceed in the other direction without any such effect.)

Turning to the offset coefficient, any error in the offset, a_0 , translates directly as an error (global bias) in estimated SST. So, the offset needs to be specified to within a tenth of a kelvin to meet the accuracy target of 0.1 K.

We wish to compare these requirements with what is possible for RTM-based coefficients. To do this, the accuracy and precision of forward modeling of channel-integrated radiances for typical sensors need to be evaluated. This is done in the following section, 2c, and we return to the question of how closely coefficients can be determined in section 2d.

c. Accuracy and precision of the forward model

Ideally, the accuracy and precision of forward modeling would be assessed by comparing simulated and observed top-of-atmosphere radiances for known states. In practice, it is difficult to know the state \mathbf{x} with sufficient precision to achieve a meaningful assessment. For our purpose in this article, it is sufficient to evaluate the likely error levels in simulated BTs by studying the sensitivity of simulations to reasonable uncertainties in \mathbf{b} and \mathbf{x} and by intercomparison of RTMs commonly used for RT-based SST.

We start by decomposing the forward model error into two elements: a bias term and a scatter term:

$$\varepsilon_y = \bar{\varepsilon} + \varepsilon'. \quad (4)$$

The bias term $\bar{\varepsilon}$, is the forward-model error averaged over the full range of states. It is the mean effect of errors in the model parameters, of errors in the state, and/or of numerical approximations in the RTM. The scatter term is zero mean, $\bar{\varepsilon}' = 0$, but is not random. It represents the differential effect that these forward-model errors have on simulated BTs for different states.

We look first at the likely error from using an approximate RTM, as is often done. The authors have performed a few unpublished comparisons of RTMs for different sensors and RTM combinations. The results imply that mean channel-integrated BTs for spectral response functions typical of SST sensors are typically

within 0.1 K of each other. To illustrate this for the reader, we show an example set of results in Table 1. Here, we performed simulations using 58 NWP profiles for four different RTMs. The profiles are carefully selected to give a full range of atmospheric variability in a small sample (Chevallier 2002). The RTMs were RTTOV-7 (Saunders et al. 1999), Moderate Resolution Transmittance Code (MODTRAN; version 4), the RTM developed for SST at the Rutherford Appleton Laboratory (the RAL model) (Zavody et al. 1995; Merchant et al. 1999), and the ‘‘Reference Forward Model’’ (RFM; described online at <http://www.atm.ox.uk/RFM>), which is functionally equivalent to General Line-by-Line Atmospheric Transmittance and Radiation Model (GENLN; Edwards 1988). RTTOV-7 is a fast channel-integrated transmittance model. MODTRAN4 is a 1 cm^{-1} band model. The RAL model runs at 0.04 cm^{-1} resolution and can be viewed as a ‘‘very narrowband’’ model. The RFM is true line-by-line code, and for the comparisons was run for the same wavenumbers as the RAL model. One of the authors (Merchant) has updated the RAL model to use the high-resolution molecular spectroscopic (HITRAN2000) database (Rothman et al. 2003), which was also used for the RFM. For RTTOV-7 and MODTRAN4 the spectroscopy is less recent; the BT differences for those models therefore include effects of differences in \mathbf{b} as well as effects of the band approximations. The simulations are performed for the instrument response functions of the Spinning Enhanced Visible and Infrared Imager (SEVIRI; <http://www.eumetsat.de>). The results are expressed in Table 1 as differences of simulated BTs of the three approximate models from the RFM BTs.

We may characterize the magnitude of systematic radiative transfer model error from an approximate RTM (e.g., a band model) for any given channel as 0.1 K or less (e.g., see the comparisons of the RAL model and RFM in Table 1). One might argue that this implies use of a full line-by-line RTM in SST simulations to make this component of error smaller. Estimation of the absolute accuracy in radiative transfer that could be achieved in this way is outside the scope of this paper, but clearly use of a line-by-line model rather than a band model will reduce the first contribution to forward-model error in Eq. (1). However, we do need also to consider the magnitude of error arising from reasonable errors in the specification of the state, \mathbf{x} , and in the model parameters, \mathbf{b} [i.e., the size of the second and third terms

TABLE 2. Effect of perturbation to forward-model parameters on brightness temperatures (ΔBT) in typical SST sensor channels,^a expressed as mean, $\bar{\epsilon}$, and standard deviation, ϵ' , and the correlation coefficient between ΔBT and unperturbed BT, r .

Perturbation	3.7- μm ΔBT (K)			11- μm ΔBT (K)			ΔSST retrieval ^g	
	$\bar{\epsilon}$	ϵ'	r	$\bar{\epsilon}$	ϵ'	r	Bias (K)	ΔSD (%)
Increase sea surface emissivity by 0.001 ^b	0.017	0.001	0.1	0.034	0.013	-0.7	-0.05	0.4
Change trace gas profiles from 1999 to 1991 values ^c	0.024	0.002	0.0	0.005	0.001	0.0	-0.03	0.03
Change water vapor continuum parameterization ^d	0.017	0.014	0.74	0.017	0.030	0.6	0.01	0.08
Reduce upper-tropospheric humidity by 15% ^e	0.019	0.015	0.4	0.024	0.035	0.2	-0.04	6.8
Modify instrument spectral response ^f	0.099	0.027	0.6	0.019	0.007	0.9	-0.12	4.4

^a These simulations were performed using spectral response functions for the ATSR-2 (available online at <http://www.atrs.rl.ac.uk/>). These are typical of meteorological SST window channels, such as AVHRR and GOES Imager series sensors.

^b This is the uncertainty in rough sea emissivity in the 10.5–12.5- μm window, if one uses emissivity models that account for varying wave-shadowing effects and for surface emission–surface reflection (Wu and Smith 1997; Watts et al. 1996).

^c In specifying the state, the volume mixing ratio profiles of trace gases needs to be defined (or, in the case of approximate RTMs, may be implicit in the model). Here, to indicate the scale of sensitivity, the trace gas profiles of CO_2 , CH_4 , N_2O , CO , and CFCs were changed from estimated global-mean profiles for 1999 to 1991. These long-term changes can be modeled and removed using time-dependent coefficients, but regional and seasonal variations that are often of similar or greater magnitude remain and are not usually accounted for with prior information.

^d The difference is between use of two parameterizations of Clough and colleagues (S. A. Clough 2002, personal communication), viz. “CKD 2.2.2” and “MT_CKD” (see <http://www.rtweb.aer.com/>).

^e Some radiosoundings may be subject to a dry bias of order 5% (Miloshevich et al. 2001). The profiles used here from the European Centre for Medium-Range Weather Forecasts may overestimate upper-tropospheric humidity (UTH) in subsiding regions and may be deficient in UTH in regions of strong convection (Soden and Bretherton 1994). Here, then we illustrate the sensitivity to a systematic error in the water vapor profile by reducing the UTH by 15% between 500 and 100 hPa. Other plausible scenarios for humidity biases could be assessed similarly. Of course, were the true distribution of UTH more precisely known, this uncertainty would reduce.

^f For illustrative purposes, we modify the profile within the specification for measured spectral response function of GOES Imager infrared channels (M. Weinreb 2003, personal communication). The central wavenumber was shifted by 0.8 and 0.2 cm^{-1} for the 3.7- and 11- μm channels, respectively, and random changes of 1% were applied to the normalized spectral response function. Note that we change the radiance-to-temperature conversion in the forward model appropriately to the “new” spectral response, so that these errors are not arising from that conversion step.

^g Quoted for errors introduced into coefficients for three-channel linear retrievals (using channels at 3.7, 11, and 12 μm), in simulation using ATSR-2 channels with a noise equivalent differential temperature of 0.07 K.

in Eq. (1)]. These are further explored below using sensitivity analysis.

Examples of sensitivity of forward-model BTs are given in Table 2. In each case, we have perturbed (a) model or state parameter(s) to a degree representative of their contemporary uncertainty or variability, and have quantified the resulting $\bar{\epsilon}$ and ϵ' (relative to the unperturbed forward model). Details of the perturbations and their bases are given in the footnotes to the table; the RAL RTM has been used to create this table. Realistic systematic uncertainty in specification of model parameters (such as emissivity) and state parameters (such as humidity) cause forward-model bias of order a few hundredths of a kelvin. A realistic upper estimate of forward-model bias related to the sensor characterization is just under a tenth of a kelvin for the 3.7- μm channel and again is a few hundredths of a kelvin for the 11- μm channel. [This is an upper estimate because sensor characteristics should be more closely specified for more modern instruments than the Geostationary Operational Environmental Satellite (GOES) sensor used as an example.] The resulting SST biases are seen to be of comparable magnitude—a few hundredths of a kelvin in each case. There are, of course, many thousands of model parameters (e.g., centers and strengths

of individual absorption lines), most of which we may assume have much smaller impacts on BTs than the perturbations illustrated. Nonetheless, we conclude that the combined error from contributions at the >0.01 K level from several influential parameters, including those illustrated in Table 1, is likely to be ~ 0.1 K for a typical forward model.

The scatter terms can be smaller by factors up to ~ 20 . This occurs when the perturbed parameters have a differential effect on simulations for different profiles that is much smaller than their common effect. Scatter terms are not smaller where the change involves the absorption of water vapor in the atmosphere; for these cases bias and scatter elements of the error are comparable. This is because absolute humidity is highly variable between different atmospheric profiles. The overall magnitude of ϵ' relative to $\bar{\epsilon}$ therefore depends on how accurately water vapor amount and water vapor absorption are represented. The change in split-window BT between the water vapor continuum parameterizations CKD2.2.2 (Han et al. 1997) and MT_CKD (S. A. Clough 2002, personal communication) is negligible for SST retrieval purposes. This perhaps indicates convergence in the spectroscopy of the continuum in this spectral region.

The increase in the simulated standard deviation of the retrieved SST that results tends to be modest.

We conclude that, overall, a forward model for SST should be capable of simulating the differences in BTs arising from different states to an accuracy (bias) of order 0.1 K. This matches the level of agreement between the RTMs that have been used for SST by the authors. Even if a full line-by-line RTM were to be used, the forward-model accuracy, though improved, would remain of this order, based on the sensitivity to uncertain parameters illustrated in Table 2. (Note that we are merely making an order of magnitude estimate here—that is, we are concluding that 0.1 K is a better estimate than 0.01 or 1 K.) The precision (scatter) is likely to be similar unless errors in water vapor profiles and water vapor absorption parameters are small.

d. How well can coefficients be specified?

The distinction between $\bar{\epsilon}$ and ϵ' introduced above is relevant because the former (the bias term) determines how well the offset can be specified, whereas the latter (the scatter term) influences the specification of the weighting coefficients.

The covariance matrix, \mathbf{S}_{yy} , used in defining the weights, \mathbf{a} [Eq. (3)], is written in full as

$$\overline{(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T}, \tag{5}$$

that is, it is the expectation of the outer product of $(\mathbf{y} - \bar{\mathbf{y}})$ with itself. What is the forward-model error in the estimate of $(\mathbf{y} - \bar{\mathbf{y}})$ that we form from our RTM? From the definitions and properties of $\bar{\epsilon}$ and ϵ' , it is seen to be only the component ϵ' . Similar considerations apply to \mathbf{s}_{yy} . Thus, the accuracy to which weighting coefficients may be specified depends on the relative precision with which RTMs can simulate BTs for different profiles (and not the absolute accuracy).

The propagation of the error, ϵ' , into the standard deviation of SST retrievals is illustrated in the last column of Table 2 for the forward-model perturbations. Here, three-channel SST coefficients derived from the perturbed BTs are applied to unperturbed BTs, to simulate SST retrievals adversely affected by forward-model error. (Here, the RAL RTM is used. The magnitudes obtained are not sensitive to the RTM. Different coefficients would give different results of a similar magnitude.) The resulting bias and standard deviation (SD) are found. The percentage increases in SD (ΔSD) over that for the “true” coefficients are small, a few percent or less. In actual application, of course, there may be many unknown forward-model errors that combine to give larger values of ΔSD . Nonetheless, this result gives grounds for optimism that with plausible forward-model errors the *relative* precision of the BTs is adequate to define the weighting coefficients near optimally.

By substituting $\mathbf{y} + \bar{\epsilon} + \epsilon'$ for \mathbf{y} in the equation defining the offset coefficient [Eq. (3)], a_0 , we can see that the error in the offset coefficient from forward-model

error is $\mathbf{a}^T \bar{\epsilon}$, that is, it depends on the *absolute* accuracy of the RTM (not the relative precision). The simulated SST biases in Table 2 are effectively evaluations of $\mathbf{a}^T \bar{\epsilon}$ for the forward-model errors indicated (and a three-channel retrieval). Of course, the elements of $\bar{\epsilon}$ are only known in simulation studies. We have argued (above) that the likely magnitude of the elements of $\bar{\epsilon}$ (i.e., the likely magnitude of the biases in modeling each channel) is ~ 0.1 K. Not all forward-model errors will have independent effects on nonoverlapping channels, but for the sake of simplicity, let us assume that the forward-model biases in the SST retrieval channels are independent and of magnitude ϵ (in kelvins). The expectation of the error in the offset coefficient is then

$$\sigma(a_0) = \epsilon \times \sqrt{\sum_i a_i^2} = \epsilon |\mathbf{a}|. \tag{6}$$

Equation (6) states that the expected error in the offset coefficient increases in proportion to the scale of bias in forward modeling, as would be expected. It further states that it also depends on the magnitude of the weighting coefficient vector, \mathbf{a} . This varies with the type of retrieval being done. For the traditional “split window” retrieval (using channels at 11 and 12 μm) near nadir, it is ~ 4 . For a three-channel single-view retrieval (as may be used at night on sensors with a 3.7- or 3.9- μm channel), it is nearer 2. Thus, we can expect a greater spread of biases in RT-based split-window retrievals than in three-channel retrievals. [The task of specifying the offset coefficient for a dual-view sensor (i.e., ATSR-series) may seem to be particularly challenging, since the magnitudes of the weighting vector are ~ 9 and ~ 3 for dual-view two-channel and three-channel coefficients, respectively. However, the forward-model biases are not independent between views, causing Eq. (6) to be an overestimate in that case. In practice, the biases in retrievals have been of order 0.2 K for ATSR (Merchant and Harris 1999) and have been similar for ATSR-2 and for AATSR (L. Horrocks 2002, personal communication).]

We can turn the above arguments around, and state that in order to have RT-based coefficients that give biases < 0.1 K, the absolute accuracy of forward modeling would need to be $< 0.1 \text{ K} / |\mathbf{a}|$, that is, of order a few centikelvins. The results in section 2c suggest that forward models do not currently meet this absolute accuracy by a factor of 2–5 (depending on the RTM, the sensor characterization, the retrieval coefficients, etc.). Therefore, at present, global biases in RT-based SSTs of order a few to several tenths of a kelvin are to be expected, unless the offset coefficient is adjusted empirically.

One can view this conclusion negatively: we cannot yet achieve with the RT approach the accuracy we would like, and offset adjustment is needed in practice. But a more positive viewpoint is that, with the best RTMs and sensor characterization, we are not far from our target,

TABLE 3. Performance of RT-based coefficients in validation.

Sensor/platform/MDB	Coefficient formulation ^a	No. in sample	Bias ^b (K)	Standard deviation (K)	Lowest possible SD ^c (K)	Empirical coeffs SD ^d (K)
AVHRR/NOAA-11/Pathfinder 1990	NLSST	3627	0.34	0.71	0.60	0.63
	Triple		0.07	0.53	0.52	0.53
AVHRR/NOAA-14/Pathfinder 1997	NLSST	4741	0.19	0.65	0.62	0.63
	Triple		0.20	0.56	0.56	0.57
AVHRR/NOAA-16/CMS 2002	NLSST	636	0.00	0.52	0.49	0.50
	Triple		0.39	0.49	0.45	0.48
GOES-8 CMS 2002	NLSST	5612	-0.55	0.59	0.57	0.58
	Triple		-0.29	0.40	0.38	0.39

^a NLSST means “nonlinear SST” and refers to retrieval using the split-window channels with coefficients that vary with a climatological prior SST. Triple means a linear three-channel algorithm.

^b The nighttime skin effect, ~ 0.2 K, has **not** been accounted for here.

^c Coefficients obtained by linear regression of in situ SST and the corresponding observed BTs on all matches.

^d Coefficients specified by linear regression on a 50% subset of matches and then applied to the remaining (independent) matches.

and future improvements in our understanding of the key forward-model parameters may well make 0.1-K accuracy achievable.

3. Practical experience

At Centre de Meteorologie Spatiale (CMS), RT modeling using MODTRAN has been used to derive coefficients for determination of SST from a number of sensors used in operational meteorology (Brisson et al. 2002; Francois et al. 2002). The satellite SSTs have been matched with in situ measurements reported on the global telecommunication system, forming a matchup database (MDB) for each sensor. In this section, we describe the extent to which the retrieval and validation results support the discussion in section 2. We summarize the characteristics of the MDBs, before describing the results obtained from them.

a. Match-up databases used

MDBs are constructed at CMS as follows, for GOES-8 and AVHRR. After cloud screening and SST retrieval are applied to imagery to create SST products, satellite data are collected for 20 km \times 20 km boxes centered on the available matching buoy locations (providing that no more than 40% of this area has been screened as cloudy). These collected data include the retrieved SSTs and the brightness temperatures averaged over only the cloud-free pixels of the validation box. The matchup time window is 3 h for the polar orbiters and half an hour for GOES. Only buoy measurements have been included in these MDBs; that is, ship temperatures are not used. To screen out infeasible in situ measurements, only buoy temperatures within 2 K of the local climatological values have been retained. For the purpose of comparing the quality of retrieval coefficients with our expectations from the discussions above, we here use only those matches where less than 10% of the validation box was screened as cloudy. (In other circumstances, where we wish to assess the retrieval quality

of the combined cloud-clearing/retrieval scheme, it is appropriate to use all the matchups.) CMS MDBs have been used for GOES-8 and the National Oceanic and Atmospheric Administration’s NOAA-16. They cover the western and the northeastern Atlantic, respectively.

When available (i.e., for NOAA-11 and NOAA-14), we have used the “Pathfinder” MDBs (Kilpatrick et al. 2001), because they are representative of the global ocean. For consistency, the pathfinder MDBs have been further screened similarly to those of CMS: in situ measurements must be within 2 K of the climatological mean, and only data showing the best cloud mask test results have been kept.

All AVHRR MDBs have been screened to keep only data with solar zenith angles greater than 110°. In the case of GOES, data from 0300 to 0900 UTC have been excluded because of the “midnight calibration problems” (Johnson and Weinreb 1996; Brisson et al. 2002) (see also section 4), while the limit on solar zenith angle is 90°.

b. Validation results

Table 3 shows validation results for several sensors obtained for matches using coefficients for “nonlinear SST” (using the split-window channels, nominally 11 and 12 μ m, with coefficients that vary with climatological SST) and “triple” (three channels, nominally 3.7, 11, and 12 μ m, with invariant coefficients). The bias and standard deviation of the RT-based coefficients are shown. For comparison, two other standard deviations are shown. The “lowest possible SD” is obtained as follows. Coefficients based on the whole of the MDB dataset are defined by multilinear regression of the matched BTs against the in situ SSTs. These coefficients give zero bias, minimum SD SST estimates when applied back to the data from which they were derived. While this minimum SD is interesting as a target for the quality of RT-based coefficients, it may not fairly represent the SD that could be obtained by empirical regression. This is represented by the “empirical coeffs

SD,” obtained as follows. The MDB is randomly divided in two, and coefficients are defined on one portion, by multivariate least squares regression, as before, but are applied to the BTs in the second portion, that is, they are applied to independent data.

The biases obtained for the RT-based coefficients range between -0.55 and $+0.34$ K for split-window SSTs and -0.29 and $+0.39$ K for triple SSTs. Since the RTM in question is an approximate band model (MODTRAN) and the usual uncertainties in model parameters apply, the total forward-model errors are likely to be at least 0.1 K, possibly somewhat more. These magnitudes of SST bias are consistent with the hypothesis of absolute biases of ≥ 0.1 K in the forward modeling of BTs leading to mis-specification of the offset coefficient by a few tenths of a kelvin via Eq. (6). Moreover, the spread of biases is greater for split-window retrievals than triple retrievals, in agreement with the prediction of Eq. (6), since $|a|$ is greater by a factor of 2 for the former. [Note, however, that the sample of different sensors is too small for statistical significance of this last result. There is also a tendency for triple retrievals to be warmer by of order a tenth of a kelvin than split window. Again, this is not statistically significant. However, it does fit with the pattern of inter-algorithm discrepancy noted in Merchant et al. (1999) using the RAL RTM applied to ATSR. We speculate that this may reflect a systematic error in the spectroscopic parameters in the vicinity of $3.7 \mu\text{m}$.]

The RT-based standard deviations are typically greater than the lowest-possible SD for each sensor by less than 10%. The exception is the case of the AVHRR on *NOAA-11* when using a split-window retrieval; we do not yet understand why this standard deviation is relatively poor (large). Generally, the RT-based standard deviations are therefore only moderately higher than the “lowest possible” and “empirical coeffs” SDs, confirming that the relative accuracy of modeled BTs is sufficient to determine effective weighting coefficients.

In summary, then, practical experience with RT-based coefficients is consonant with the theoretical considerations presented in section 2. The precision with which RTMs simulate relative BT differences (the determining factor for the weight coefficients that affect the retrieval SD) is sufficient to give near-optimal SST SDs in validation data. However, since mean bias in SST retrievals depends on accuracy in the offset coefficient, RT-based coefficients require a further step of bias adjustment of the order of tenths of a kelvin—the absolute accuracy of the forward modeling process is not at the $\ll 0.1$ K level required to obtain biases of less than this magnitude.

4. Offset adjustment: General comments

The basic concept of how to adjust the offset term, a_0 , is straightforward: satellite SSTs are matched against in situ SSTs, and the offset term is modified in the light

of the mean residual between the two. There are, however, a few subtleties to be considered, in the context of the purpose to which the satellite SSTs are to be put.

From the beginning, one must be clear about the definition of the satellite SSTs that are required. Satellites sensors are actually sensitive to the radiometric temperature of the sea surface. This radiometric temperature is essentially the same as that of the air–sea interface temperature, or “ocean skin” (Saunders 1967). The skin temperature in general differs from the temperature a few centimeters below the surface because of a skin effect (a thermal gradient across the sublayers of the air–sea interface). There may also be diurnal warm-layer effect (Fairall et al. 1996) causing temperature stratification in the upper meters of the ocean, as a result of near-surface heating by solar irradiance during daytime. These differences are detectable in comparisons of accurate satellite SSTs and high quality in situ measurements (Murray et al. 2000).

One approach is to take the view that satellite sensors cannot see the bulk SST, and that it should not be pretended that they can. The satellite SST is then a skin SST product, and any offset adjustment should be made with reference to the radiometric temperature. The problem here is that radiometric in situ measurements are rather rare [although becoming more common (Donlon et al. 2002; Kearns et al. 2000)]. For practical purposes, it is sufficient to assume that the nighttime skin–bulk difference comprises only a skin effect. There remains some disagreement about the size and wind speed dependence of that skin effect (Donlon et al. 2002, 1999; Emery et al. 2001; Fairall et al. 1996), but for moderate wind speeds ($4\text{--}10 \text{ m s}^{-1}$) the consensus is that the skin effect is fairly constant and likely to be within 0.05 of -0.2 K (i.e., radiometric temperature cooler than bulk). A practical strategy is, therefore, to derive an offset adjustment for skin SST coefficients (both day- and nighttime coefficients) by setting a_0 such that the mean residual for nighttime moderate-wind matches is -0.2 K.

Alternatively, one may decide to derive bulk SSTs from satellite observations. This can be done by adjusting the satellite retrieval by some model or parameterization of skin and warm-layer effects. This requires auxiliary meteorological information. Bias adjustment then consists in choosing a_0 such that the mean residual of matches is 0.0 K. Note that the offset adjustment is then simultaneously a correction of both the skin–bulk conversion model and the retrieval coefficients, and therefore has no clear physical meaning.

5. Offset adjustment: Example using *GOES-8*

The practicalities of adjusting a_0 are illustrated here by the case of *GOES-8*. The following points are discussed below:

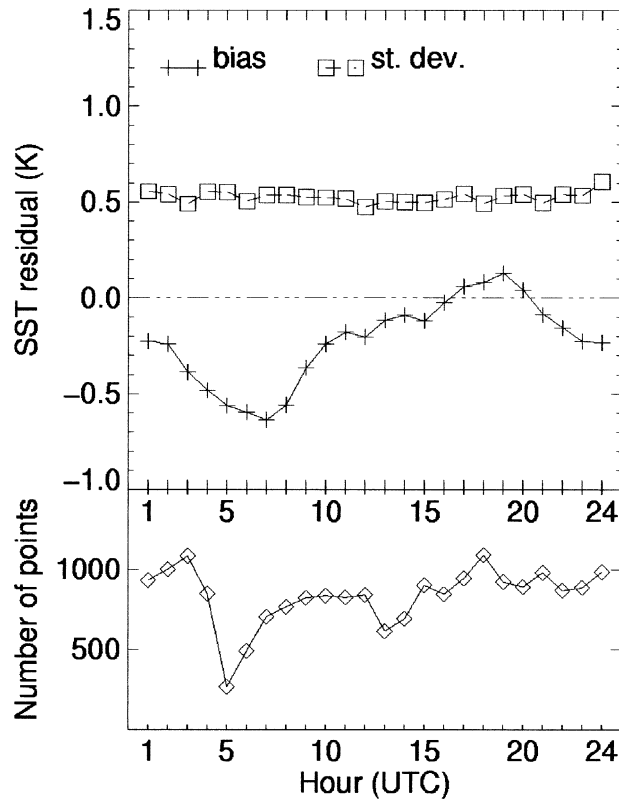


FIG. 2. SST residual (satellite estimate – moored-buoy measurement) as a function of the satellite measurement hour (UTC).

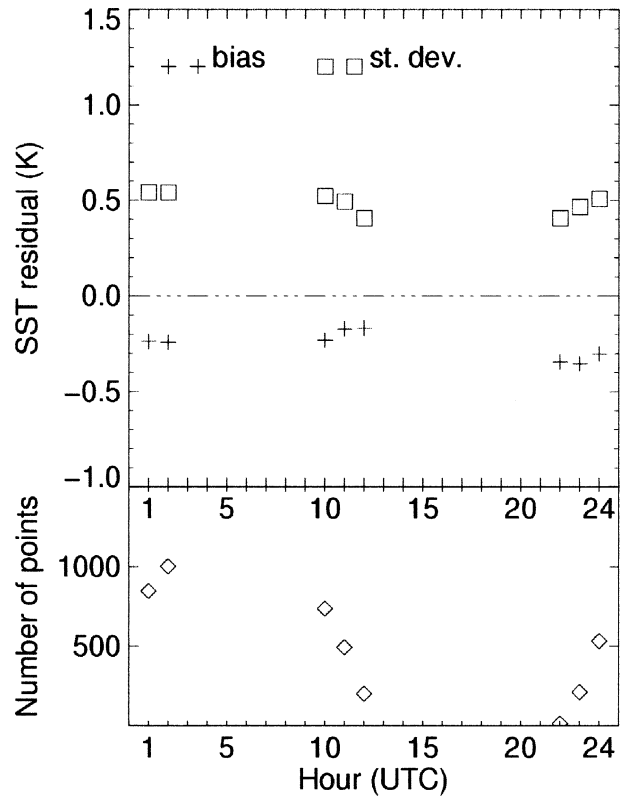


FIG. 3. Same as Fig. 2, except that only nighttime data and data uncontaminated by the midnight calibration error have been kept.

- screening major artifacts in the data: diurnal heating and calibration errors;
- using drifting buoys, moored buoys, or both for the in situ data;
- filtering data with respect to wind speed; and
- refining the elimination of cloud contaminated pixels.

The MDB considered includes matchups from 1 October 2001 to 30 September 2002. Again, we use only matchups where <10% of the validation box was flagged as cloud contaminated by the cloud-screening procedures. This restriction was originally proposed to provide a reasonable guarantee that the clear-sky BTs of the matchups considered were not cloud contaminated.

a. Screening major artifacts

Figure 2 shows the mean difference between SST calculated by the RTM-derived split-window algorithm and coincident moored-buoy measurements (21 112 points), as a function of the hour of day (UTC). Moored-buoy data have been chosen as best adapted to determine diurnal-cycle characteristics (Brisson et al. 2002). Minimum values of about -0.6 K are found around 0500 UTC. These are not geophysically plausible, and reflect the so-called midnight calibration error. Maximum values of about 0.1 K, due to the mean diurnal heating of the water between the in situ sensor and the surface,

occur around 1900 UTC (when most of the ocean matchups have a local solar time of midday to midafternoon).

To determine a_0 (i.e., to correctly adjust the retrieval coefficients) it is thus necessary to eliminate daytime data (having solar zenith angle less than 90°) and data from 0300 to 0900 UTC. (The ranges of hours to exclude are matters of judgement. The ranges we have used are probably conservative.) Figure 3 shows the distribution with time of the remaining data (4602 points).

b. Choice of in situ measurements

When all known artifacts are screened out, the mean difference between calculation and moored-buoy measurements is -0.24 K and the standard deviation is 0.51 K. The bias value is significantly different (0.3 K warmer) than that presented in Table 3 (obtained under similar conditions but with drifting buoys).

There are two main reasons for this difference. First, the spatial distribution of the measurements differs (Fig. 4) in a way that is likely to correlate with atmospheric conditions: the moorings are in more “continental” locations. Second, the in situ measurements are actually slightly different in nature: moored buoys do record temperatures slightly colder than drifters. This was highlighted recently by Emery et al. (2001), albeit the difference they observed was smaller. Since, in this MDB,

the choice between moored buoys and drifters leads to values of a_0 that may differ by 0.2–0.3 K, we conclude that drifters and moorings should not be merged for GOES-East validation or offset adjustment, contrary to the more general conclusion reached by Emery et al. However, these authors also make the excellent point that good meridional coverage is highly desirable, which would suggest that drifters are far more appropriate in the present case. The origin of the drifter–mooring differences in this MDB requires further investigation.

c. Filtering with respect to wind speed

As seen in Fig. 4, the moored buoys are mainly located on the continental shelf and thus could be argued to be not representative of the entire ocean. However, in investigating the effects of wind speed variability, they have a significant advantage over the drifting buoys because they are better instrumented (and maintained?). Of the moored-buoy SSTs, 98% have a coincident wind speed measurement, which greatly facilitates filtering with respect to wind speed. Wind speed is available only for 26% of drifting-buoy matchups in the MDB.

Figure 5 shows the distribution of the SST residuals as a function of wind speed in bins of 3 m s^{-1} . For wind speeds up to 12 m s^{-1} , the numbers of points per bin are sufficient to make statistically significant statements about variation of the SST residual against wind speed. For the range $3\text{--}12 \text{ m s}^{-1}$ the mean residuals are, to within the standard errors, constant against wind speed at -0.23 K . For the $<3 \text{ m s}^{-1}$ range the residual is 0.12 K cooler at -0.35 K . This is a modest difference, but it is statistically significant at the 99.9% confidence level, it is geophysically plausible, and it matches the results observed in completely independent data (Murray et al. 2000). We therefore cautiously attribute this cooling at low wind speed to thickening of the skin layer. Formally, these results support our suggestion that only matchups corresponding to wind speeds between 4 and 10 m s^{-1} should be retained for offset adjustment. In practice, if the proportion of matches outside this range is rather small (as here) this may make only a modest difference to the offset calculated (much less than 0.1 K).

The spatial distribution of drifters is more representative of the ocean, but only 26% of the data have coincident wind and SST measurements. The wind dependence of the residuals is not as clearly interpretable for drifters as for moored buoys (not shown).

d. Refining the elimination of cloud-contaminated pixels

As mentioned above, only validation boxes with fewer than 10% of pixels flagged “cloud” are included among the matchups used in this example. The unflagged pixels used to derive the SSTs are nominally not cloudy, but, in addition, a semiprobabilistic assessment of the quality of each pixel for SST retrieval has been defined

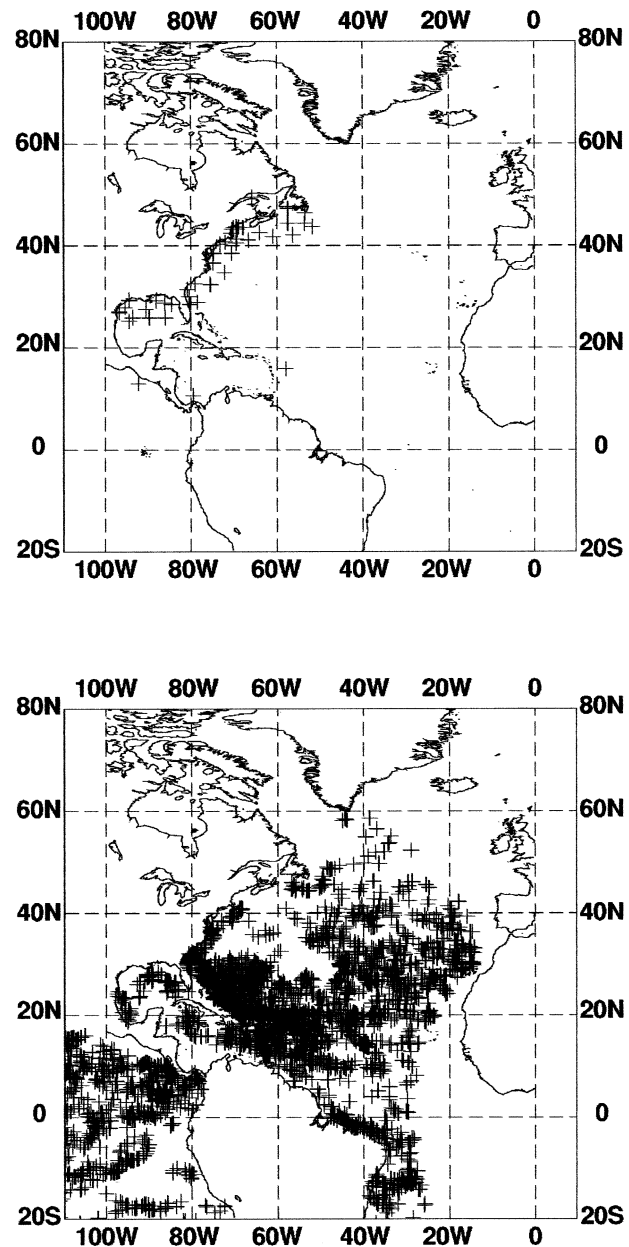


FIG. 4. Measurement distributions: (top) moored buoys, (bottom) drifters.

at CMS (Brisson et al. 2001). This is expressed on a qualitative five-level scale: 1) cloudy (i.e., screened), 2) bad, 3) acceptable, 4) good, and 5) excellent. “Bad” implies that there is some likelihood of residual cloud contamination, or of other contamination such as dust, but not sufficient to trigger the cloud screen. Such pixels are not considered reliable for SST consideration for quantitative purposes but are usable for generating SST fields for visualization purposes (e.g., locating fronts by eye). Categories 3–5 are considered acceptable for SST for quantitative applications. “Excellent” implies that

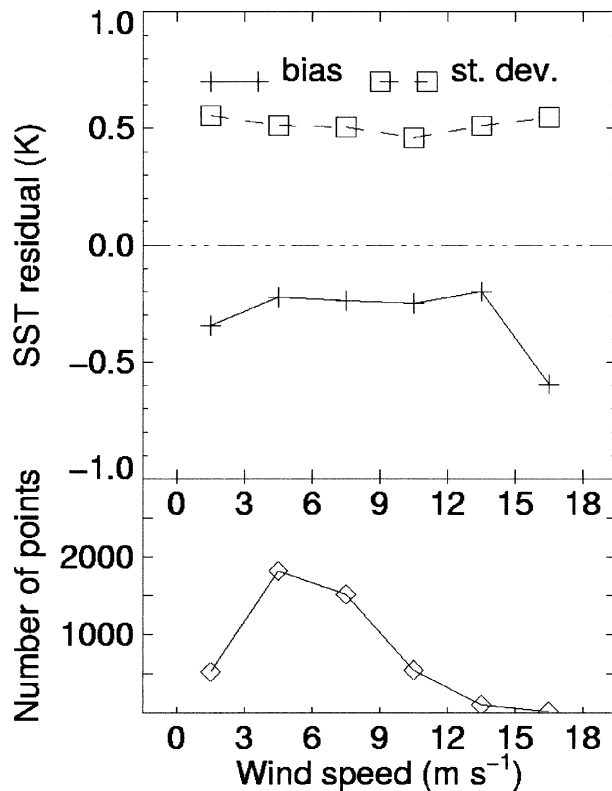


FIG. 5. SST residual (satellite estimate – moored-buoy measurement) as a function of the in situ wind speed.

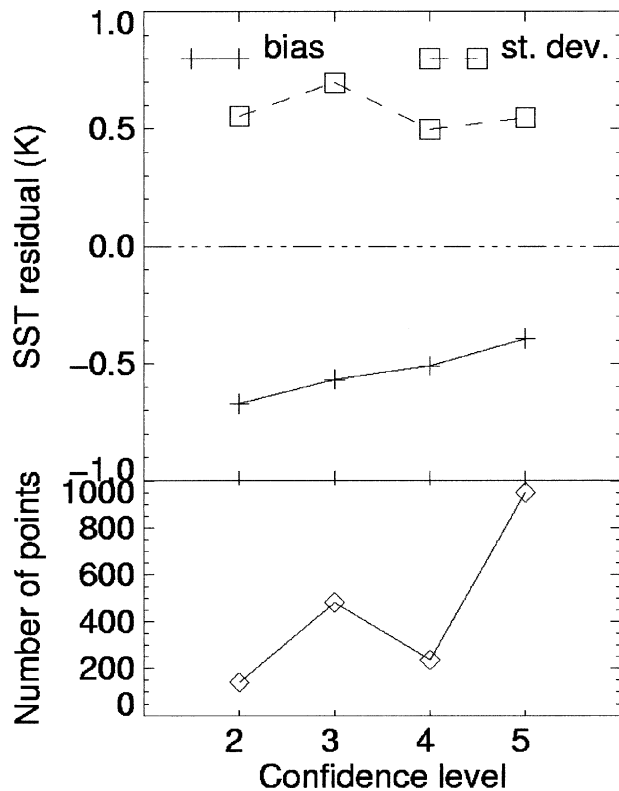


FIG. 6. SST residual (satellite estimate – drifting-buoy measurement) as a function of confidence level.

there is very little likelihood that the pixel is anything other than a truly clear-sky observation.

In Fig. 6 we show the residuals between satellite and drifting-buoy SSTs, plotted as a function of the confidence level. (Since the pixels in the validation box are given a confidence level individually, only cases where $>75\%$ of the pixels have the same confidence level are used.) This graph shows that, although most of the matchups are dominated by excellent (level 5) pixels, many matchups largely comprise pixels in the poor to good categories. The mean residual cools significantly as the confidence level decreases, from -0.39 K for excellent matchups to -0.65 for bad matchups. The sign of the trend is consistent with there being an increasing risk of cloud or other contamination in the lower confidence levels. (The results obtained with moored buoys are quite similar and are not shown here.)

For bias adjustment of RT-based coefficients, one should choose to retain only the highest quality pixels, that is, those equivalent to the CMS category of “excellent.” The adjustment is then the best available estimate of the required correction to a_0 for those retrieval coefficients. (Note, however, that the overall retrieval scheme includes both the cloud-screening step and the SST-estimation step using the retrieval coefficients. One could instead require an offset adjustment that best suits the retrieval scheme as a whole, to account in a mean

sense for the residual uncertainties in cloud screening. It would be appropriate to use the pixels for every confidence level for which quantitative SSTs are generated. This practical expedient may sometimes be necessary, but it is unsatisfactory from a methodological viewpoint. Correcting RT-related biases and taking account of residual cloud errors are really two distinct problems that are more cleanly addressed separately. An approach that does not separate the problems has the disadvantage of folding into the offset adjustment an influence of the cloud-screening scheme. If the cloud screening were changed, the offset adjustment would need to be updated, and in that sense the offset adjustment would have no clear physical meaning.)

e. Lessons learned about doing offset adjustment

In summary, from this example of offset adjustment for *GOES-8*, we observe the following.

- The mean residual varies with the time of day. Daytime data must be excluded to avoid confounding by the diurnal signal. In the case of *GOES-8*, data affected by a diurnal calibration error must also be removed; in general, other sensors may be subject to different artifacts that need to be removed.
- Residuals relative to moored buoys are warmer by 0.3 K than residuals relative to drifters. (This size of

difference may be specific to the *GOES-8* situation, but the possibility of such a difference is a general lesson.) This is a considerable effect and requires further study. Drifters might be preferred as more representative of the global ocean.

- Eliminating low and high wind speed values (<4 and >10 m s⁻¹) is preferable but may not have a significant impact on the resulting offset where low wind speed conditions are relatively rare.
- Residual cloud can affect the offset adjustment at the 0.1-K level, even within the context of a very rigorous cloud-screening scheme.

6. Concluding remarks

Radiative transfer simulation is a powerful approach to defining SST retrieval coefficients. Its advantages over a purely empirical approach include much greater insight into retrieval process, the potential to integrate retrieval definition and cloud-screening design, a clear framework for investigating problems in retrieved SSTs and obtaining solutions, the ability to define SST retrieval coefficients before launch, and, perhaps most importantly, high confidence in the SSTs obtained in areas lacking in situ measurements. In this article we have highlighted, with theoretical arguments and practical experience, one of the limitations of this approach: the forward-modeling process is not yet absolutely accurate enough to specify the offset coefficient to 0.1 K. So, we conclude that empirical adjustment of the offset based on validation data is needed for this accuracy. Recourse to an empirical adjustment, however, does not compromise high confidence in the SSTs obtained in regions that lack in situ data. The offset adjustment we propose is global, and the spatiotemporal distribution of the validation data on which it is based is not as critical as when defining SST retrieval coefficients entirely by empirical means.

Acknowledgments. The authors gratefully acknowledge that this collaboration has been aided by funding under a Visiting Scientist program of EUMETSAT. We also thank the anonymous reviewers and editors of this paper for comments that helped us significantly clarify the presentation of our work.

REFERENCES

- Brisson, A., S. Eastwood, P. Le Borgne, and A. Marsouin, 2001: O&SI SAF sea surface temperatures: Pre-operational results. *Proc. 2001 EUMETSAT Meteorological Satellite Data Users' Conf.*, Darmstadt, Germany, EUMETSAT, 152–159.
- , P. Le Borgne, and A. Marsouin, 2002: Results of one year of preoperational production of sea surface temperatures from *GOES-8*. *J. Atmos. Oceanic Technol.*, **19**, 1638–1652.
- Chevallier, F., 2002: Sampled databases of 60-level atmospheric profiles from the ECMWF analyses. EUMETSAT Tech. Rep. NWP-SAF-EC-TR-004, Version 1.0, 27 pp.
- Deschamps, P. Y., and T. Phulpin, 1980: Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 μm . *Bound.-Layer Meteor.*, **18**, 131–143.
- Donlon, C. J., T. J. Nightingale, T. Sheasby, J. Turner, I. S. Robinson, and W. J. Emery, 1999: Implications of the oceanic thermal skin temperature deviation at high wind speed. *Geophys. Res. Lett.*, **26**, 2505–2508.
- , P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369.
- Edwards, D. P., 1988: Atmospheric transmittance and radiance calculations using line-by-line computer models. *Proc. SPIE*, **298**, 1–23.
- Emery, W. J., D. J. Baldwin, P. Schlüssel, and R. W. Reynolds, 2001: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *J. Geophys. Res.*, **106**, 2387–2405.
- Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.*, **101**, 1295–1308.
- Francois, C., A. Brisson, P. Le Borgne, and A. Marsouin, 2002: Definition of a radiosounding database for sea surface brightness temperature simulations—Application to sea surface temperature retrieval algorithm determination. *Remote Sens. Environ.*, **81**, 309–326.
- Gentemann, C. L., C. J. Donlon, A. Stuart-Menteth, and F. J. Wentz, 2003: Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.*, **30**, 1140, doi:10.1029/2002GL016291.
- Han, Y., J. A. Shaw, J. H. Churnside, P. D. Brown, and S. A. Clough, 1997: Infrared spectral radiance measurements in the tropical Pacific atmosphere. *J. Geophys. Res.*, **102**, 4353–4356.
- Johnson, R. X., and M. Weinreb, 1996: *GOES-8* imager midnight effects and slope correction. *Proc. SPIE*, **2812**, 596–607.
- Kearns, E. J., J. A. Hanafin, R. H. Evans, P. J. Minnett, and O. B. Brown, 2000: An independent assessment of Pathfinder AVHRR sea surface temperature accuracy using the Marine-Atmosphere Emitted Radiance Interferometer (M-AERI). *Bull. Amer. Meteor. Soc.*, **81**, 1525–1536.
- Kilpatrick, K. A., G. P. Podesta, and R. Evans, 2001: Overview of the NOAA/NASA Advanced Very High Resolution Radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, 9179–9197.
- Llewellyn-Jones, D. T., P. J. Minnett, R. W. Saunders, and A. M. Zavody, 1984: Satellite multichannel infrared measurements of sea-surface temperature of the NE Atlantic Ocean using AVHRR2. *Quart. J. Roy. Meteor. Soc.*, **110**, 613–631.
- McClain, E. P., W. G. Pichel, and C. C. Walton, 1985: Comparative performance of AVHRR-based multichannel sea-surface temperatures. *J. Geophys. Res.*, **90**, 1587–1601.
- Merchant, C. J., and A. R. Harris, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements. *J. Geophys. Res.*, **104**, 23 579–23 590.
- , A. R. Harris, M. J. Murray, and A. M. Zavody, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modeling and interalgorithm comparison. *J. Geophys. Res.*, **104**, 23 565–23 578.
- Milosheovich, L. M., H. Vömel, A. Paukkunen, A. J. Heymsfield, and S. J. Oltmans, 2001: Characterization and correction of relative humidity measurements from Vaisala RS80-A radiosondes at cold temperatures. *J. Atmos. Oceanic Technol.*, **18**, 135–156.
- Murray, M. J., M. R. Allen, C. J. Merchant, A. R. Harris, and C. J. Donlon, 2000: Direct observations of skin-bulk SST variability. *Geophys. Res. Lett.*, **27**, 1171–1174.
- Rodgers, C. D., 1976: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.*, **14**, 609–624.
- , 1990: Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.*, **95**, 5587–5595.

- Rothman, L. S., and Coauthors, 2003: The Hitran molecular spectroscopic database: Edition of 2000 including updates through 2001. *J. Quant. Spectros. Radiat. Transfer*, **82**, 5–44.
- Saunders, P. M., 1967: The temperature at the ocean–air interface. *J. Atmos. Sci.*, **24**, 269–273.
- Saunders, R., M. Matricardi, and P. Brunel, 1999: An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quart. J. Roy. Meteor. Soc.*, **125**, 407–425.
- Soden, B. J., and F. P. Bretherton, 1994: Evaluation of water vapor distribution in general circulation models using satellite observations. *J. Geophys. Res.*, **99**, 1187–1210.
- Walton, C. C., W. G. Pichel, J. F. Sapper, and D. A. May, 1998: 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.*, **103**, 27 999–28 012.
- Watts, P. D., M. R. Allen, and T. J. Nightingale, 1996: Wind speed effects on sea surface emission and reflection for the along track scanning radiometer. *J. Atmos. Oceanic Technol.*, **13**, 126–141.
- Wu, X. Q., and W. L. Smith, 1997: Emissivity of rough sea surface for 8–13 μm : Modeling and verification. *Appl. Opt.*, **36**, 2609–2619.
- Zavody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones, 1995: A radiative transfer model for sea-surface temperature retrieval for the along-track scanning radiometer. *J. Geophys. Res.*, **100**, 937–952.