Improvements of Satellite SST Retrievals at Full Swath

Walton McBride[1], Robert Arnone[2], Jean-Francois Cayula[3]

[1]*Naval Research Lab, Code 7333, 1009 Balch Blvd,*
Stennis Space Center, MS 39529

[2] *University of Southern Mississippi, Department of Marine Sciences,*
Stennis Space Center, MS 39529

[3] *QinetiQ North America, Services & Solution Group,*
Stennis Space Center, MS 39529

[*]*Corresponding author:* walt.mcbride@nrlssc.navy.mil

ABSTRACT

The ultimate goal of the prediction of Sea Surface Temperature (SST) from satellite data is to attain an accuracy of 0.3°K or better when compared to floating or drifting buoys located around the globe. Current daytime SST algorithms are able to routinely achieve an accuracy of 0.5°K for satellite zenith angles up to 53°. The full scan swath of VIIRS (Visible Infrared Imaging Radiometer Suite) contains satellite zenith angles up to 70°, so that successful retrieval of SST from VIIRS at these higher satellite zenith angles would greatly increase global coverage. However, the accuracy of the SST algorithms steadily degrades to nearly 0.7°K as the satellite zenith angle reaches its upper limit, due mostly to the effects of increased atmospheric path length. Both MCSST (Multiple-Channel) and NLSST (Non-Linear) algorithms were evaluated using a global data set of in-situ buoy and satellite brightness temperatures, in order to determine the impacts of satellite zenith angle on accuracy. Results of our analysis showed how accuracy in SST retrievals is impacted by the aggressiveness of the pre-filtering of buoy matchup data, and illustrated the importance of fully exploiting the information contained in the first guess temperature field used in the NLSST algorithm. Preliminary results suggested that SST retrievals could be obtained using the full satellite swath with a 30% improvement in accuracy at large satellite zenith angles and that a fairly aggressive pre-filtering scheme could help attain the desired accuracy of 0.3°K or better using over 75% of the buoy data.

Keywords: sea surface temperature, VIIRS satellite, remote sensing, full swath.

INTRODUCTION

The ultimate goal of the prediction of Sea Surface Temperature (SST) from satellite data is to attain an accuracy of 0.3°K or better when compared to floating or drifting buoys located around the globe. Current daytime SST algorithms are able to routinely achieve an accuracy of 0.5°K for satellite zenith angles up to 53°. The full scan swath of VIIRS (Visible Infrared Imaging Radiometer Suite) contains satellite zenith angles up to 70°, so that successful retrieval of SST from VIIRS at these higher satellite zenith angles would greatly increase global coverage. However, the accuracy of the best SST algorithm steadily degrades to nearly 0.7°K as the satellite zenith angle reaches its upper limit, due mostly to the effects of increased atmospheric path length.

Ideally, the goal of 0.3°K accuracy for angles up to 70° is to be achieved by using only data retrieved from satellites such as VIIRS. Historically, the MCSST algorithm was derived with this approach in mind [1], where the coefficient of

the $\Delta T = T_{11}-T_{12}$ term is known in the literature as Gamma. However, the use of either climatology or first guess temperature field, T_{field}, is invoked by Walton [1], to express Gamma as a linear function of surface temperature: Gamma = Constant*T_{field}. This is justified by experimental data gathered by Walton, who shows that its incorporation into MCSST to accentuate the linear dependence of Gamma on surface temperature, known as NLSST, results in better accuracy. This slight improvement, however, also results in the mingling of both Celsius and Kelvin temperatures into the NLSST algorithm. Previous VIIRS SST algorithm studies show that T_{field} should be in Celsius rather than Kelvin for best accuracy in the daytime NLSST, while the reverse is true for nighttime NLSST.

Recently, Cayula and May [2] revisited this problematic inconsistency in the choice of units by investigating the effect of adding an offset from 0° to 300° to T_{field} in both the daytime and nighttime NLSST. This new version, coined the expanded NLSST, shows a slightly better accuracy with offsets of 16.2° and 43.3° for daytime and nighttime expanded NLSST, respectively.

The use of T_{field} by Walton [1] to characterize the Gamma coefficient was in effect equivalent to using it as a predictor. However, the full information provided by T_{field} is somewhat tamed by the fact that it is a multiplier to another predictor, the ΔT term, which influences their synergy as a combined predictor. We were therefore motivated to investigate the use of T_{field} as an additional, separate predictor in an effort to extract the maximum amount of information it contains with an eye to improving the accuracy of SST predictions up to 53° Sun zenith angles, as well as exploring the full swath up to 70°. Only the MCSST and NLSST daytime algorithms will be addressed here.

After first writing out the equations for MCSST, NLSST and TfieldSST, the SST algorithm presented here that includes T_{field} as a separate predictor, our approach is guided visually through gradual attempts to increase the agreement between scatterplots involving buoy data, $T_{buoy}-T_{11}$ versus $\Delta T = T_{11}-T_{12}$, and scatterplots involving SST predictions, SST-T_{11} versus $\Delta T = T_{11}-T_{12}$. Here, T_{buoy} represents temperatures in Kelvin units recorded at buoy locations around the globe, while SST are predictions from a particular SST algorithm and are also in Kelvin units. Three characterizations of T_{field} are used, differing in their spatial resolution, in order to assess the associated change in accuracy generated by TfieldSST. Analysis of relationships between the regression coefficients is performed, resulting in a satisfying and intuitive representation of the TfieldSST algorithm in terms of SST_{MC} and T_{field}.

RESULTS

We begin by first presenting the equations for MCSST, NLSST and TfieldSST in that sequence in order to define the notation and conventions for the terminology used in this paper:

$$SST_{MC} = c_1^{MC}T_{11} + c_2^{MC}\left(T_{11}-T_{12}\right) + c_3^{MC}S\left(T_{11}-T_{12}\right) - c_4^{MC}$$

$$SST_{NL} = c_1^{NL}T_{11} + c_2^{NL}T_{field}\left(T_{11}-T_{12}\right) + c_3^{NL}S\left(T_{11}-T_{12}\right) - c_4^{NL}$$

$$SST_{Tfield} = c_1^{Tfield}T_{11} + c_2^{Tfield}\left(T_{11}-T_{12}\right) + c_3^{Tfield}S\left(T_{11}-T_{12}\right) - c_4^{Tfield} + c_5^{Tfield}T_{field}$$

where $S = \sec\theta_{zenith} - 1$. Note the negative sign in front of the c_4 coefficients, in order to be able to only have positive values in coefficient tables to be presented later.

A buoy data set for the month of June 2012 created by NAVOCEANO [2] was used to produce all of the results presented here. The resulting regression coefficients for the MCSST algorithm when the buoy and SST prediction temperatures are in °Celsius were as follows:

$$c_1^{MC} = 1.009 \quad c_2^{MC} = 2.475 \quad c_3^{MC} = 1.282 \quad c_4^{MC} = 274.9$$

while, when the buoy and SST prediction temperatures are in °Kelvin (= °Celsius + 273.15°):

$$c_1^{MC} = 1.009 \quad c_2^{MC} = 2.475 \quad c_3^{MC} = 1.282 \quad c_4^{MC} = 1.75$$

Note that only the value of c_4^{MC} is affected when comparing MCSST predictions against different buoy temperature scales. Subtracting T_{11} from both sides of this last equation, the MCSST algorithm can be rewritten as:

$$SST_{MC} - T_{11} = \left(c_1^{MC} - 1\right)T_{11} + c_2^{MC}\left(T_{11} - T_{12}\right) + c_3^{MC} S\left(T_{11} - T_{12}\right) - c_4^{MC}$$

or, in a form suitable for scatterplot analysis:

$$SST_{MC} - T_{11} = \left(c_2^{MC} + c_3^{MC} S\right)\left(T_{11} - T_{12}\right) + Yoffset$$

where $Yoffset = \left(c_1^{MC} - 1\right)T_{11} - c_4^{MC}$ is the narrow range of vertical intercepts when T_{11}-T_{12} is equal to 0.0, since it depends on T_{11}. It is narrow because its T_{11} coefficient in our case is 0.009. As will be seen, the value of this offset is not as important as are the potential clues that can be gathered from comparing the shape and structure of the buoy data and SST predictions scatterplots. Side-by-side comparison should reveal the similarities and differences between the buoy data and SST predictions, as well as suggest an approach to improve the accuracy of SST algorithms. Figure 1 displays the challenge of SST predictions of sea surface temperature. The filtering parameters used are shown in the upper left and the buoy data scatterplot, T_{buoy}-T_{11} versus T_{11}-T_{12} in the top middle. The lower plots show the scatterplots predicted by MCSST and NLSST, from left to right.

$$SST_{MC} = c_1^{MC} T_{11} + c_2^{MC}\left(T_{11} - T_{12}\right) + c_3^{MC} S\left(T_{11} - T_{12}\right) - c_4^{MC} \qquad SST_{NL} = c_1^{NL} T_{11} + c_2^{NL}\left(T_{field}\right)\left(T_{11} - T_{12}\right) + c_3^{NL} S\left(T_{11} - T_{12}\right) - c_4^{NL}$$
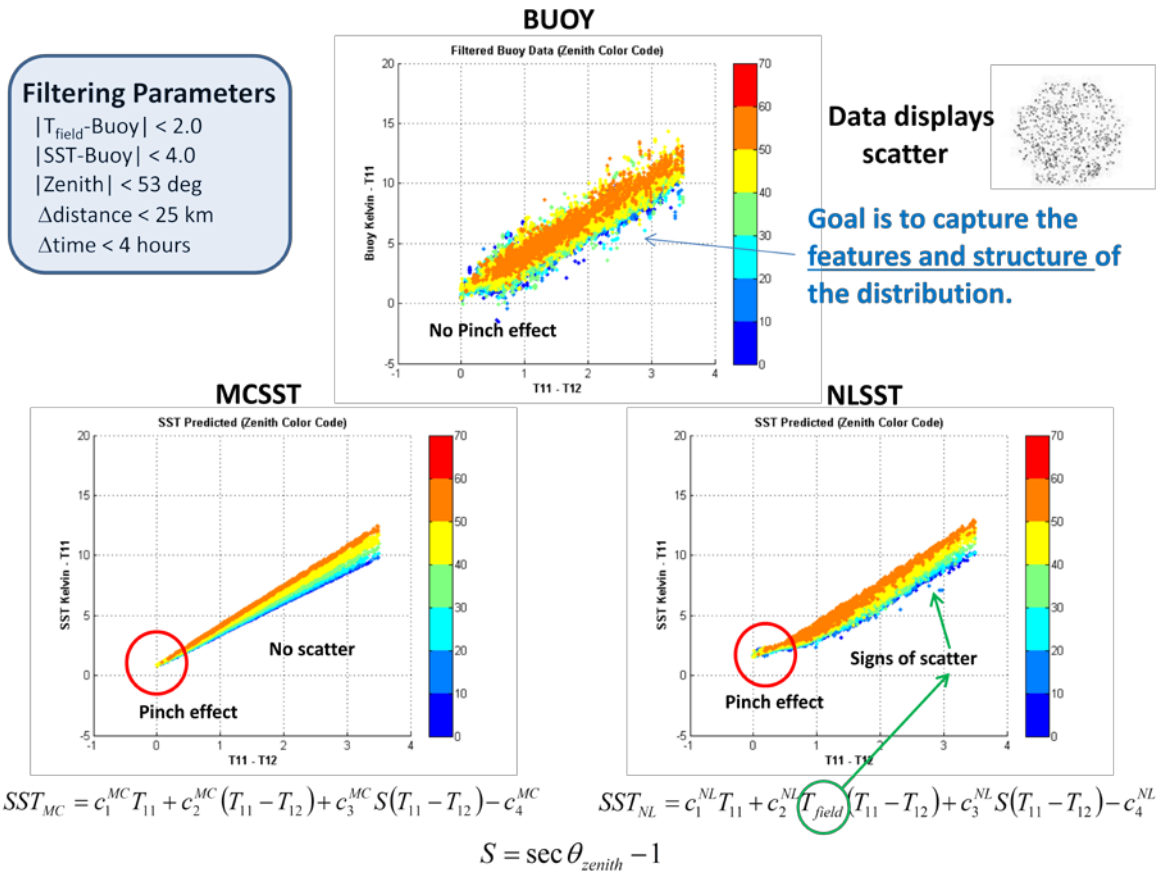
$$S = \sec\theta_{zenith} - 1$$

Figure 1. Visual comparison between scatterplots involving buoy data and SST predictions for Sun zenith angles up to 53°.

The most striking visual differences occur when comparing the buoy data scatterplot with the MCSST scatterplot on the bottom left. The "pinch effect" and the divergence of the separate 10° zenith bands can be identified with the form of the MCSST equation. The main discrepancy appears to be the absence of "scatter" when compared to the buoy data scatterplot.

Interestingly, the inclusion of T_{field} in the NLSST equation as part of the second term somehow introduced a slight amount of scatter as can be seen in the lower right NLSST scatterplot. The "pinch effect" is still there, but moved up slightly due to the non-linearity introduced by Tfield as a multiplier.

As strongly suggested by this visual comparison between buoy and SST predictions, efforts to reduce the rms errors down to 0.3°K from the present daytime values hovering around 0.5°K will undoubtedly require the introduction of scatter and randomness into future SST algorithms. The question therefore arises as to whether, if used as a separate predictor in the MCSST algorithm, T_{field} can provide more pronounced scatter and randomness? However, inclusion of an additional predictor requires precaution vis-à-vis issues of multi-collinearity, as well as measures of its accuracy.

Characterization of T_{field}
For this study, three characterizations of T_{field} are used in order to gauge the effects of increasing spatial resolution and accuracy: Clim, K100 and K10. Both Clim and K100 have 100 km spatial resolution, with K10 enjoys a 10 km resolution. The global rms errors are taken here to be roughly 1.0°K, 0.7°K and 0.55°K, respectively [2]. Both K100 and K10 are updated daily at NAVOCEANO, where there is an effort to create K2 at 2 km spatial resolution. Aside from Clim, they are all derived from satellite data and should provide additional information to the MCSST algorithm to reduce its present rms error. As previously mentioned, this new version of MCSST is referred to here as TfieldSST. However, if this additional information is redundant, issues of multi-collinearity will surface and perhaps make the algorithm unstable.

Multicollinearity of Tfield with Existing Predictors
The fact that the inclusion of T_{field} as a product term in NLSST generated randomness in the corresponding scatterplot invites an investigation into considering its effect as a separate predictor. Before we do this, however, issues of its potential multicollinearity must be addressed. Collinearity is defined as the correlation among the predictors in a multiple Regression, and it therefore entangles the effects of the predictors, complicating the interpretation.

"Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. In this situation the coefficient estimates may change erratically in response to small changes in the model or the data. Multicollinearity does not reduce the predictive power or reliability of the model as a whole, at least within the sample data themselves; it only affects calculations regarding individual predictors. That is, a multiple regression model with correlated predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others. A high degree of multicollinearity can also prevent computer software packages from performing the matrix inversion required for computing the regression coefficients, or it may make the results of that inversion inaccurate" [3].

If the correlation is too high, then there is no unique solution to the regression coefficients. There is no rigorous threshold as to what "too high" means. However, unstable results are expected if no additional information is associated with the new predictor, and existing coefficients change markedly when collinear predictor is added. Nevertheless, stability may still occur if the correlation with the dependent variable is greater than the correlation with existing predictors.

Figure 2 displays scatterplots of each of the three characterizations of T_{field} versus $T_{11.}$ The plots on the left side are color-coded with respect to the Sun zenith angle, while the ones on the right are color-coded with respect to the absolute value of latitude. The corresponding correlation coefficients are shown at the far left. It can be seen that a high degree of correlation exists for all three cases, the highest being associated with the greatest spatial resolution, K10. Therefore, inconsistent predictions may be generated, if the additional information provided by T_{field} is highly redundant compared to the satellite data.
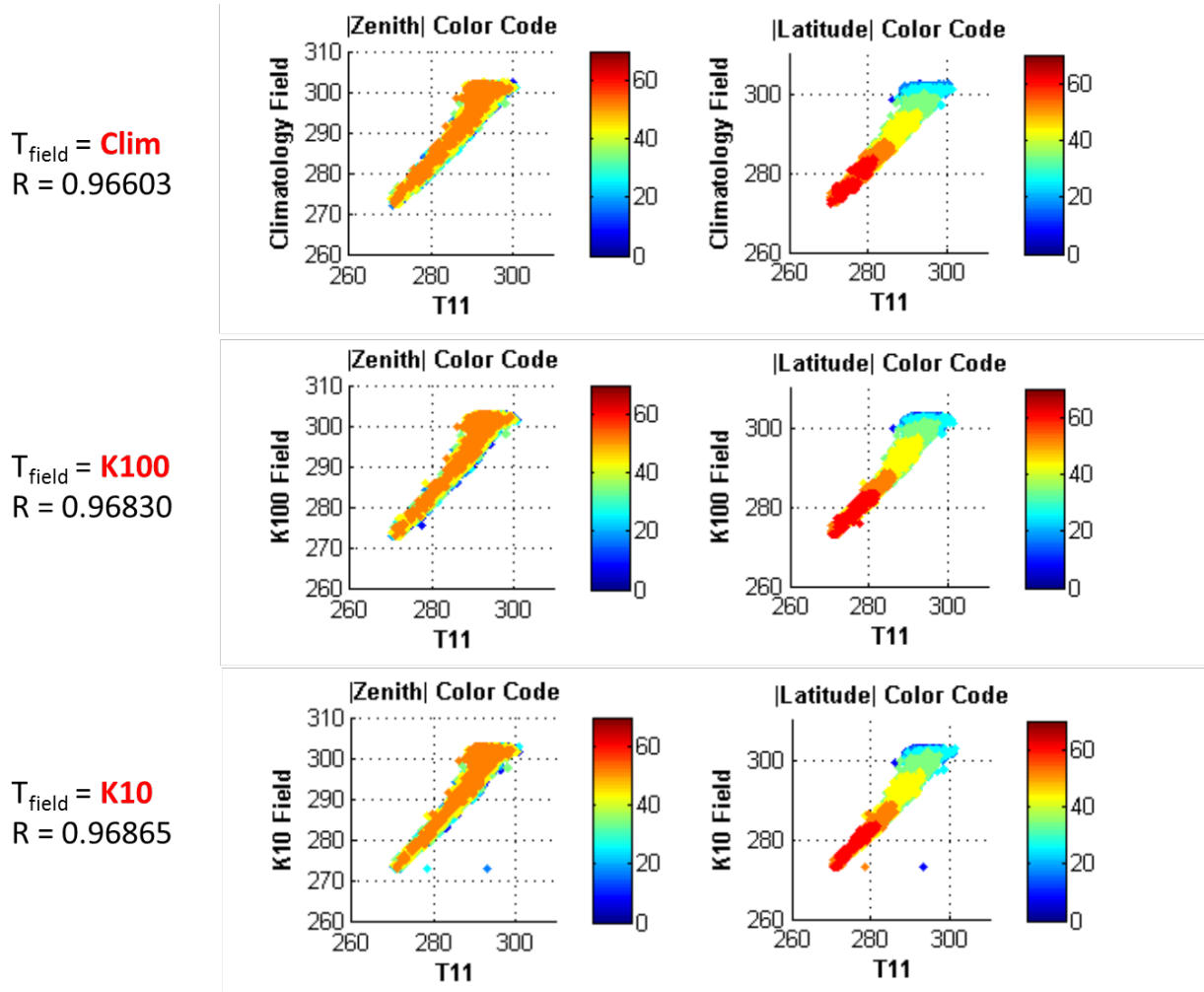
Figure 2. Scatterplots of the three Tfield characterizations versus T_{11}, and associated correlation coefficients.

Figure 2 reveals that all the scatter plots display very high correlation between T_{field} and T_{11}. As expected due to the increased spatial resolution and accuracy, the correlation coefficient R is very high and slowly increase from T_{field} = Clim to K100 to K10. These high values of R should result in competition between the values of the coefficients of T_{11} and Tfield: c_1^{Tfield} and c_5^{Tfield} values should move in opposite directions in a complementary fashion. The larger issue is whether or not the additional information provided by T_{field} is redundant, resulting in an unstable algorithm. Only an implementation of the algorithm is able to address this issue.

Effect of Gradual Increase of T_{field} Contribution to NLSST
Armed with the above warning signs, we performed an investigation similar to Cayula and May [2] using NLSST, but adding an offset Tdiffoffset to the ΔT and SΔT terms instead of Tfield.

$$SST_{NL} = c_1^{NL}T_{11} + c_2^{NL}T_{field}(T_{11} - T_{12} + Tdiffoffset) + c_3^{NL}S(T_{11} - T_{12} + Tdiffoffset) - c_4^{NL}$$

Since T_{field} is a multiplier of ΔT, this will allow us to visually evaluate the effect of gradually increasing the effect of T_{field} and to gauge its effect on the pattern of a scatterplot of SST_{NL}-T_{11} versus T_{11}-T_{12}.
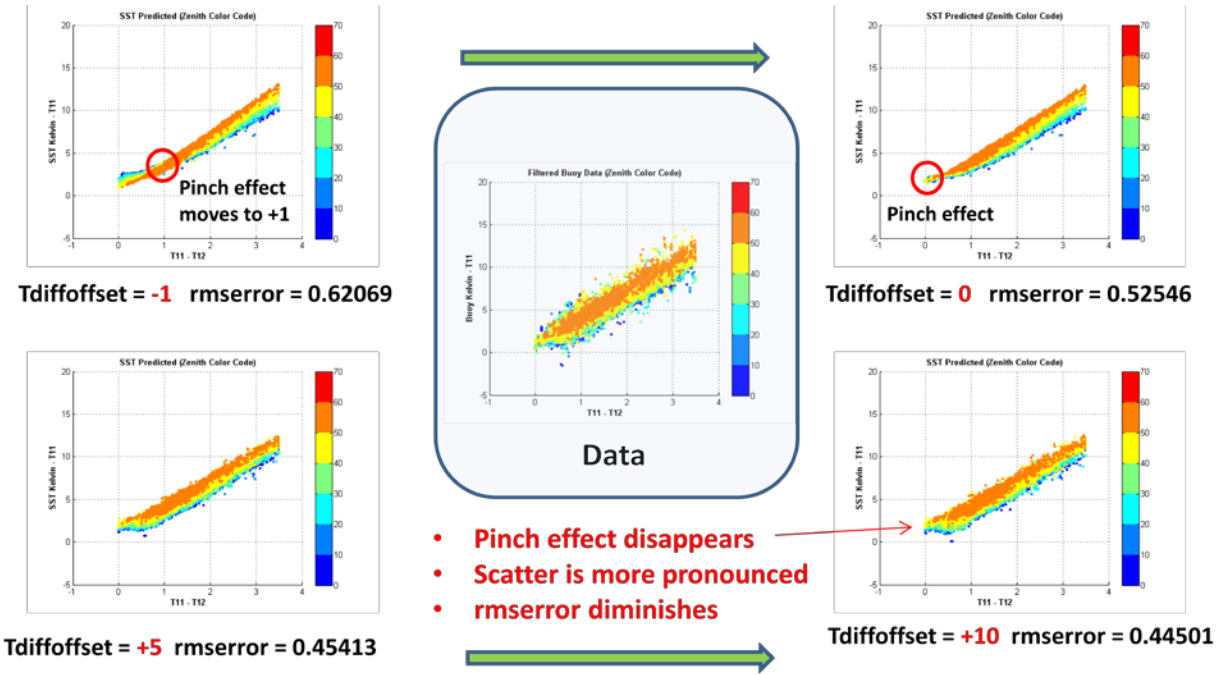
5

Figure 3. Effects on NLSST-generated scatterplots due to addition of Tdiffoffset to T_{11}-T_{12}.

The original NLSST with no Tdiffoffset is located at the upper right corner of Figure 3. The upper left corner plot shows the effect of Tdiffoffset = -1, which mostly serves to translate the "pinch effect" horizontally to where T_{11}-T_{12} = 1. However, increasing Tdiffoffset to 5 and 10 on the lower part of Figure 3 results in increased scatter and randomness, as well as decreasing the rms error from 0.52546 to 0.45413 to 0.44501. These are the effects mentioned previously that we consider to be essential ingredients for future efforts to reduce rms error. These desired trends are a strong indication that T_{field} could be the vehicle to introduce scatter and randomness towards this goal and reduce the rms error.

Adding T_{field} as a Separate Predictor in MCSST
From this point forward, MCSST will be used instead of NLSST in order to clearly isolate the effects of adding T_{field} as a separate predictor. In this section, the effects of the level of aggressiveness of pre-filtering of the buoy dataset for all three characterizations of T_{field} will be tabulated and overall relationships between regression coefficients that emerge from this investigation will be presented. In all of the following tables, c_4 values are shown for SST predictions against buoy temperatures in °Celsius in order to clearly see the trend in its value due to its higher magnitude.

Table 1 is a repository for 18 runs of TfieldSST, which includes T_{field} as a separate predictor. These runs are broken down as follows: for each the three characterizations of T_{field} (Clim, K100, K10), 6 runs representing aggressive to non-aggressive pre-filtering of the buoy data with respect to T_{field} were performed. The filtering bands are shown in the leftmost column of Table 1. The TfieldSST equation is shown at the top without superscripts on the coefficients due to lack of space in trying to line them up with their associated column. Each column represents the values of the TfieldSST coefficients. The sum of c1 and c5, which we anticipate to be complementary due to the high correlation between T_{11} and T_{field}, is shown in the next-to-last column, while the associated rms error is displayed as the end column.

The overall gradual trends in the values of the coefficients serve to ease the initial concerns about unstability of the TfieldSST algorithm. The rms error gradually increases as the aggressiveness of the pre-filtering diminishes for all three characterizations of T_{field}. It can be seen that the rms error also diminishes as T_{field} increases in accuracy, from Clim to K100 to K10. Remarkably, and a strong confirmation of the stability of the TfieldSST algorithm, the sum of c_1 and c_5

appears to be constant, of magnitude very close to 1.0, illustrating their complementarity due to their high correlation. As a shorthand in notation, all rms errors will be understood to be reported in Kelvin units.

One should keep in mind that, although aggressive pre-filtering serves to reduce the rms error, it also can severely reduces the number of buoy data points with which the regression analysis can be performed. The rows corresponding to the value of 2.0 for pre-filtering, presently being used, is highlighted in blue.

Filtering

$$SST = c_1 T_{11} + c_2(T_{11} - T_{12}) + c_3 S(T_{11} - T_{12}) - c_4 + c_5 T_{field}$$

Clim, K100, K10

Clim

	$c_1 T_{11}$	$c_2(T_{11}-T_{12})$	$c_3 S(T_{11}-T_{12})$	c_4	$c_5 T_{field}$	$c_1 + c_5$	rmserror
$\lvert T_{field}\text{-Buoy}\rvert < 0.1$	0.021	0.057	0.026	5.921	0.978	0.999	0.07175
$\lvert T_{field}\text{-Buoy}\rvert < 0.2$	0.074	0.194	0.088	20.33	0.927	1.001	0.12447
$\lvert T_{field}\text{-Buoy}\rvert < 0.5$	0.245	0.643	0.261	66.82	0.759	1.004	0.25600
$\lvert T_{field}\text{-Buoy}\rvert < 1.0$	0.471	1.218	0.535	128.5	0.541	1.012	0.36675
$\lvert T_{field}\text{-Buoy}\rvert < 2.0$	0.649	1.663	0.736	177.0	0.364	1.013	0.45077
$\lvert T_{field}\text{-Buoy}\rvert < 5.0$	0.752	1.914	0.853	204.9	0.259	1.011	0.49914

K100

	$c_1 T_{11}$	$c_2(T_{11}-T_{12})$	$c_3 S(T_{11}-T_{12})$	c_4	$c_5 T_{field}$	$c_1 + c_5$	rmserror
$\lvert T_{field}\text{-Buoy}\rvert < 0.1$	0.009	0.018	0.005	2.458	0.991	1.000	0.06968
$\lvert T_{field}\text{-Buoy}\rvert < 0.2$	0.043	0.096	0.055	11.86	0.957	1.000	0.12447
$\lvert T_{field}\text{-Buoy}\rvert < 0.5$	0.195	0.475	0.229	53.23	0.806	1.001	0.25554
$\lvert T_{field}\text{-Buoy}\rvert < 1.0$	0.389	0.958	0.467	106.1	0.615	1.004	0.36104
$\lvert T_{field}\text{-Buoy}\rvert < 2.0$	0.527	1.310	0.633	143.6	0.479	1.006	0.42815
$\lvert T_{field}\text{-Buoy}\rvert < 5.0$	0.585	1.462	0.697	159.5	0.420	1.005	0.46373

K10

	$c_1 T_{11}$	$c_2(T_{11}-T_{12})$	$c_3 S(T_{11}-T_{12})$	c_4	$c_5 T_{field}$	$c_1 + c_5$	rmserror
$\lvert T_{field}\text{-Buoy}\rvert < 0.1$	0.013	0.032	0.011	3.600	0.986	0.999	0.06991
$\lvert T_{field}\text{-Buoy}\rvert < 0.2$	0.055	0.130	0.063	15.10	0.945	1.000	0.12362
$\lvert T_{field}\text{-Buoy}\rvert < 0.5$	0.188	0.454	0.239	51.21	0.814	1.002	0.24467
$\lvert T_{field}\text{-Buoy}\rvert < 1.0$	0.326	0.810	0.406	89.01	0.677	1.003	0.33348
$\lvert T_{field}\text{-Buoy}\rvert < 2.0$	0.419	1.049	0.518	114.3	0.585	1.004	0.39268
$\lvert T_{field}\text{-Buoy}\rvert < 5.0$	0.467	1.169	0.571	127.2	0.539	1.006	0.43082

Table 1. Effects of pre-filtering buoy data and increased spatial resolution of T_{field} using MCSST for Sun zenith angles from 0° to 53°.

Changes in Scatterplots due to T_{field} as a Separate Predictor
The addition of T_{field} as a separate predictor also results in the addition of significant randomness to the scatterplot, a crucial feature mentioned earlier towards increasing the agreement between SST predictions and the buoy data. Figure 4 compares the buoy data scatterplot versus the scatterplot predicted by MCSST, basically TfieldSST without T_{field} as a separate predictor, for zenith angles from 0° to 70°, in 10° bands.

Because the plotting routine of each band occurs sequentially, some of the scatter in one band overlaps the previous band. As a way for the reader to assess the amount of overlap between zenith bands, the plotting sequence is shown for both Blue to Red and Red to Blue (top for buoy data, bottom for MCSST). The resulting dual scatterplots for MCSST reveal that there is practically no overlap of zenith band scatter, except for low values of T_{11}-T_{12}.

In sharp contrast, however, the buoy data scatterplot shows pronounced overlap of zenith band scatter. Figure 5 is similar to Figure 4, but with the lower scatterplots predicted by TfieldSST using T_{field} = K10. As can be seen, features very

similar to the buoy data scatterplot emerge, thereby establishing a strong link between reduced rms error and the presence of scatter and randomness in the predicted scatterplot, as was conjectured earlier.
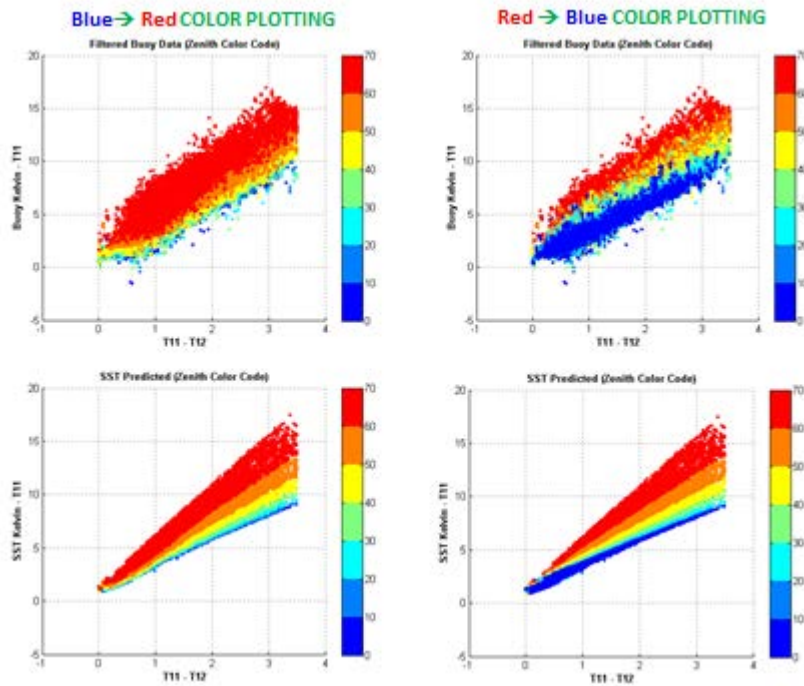

Figure 4. Comparison of MCSST predictions with Buoy Data scatterplots, without T$_{field}$, for Sun zenith angles from 0° to 70°.
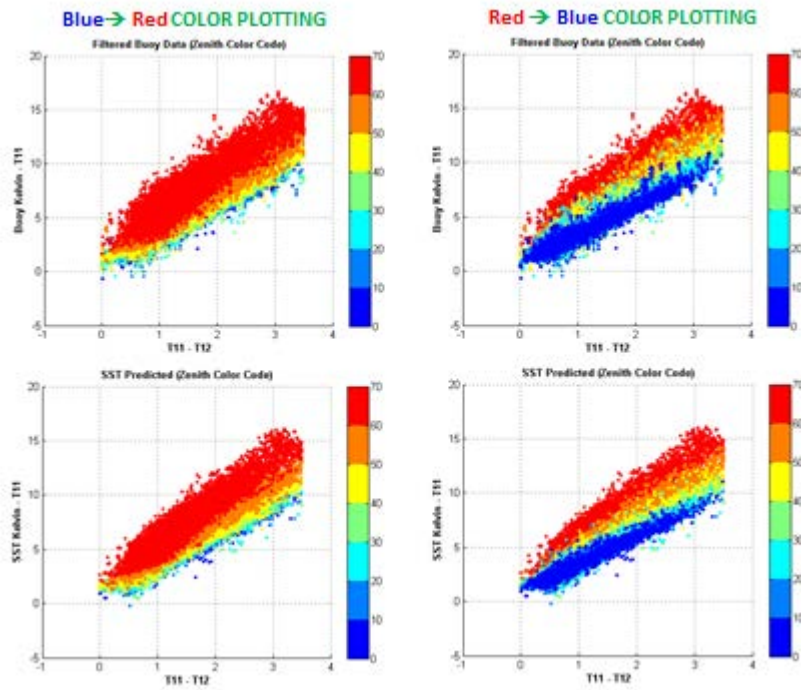

Figure 5. Comparison of MCSST predictions with Buoy Data scatterplots, with T$_{field}$ = K10, for Sun zenith angles from 0° to 70°.

Dependence of Regression Coefficients on Zenith and Latitude

The extreme stability of the TfieldSST algorithm when T_{field} = K10 is also evident from the gradual changes in the values of the regression coefficients as either zenith angle or latitude is broken up into equally spaced bands. Besides the steady increase of rms error as the zenith angle increases, Table 2 below shows this general gradual change in the values of the coefficients with a single glaring exception, such as the value of -0.011 for c_3 for the 20°-30° zenith band. It is possible that some outlier buoy data values creeped into the original creation of the data file.

In addition, global coefficient values over the entire zenith span from 0° to 70° is shown in order to compare them with the global coefficient values derived from MCSST shown in green. Again, the rms error shows a gradual increase in rms error over the 10° zenith bands when using these global coefficients, contrasted with the MCSST rms errors in green.

From GUI results

|Zenith| Band	#pts	rmserror	bias	c_1	c_2	c_3	c_4	c_5	(c_1+c_5)
+60 to +70	21213	0.41834	-8.37e-13	0.230	0.292	0.296	62.7	0.794	1.024
+50 to +60	19667	0.39780	-2.03e-13	0.296	0.621	0.349	80.9	0.721	1.017
+40 to +50	15353	0.39463	2.83e-13	0.343	0.814	0.447	93.4	0.665	1.008
+30 to +40	12415	0.38942	-9.51e-13	0.407	1.004	0.742	110.8	0.594	1.001
+20 to +30	10231	0.38372	-2.92e-12	0.469	1.264	-0.011	127.7	0.533	1.002
+10 to +20	9477	0.37942	7.47e-13	0.521	1.325	1.087	141.9	0.482	1.003
0 to +10	9151	0.37119	-6.32e-13	0.493	1.292	0.922	134.3	0.506	0.999

Using global coefficients

	#pts	rmserror	bias	c_1	c_2	c_3	c_4	c_5	(c_1+c_5)
	97507	0.41055	3.66e-12	0.294	0.610	0.388	80.10	0.719	1.013
no K10	0.65727	1.09e-11	1.034	2.204	1.363	281.8	0.000	1.034	

|Zenith| Band	#pts	rmserror	bias	rmserror	bias
+60 to +70	21213	0.44234	-0.04142	0.86911	-0.16031
+50 to +60	19667	0.39847	-0.00778	0.63142	0.01057
+40 to +50	15353	0.39828	0.00676	0.58992	0.05601
+30 to +40	12415	0.40274	0.02223	0.58093	0.06392
+20 to +30	10231	0.40334	0.01332	0.53541	0.05429
+10 to +20	9477	0.40560	0.03403	0.50201	0.06380
0 to +10	9151	0.39493	0.02127	0.52607	0.04228

Table 2. Dependence of regression coefficients on Sun zenith angle, with T_{field} = K10.

The same kind of analysis was performed in 20° latitude bands and the values of the resulting coefficients are presented in Table 3. A gradual increase in rms error can be seen as the latitude bands are further removed from the equator, except for the two extreme bands near the poles. The sum of c_1 and c_5 again shows remarkable constancy, except for the value of 0.914 at the equatorial band.

As was done in Table 2, the values of global coefficients for both TfieldSST with K10 and MCSST (green) are displayed for comparison. The use of these global coefficients to calculate the rms errors over the latitude bands reveals several outliers, whose large biases are highlighted in blue. The rms errors associated with these large biases, although still within the range of values of the other rms errors, are also relatively high. The MCSST global coefficients also predicted large biases (green) highlighted in blue.

From GUI results

Latitude Band	#pts	rmserror	bias	c_1	c_2	c_3	c_4	c_5	(c_1+c_5)
+50 to +70	7112	0.36861	-1.38e-12	0.375	0.980	0.314	102.2	0.626	1.001
+30 to +50	17404	0.39976	5.79e-13	0.513	1.191	0.626	139.8	0.497	1.010
+10 to +30	12910	0.3532	-6.44e-13	0.279	0.672	0.339	76.3	0.731	1.010
-10 to +10	4856	0.3521	-1.66e-13	0.242	0.795	0.240	63.9	0.672	0.914
-30 to -10	8947	0.34733	-1.81e-13	0.432	1.080	0.554	117.9	0.580	1.012
-50 to -30	9098	0.38322	-6.67e-13	0.514	1.054	0.612	140.0	0.501	1.015
-70 to -50	1468	0.36599	-2.07e-14	0.359	0.777	0.359	98.1	0.673	1.032

Using global coefficients

	#pts	rmserror	bias	c_1	c_2	c_3	c_4	c_5	
	61795	0.3894	3.05e-12	0.416	1.039	0.515	113.4	0.589	1.005
no K10		0.5305	6.62e-12	1.006	2.550	1.174	274.1	0.000	1.006

Latitude Band	#pts	rmserror	bias	rmserror	bias
+50 to +70	7112	0.37013	-0.02832	0.5056	-0.07924
+30 to +50	17404	0.40451	0.12566	0.51462	0.11358
+10 to +30	12910	0.36430	-0.01473	0.57448	-0.11061
-10 to +10	4856	0.38496	-0.11888	0.58587	-0.11294
-30 to -10	8947	0.34863	-0.07716	0.47300	0.00595
-50 to -30	9098	0.38908	-0.06359	0.49011	0.05780
-70 to -50	1468	0.37408	0.03549	0.47590	-0.00664

Table 3. Dependence of regression coefficients on latitude bands for Sun zenith angles from 0° to 53°, with T_{field} = K10.

Reduction of rms error as a function of Sun zenith angle bands

Present SST algorithms only consider Sun zenith angles from 0° to 53°. In an effort to quantify the improvement in accuracy afforded with the use of T_{field} as a separate predictor, Table 4 displays the rms errors predicted by MCSST (**Present**) and by the TfieldSST algorithm when the three characterizations of T_{field} are used. The percent reduction for each characterization is shown in green. The reduction in rms error in the zenith band from 0° to 53° is clearly seen from left to right, as the spatial resolution of T_{field} increases from Clim to K100 to K10. In addition, results for the zenith bands of 53° to 70° and 0° to 70° representing the missing part of the full swath are shown in the last 2 rows of Table 4. The rms error percent reduction over the 53° to 70° zenith band is quite remarkable as the rms error drops from 0.73712 down to 0.41298 for T_{field} = K10. The full swath zenith band of 0° to 70° shows a percent reduction of over 30% with K100 and even more with K10.

Zenith	Present	Tfield = Clim		Tfield = K100		Tfield = K10	
0° to 53°	0.53052	0.45077	15.0%	0.42551	19.8%	0.38937	26.6%
53° to 70°	0.73712	0.51667	29.9%	0.45622	38.1%	0.41298	43.9%
0° to 70°	0.65627	0.49989	23.8%	0.45507	30.7%	0.41053	37.4%

Table 4. Percent reduction in rms error over Sun zenith angle bands for the three characterizations of T_{field}.

Global Improvement of SST predictions against Buoy data

Another way to visualize the improvements over the 3 zenith bands in Table 4 is to produce worldmaps of SST_{Tfield}-T_{buoy} corresponding to these zenith bands side-by-side with worldmaps of SST_{MC}-T_{buoy}. Figures 6, 7 and 8 use the Blue to Red and Red to Blue method to plot results for 0° to 53°, 53° to 70° and 0° to 70°, respectively.
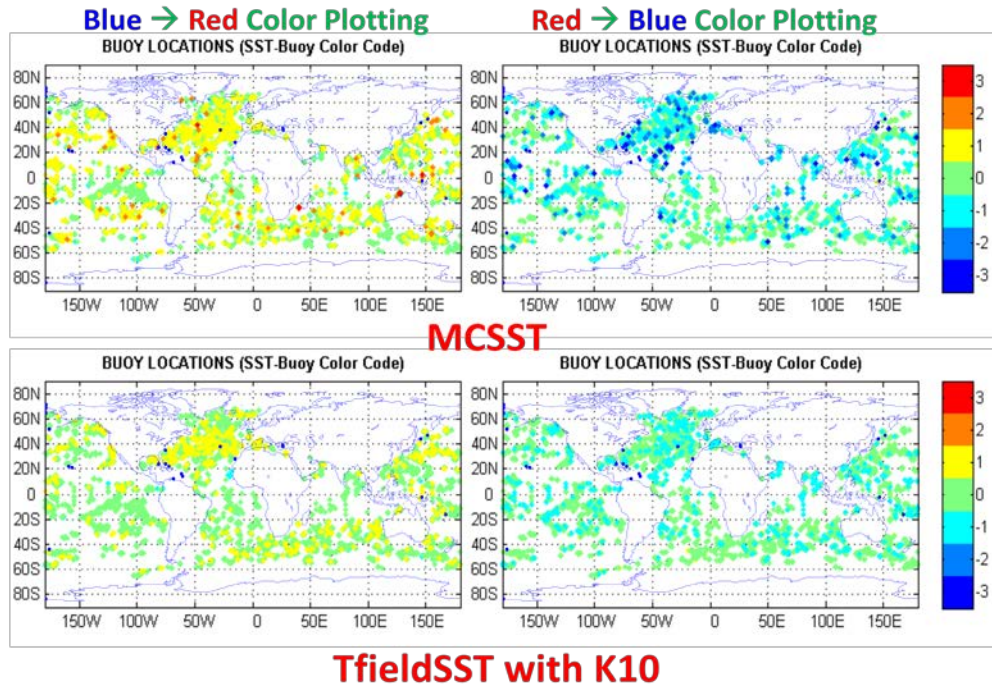

Figure 6. Global map of SST_{MC}-T_{buoy} and SST_{Tfield}-T_{buoy} for Sun zenith angles from 0° to 53°.
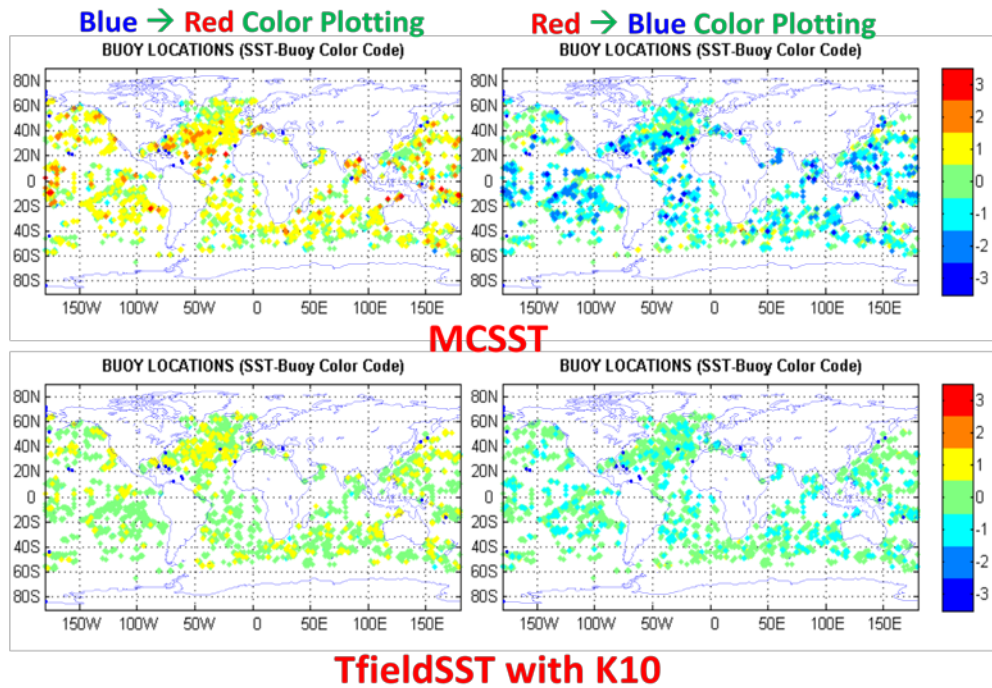
Figure 7. Global map of SST_{MC}-T_{buoy} and SST_{Tfield}-T_{buoy} for Sun zenith angles from 53° to 70°.

11

For all three figures, the top two worldmaps represent predictions from MCSST, while the bottom two worldmaps are the TfieldSST predictions using T_{field} = K10. The reduction in rms errors described previously is evidenced by the ability of the TfielSST algorithm to remove extreme differences that are present in MCSST results, both positive (towards the red) and negative (towards the blue). Note the impressive confinement of differences to the yellow and pale blue colors upon close inspection of the TfieldSST worldmaps.
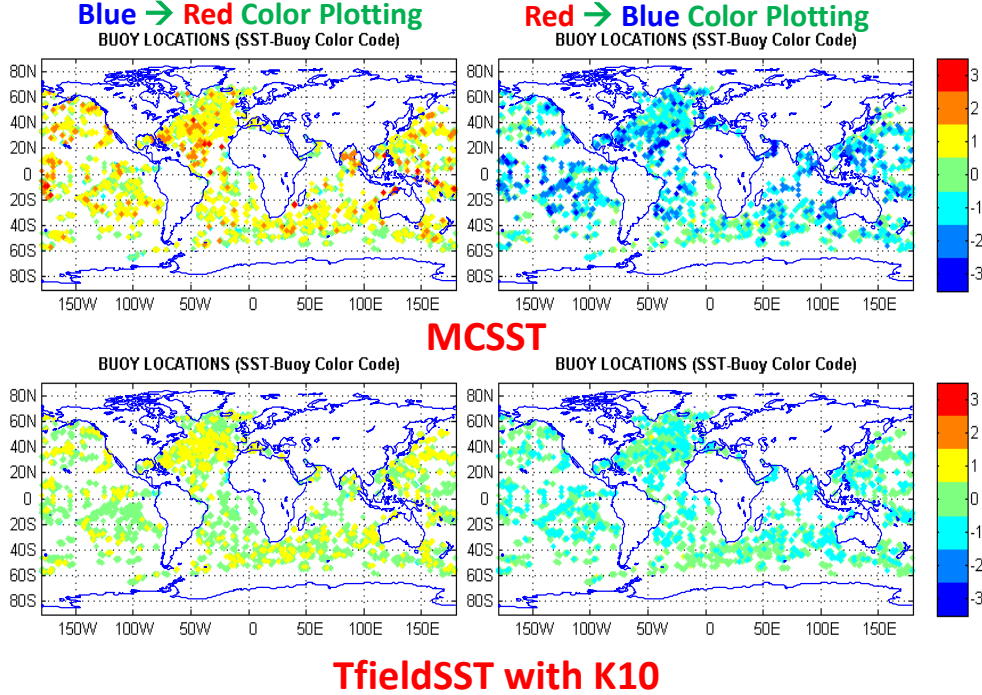
Figure 8. Global map of SST_{MC}-T_{buoy} and SST_{Tfield}-T_{buoy} for Sun zenith angles from 0° to 70°.

DISCUSSION

Although the ultimate goal is to predict sea surface temperature from satellite data, an investigation about using T_{field} as a separate predictor, and therefore combining satellite and previously collected buoy data, revealed a few salient features and shortcomings that bear discussing. In addition, we present an interpretation of the results that clearly shows the synergy in the combination of satellite and buoy data sets towards increasing the accuracy of SST predictions of sea surface temperature. The major differences between pre-filtering of the buoy data set and the use of higher spatial resolution versions of T_{field} are highlighted. Advantages of using T_{field} as a separate predictor are discussed, together with the ultimate goal of not using it by depending solely on satellite data.

Interpretation of Results
Equations for the two SST algorithms compared in this paper, MCSST and TfieldSST, are again displayed here for easy reference in this discussion:

$$SST_{MC} = c_1^{MC} T_{11} + c_2^{MC}(T_{11} - T_{12}) + c_3^{MC} S(T_{11} - T_{12}) - c_4^{MC}$$

$$SST_{Tfield} = c_1^{Tfield} T_{11} + c_2^{Tfield}(T_{11} - T_{12}) + c_3^{Tfield} S(T_{11} - T_{12}) - c_4^{Tfield} + c_5^{Tfield} T_{field}.$$

As shown in Results section, the sum of the T_{11} and T_{field} coefficients remained remarkably steady around the value of unity:

$$c_1^{Tfield} + c_5^{Tfield} = 1 \,.$$

In a search for other relationships that could possibly offer additional insight into an interpretation of results, a separate investigation of possible relationships between the ratios of TfieldSST coefficients and MCSST coefficients was performed. The results are shown in Tables 5 and 6, which differ only in the range of the Sun zenith angles. The values in red were obtained by dividing the TfieldSST coefficients 2, 3 and 4 by the first TfieldSST coefficient, c_1^{Tfield}.

Because of the general trends in coefficient values presented earlier and the relatively good agreement in values between the MCSST coefficients and these ratios, the following relationship suggests itself:

$$c_i^{MC} = c_i^{Tfield} / c_1^{Tfield}$$

or

$$c_i^{Tfield} = c_1^{Tfield} c_i^{MC} \,.$$

	c1	c2	c3	c4	c5	rmserror
MCSST	1.006	2.541	1.173	274.10	----	0.5058
TfieldSST	0.251	0.617	0.312	68.42	0.752	0.2869
	1.000	2.467	1.243	272.58	----	

Table 5. Comparison of coefficients of MCSST and TfieldSST, as well as their ratios .
Sun zenith angles are from 0° to 53°.

	c1	c2	c3	c4	c5	rmserror
MCSST	1.034	2.201	1.360	281.98	----	0.6382
TfieldSST	0.163	0.326	0.221	44.48	0.844	0.2987
	1.000	2.001	1.356	272.88	----	

Table 6. Same as Table 5, except that Sun zenith angles are from 0° to 70°.

Substituting this relationship into the equation for SST_{Tfield}:

$$SST_{Tfield} = c_1^{Tfield}\left[c_1^{MC}T_{11} + c_2^{MC}(T_{11} - T_{12}) + c_3^{MC}S(T_{11} - T_{12}) - c_4^{MC}\right] + c_5^{Tfield}T_{field}$$

Resulting in a relationship that clearly illustrates a linear weighted average between only two predictors:

$$SST_{Tfield} = c_1^{Tfield} SST_{MC} + \left(1 - c_1^{Tfield}\right)T_{field} \,.$$

The latter equation can be rewritten as a linear function of the difference between SST_{MC} and T_{field}:

$$SST_{Tfield} = T_{field} + c_1^{Tfield}\left(SST_{MC} - T_{field}\right).$$

A very simple and insightful interpretation now emerges. The TfieldSST algorithm is searching for a correction to the background (T_{field}) which is a linear function of the difference between observation (satellite data through SST_{MC}) and background (T_{field}). Of great relevance and guidance here is the established theory about the relationship between the

individual variances for this type of equation. If, for the sake of simplicity in the resulting relationships, the observation and background *errors* are now assumed to be uncorrelated [5]:

$$\sigma^2_{SST_{Tfield}} = \left(1 - c_1^{Tfield}\right)^2 \sigma^2_{Tfield} + \left(c_1^{Tfield}\right)^2 \sigma^2_{SST_{MC}}.$$

The optimal c_1^{Tfield} that minimizes the error variance $\sigma^2_{SST_{Tfield}}$ is obtained by setting the derivative with respect to c_1^{Tfield} equal to 0.0, resulting in:

$$c_1^{Tfield} = \frac{\sigma^2_{Tfield}}{\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}} \quad \text{and} \quad 1 - c_1^{Tfield} = c_5^{Tfield} = \frac{\sigma^2_{SST_{MC}}}{\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}}.$$

When this optimal value is used, a simple relationship between the three variances emerges:

$$\sigma^2_{SST_{Tfield}} = \left(\frac{\sigma^2_{SST_{MC}}}{\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}}\right)^2 \sigma^2_{Tfield} + \left(\frac{\sigma^2_{Tfield}}{\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}}\right)^2 \sigma^2_{SST_{MC}}$$

$$\sigma^2_{SST_{Tfield}} = \frac{\sigma^2_{Tfield}\sigma^2_{SST_{MC}}}{\left(\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}\right)^2}\left(\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}\right) = \frac{\sigma^2_{Tfield}\sigma^2_{SST_{MC}}}{\sigma^2_{Tfield} + \sigma^2_{SST_{MC}}}$$

or

$$\frac{1}{\sigma^2_{SST_{Tfield}}} = \frac{1}{\sigma^2_{Tfield}} + \frac{1}{\sigma^2_{SST_{MC}}}.$$

Note that the resulting variance of SST_{Tfield} will **always be smaller** than either variance, due to the "resistors in parallel" characteristic of this last relationship. Clearly, should the variances of T_{field} and SST_{MC} be equal, the resulting variance of SST_{Tfield} will be half of each, resulting in its rms error being a factor of $\sqrt{2}/2$ smaller than their rms errors. For example, should the rms errors of T_{field} and SST_{MC} each be equal to 0.5, the resulting rms error of SST_{Tfield} would be 0.3535, a substantial reduction that is in the neighborhood of TfieldSST rms errors reported in this article. This interpretation and its consequences on the variance of SST_{Tfield} correctly and neatly encapsulates the nature, magnitude and overall trend of the results presented earlier.

A couple of other interesting relationships which relate the rms errors between SST_{Tfield} and its predictors SST_{MC} and T_{field} immediately follow from the above equations:

$$\sigma^2_{SST_{Tfield}} = c_1^{Tfield}\sigma^2_{SST_{MC}} \quad \Rightarrow \quad \sigma_{SST_{Tfield}} = \sqrt{c_1^{Tfield}}\,\sigma_{SST_{MC}}$$

and

$$\sigma^2_{SST_{Tfield}} = c_5^{Tfield}\sigma^2_{Tfield} \quad \Rightarrow \quad \sigma_{SST_{Tfield}} = \sqrt{c_5^{Tfield}}\,\sigma_{Tfield}.$$

From this, a ballpark estimate of the percent reduction in SST_{MC} rms error is conveniently given by $100\left(1 - \sqrt{c_1^{Tfield}}\right)$. Introduction of a small correlation between SST_{MC} and T_{field} errors with respect to T_{buoy} will introduce minimal changes in these ballpark estimates.

14

The global rms error for the Clim, K100 and K10 fields were estimated to be about 1.0, 0.7 and 0.55, respectively [2]. Substituting the rms error for SST_{MC} from Table 5, we arrive at SST_{Tfield} rms error values of 0.450, 0.409 and 0.372 for zenith angles from 0° to 53° .Using the SST_{MC} rms error value from Table 6, SST_{Tfield} rms error values of 0.538, 0.472 and 0.4166 are obtained when zenith angles from 0° to 70° are considered. These values again are consistent with the results concerning the three different characterizations of T_{field} were used: Clim, K100 and K10.

With the above explanation of the reason why the TfieldSST rms error turns out to be substantially smaller than either the SST_{MC} or T_{field} rms errors, we now turn to a discussion of the achievability of the 0.3°K threshold for rms error from any SST algorithm. Figure 9 displays several curves which represent the desired rms error as a function of both SST_{MC} and T_{field} rms errors. The bold red line represents the desired 0.3°K threshold for rms error. The black circle corresponds to a Desired rms error of 0.3535 just discussed in the example above.
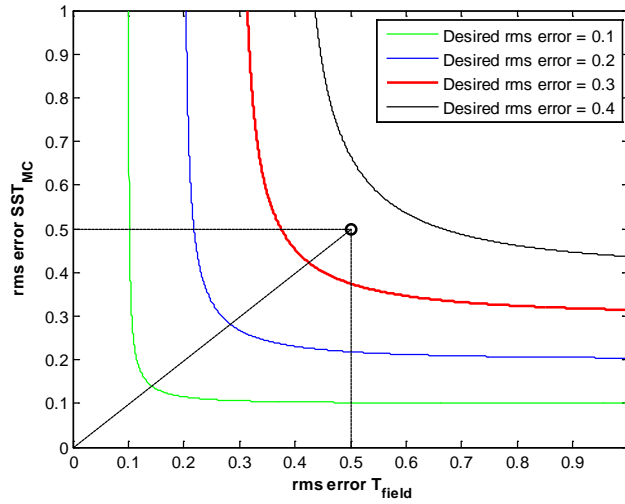

Figure 9. Curves of Desired rms errors.

The diagonal dashed line represents the path taken to reduce the Desired rms error with a common reduction factor for both SST_{MC} and T_{field}. Its intersection with the 0.3°K threshold curve (red) requires that each rms error be reduced to a value of $0.3 / (\sqrt{2}/2)$ or 0.425, a reduction of about 15% from existing rms errors of roughly 0.5 for both SST_{MC} and T_{field}. From the relationship between variances given earlier, the path along the vertical dashed line would require a reduction in the SST_{MC} rms error to a value of 0.375, while keeping the T_{field} rms error at 0.5. Due to the symmetric nature of the variances relationship that is also illustrated in Figure 9, the path along the horizontal dashed line would require a reduction of T_{field} rms error to a value of about 0.375 as well, while keeping the SST_{MC} rms error at 0.5. This represents a reduction of about 25% from existing T_{field} rms error of roughly 0.5.

The latter path seems to be the easiest and quickest way to achieve the 0.3°K threshold, as efforts are now under way at NAVOCEANO [2] to develop a K2 characterization of T_{field}, with a spatial resolution of 2 km. A decrease in T_{field} rms error is expected with this new product, due to its higher spatial resolution.

Interestingly, if one is willing to sacrifice a small percentage of the filtered buoy data points, the 0.3°K threshold can be easily achieved presently without any improvements to the present SST_{MC} or T_{field} rms errors. More aggressive pre-filtering of the buoy data is analogous to artificially increasing the accuracy of T_{field}, at the cost of eliminating some buoy data points. We therefore now present the results of a tradeoff study between pre-filtering aggressiveness, spatial resolution of T_{field}, range of zenith angles and resulting number of filtered buoy data points available for ingestion into

15

the TfieldSST algorithm, with the ultimate goal of achieving the 0.3°K threshold in Figures 10, 11 and 12, where runs were made for values of $| T_{field} - T_{buoy} | = 0.1, 0.2, 0.5, 1.0, 2.0, 3.0, 4.0$ and 5.0.

These figures only differ with respect to zenith bands: 0° to 53°, 53° to 70° and full swath 0° to 70°. For each figure, the horizontal axis represents the aggressiveness of the pre-filtering, increasing from right to left. The plot on the left of each figure is the rms error resulting from SST_{MC} (bold red), and the three characterizations of T_{field}. The plot on the right of each figure displays two variables: the percent of buoy data left over after the pre-filtering (solid lines), and the percent contribution that T_{field} makes to the TfieldSST algorithm, expressed as c_5^{Tfield} in percent units. The 0.3°K threshold is represented by the horizontal dashed black line. The vertical dashed line is a visual estimate of the required $| T_{field} - T_{buoy} |$ required to achieve this goal, and is drawn for a value of 0.7°K. A noteworthy feature is that the K100 and K10 characterizations of T_{field} seem to converge at that value, while Clim requires more aggressive filtering to reach the 0.3°K threshold.
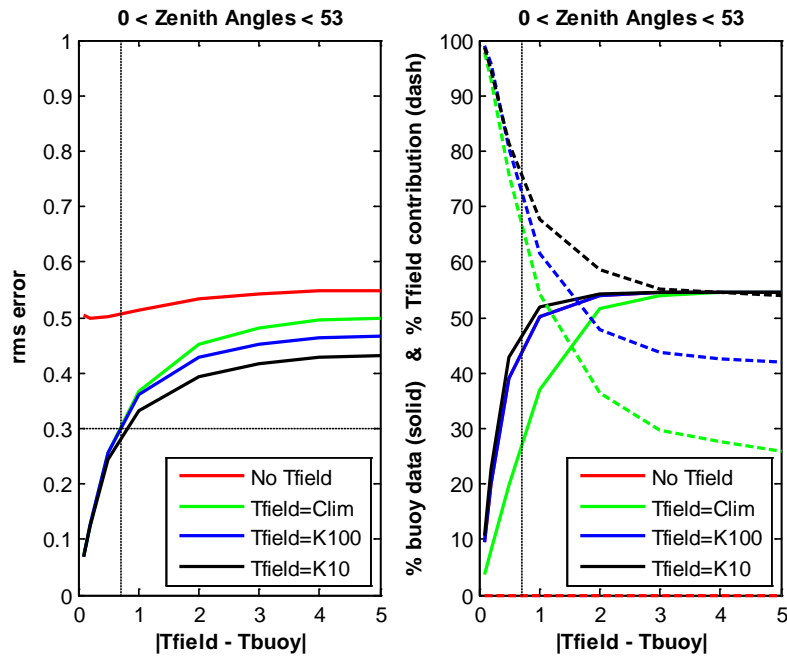

Figure 10. Plots of rms error, % of buoy data points filtered and % of T_{field} contribution to TfieldSST, $(100\, c_5^{Tfield})$, for Sun zenith angles from 0° to 53°.

It will be recalled from the tabular data presented in the Results section, that overall trends showed that the value of c_5^{Tfield} depended on the aggressiveness of the pre-filtering of the buoy dataset against T_{field}, the resolution of the T_{field} data (Clim, K100, K10), as well as the Sun zenith angle. In general, for a given level of pre-filtering, the value of c_5^{Tfield} increased and the rms error decreases as the spatial resolution of the T_{field} data increased from Clim to K100 to K10. This indicates that the TfieldSST algorithm automatically adjusts the mix of satellite and T_{field} data to produce rather drastic increases in accuracy of about 30% (equivalent to the multiplicative factor of $\sqrt{2}/2$ found previously) . Results also revealed that the value of c_5^{Tfield} increases with zenith angle, showing that present regression algorithms that use only satellite data are, by themselves, less capable and less accurate as the zenith angle increases.
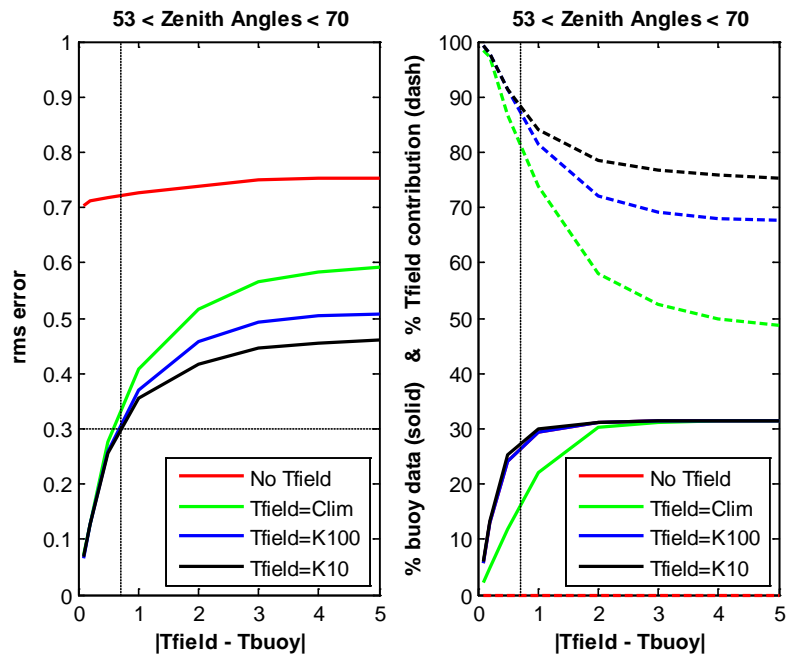
Figure 11. Plots of rms error, % of buoy data points filtered and % of T_{field} contribution to TfieldSST, ($100\, c_5^{Tfield}$), for Sun zenith angles from 53° to 70°.
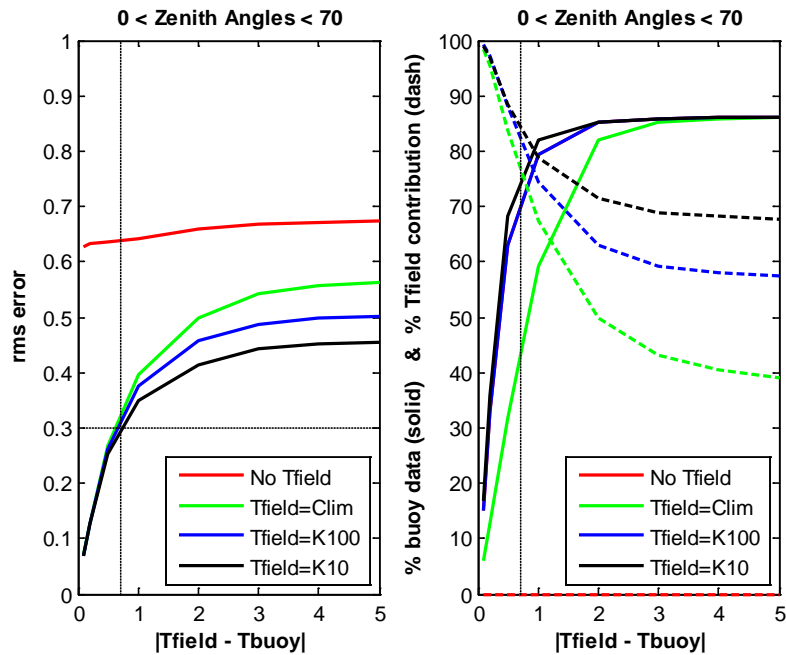

Figure 12. Plots of rms error, % of buoy data points filtered and % of T_{field} contribution to TfieldSST, ($100\, c_5^{Tfield}$), for Sun zenith angles from 0° to 70°.

All of these gradual trends are reflected in Figures 10 through 12 and therefore reinforce our assumption that TfieldSST is basically a linear weighting of SST_{MC} (satellite) and T_{field} (background) data. It is only for high values of $|T_{field} - T_{buoy}|$, corresponding to less aggressive pre-filtering, that differences in SST_{Tfield} rms errors reveal themselves between each characterization of T_{field}. This can be seen on all the plots of rms error vs $|T_{field} - T_{buoy}|$ in Figures 10 through 12. Unlike the T_{field} rms errors, the SST_{MC} rms error shown in bold red is relatively insensitive to the aggressiveness in pre-filtering, for all zenith bands shown in Figures 10 through 12. For the 53° to 70° band of zenith angles shown in Figure 11, the rms error averages about 0.725, for a combined full swath rms error average of about 0.650 shown in Figure 12. The gradual decrease in SST_{Tfield} rms error from right to left for all three characterizations of T_{field} is consistent with the artificial reduction of T_{field} rms error achieved through gradually more aggressive pre-filtering.

The plots on the right of each figure are more pertinent to our tradeoff analysis, however, where the same vertical dashed line with a value of 0.7°K of $|T_{field} - T_{buoy}|$ required to achieve the goal of 0.3°K threshold is replicated. It should be mentioned that Clim is also used for MCSST pre-filtering of buoy data, explaining why there is no solid red curve and that the green solid curve applies for MCSST pre-filtering. In addition, since there is no T_{field} term in MCSST, the red dashed curve for MCSST represents a 0.0 value for $100\,c_5^{Tfield}$.

Figure 10 for the 0° to 53° zenith band shows that only about 55% of the data buoy points filtered through at the less aggressive range of $|T_{field} - T_{buoy}|$ for all characterizations of T_{field}. In order for the TfieldSST algorithm to achieve the 0.3°K threshold required with value 0.7°K of $|T_{field} - T_{buoy}|$, only about 25% of the buoy data points for Clim are left to be ingested into the algorithm, while both K100 and K10 only suffer a more modest decline to 42% and 46%, respectively.

Figure 11 for the 53° to 70° zenith band shows that only about 30% of the data buoy points filtered through at the less aggressive range of $|T_{field} - T_{buoy}|$ for all characterizations of T_{field}. In order for the TfieldSST algorithm to achieve the 0.3°K threshold required with value 0.7°K of $|T_{field} - T_{buoy}|$, only about 15% of the buoy data points for Clim are left to be ingested into the algorithm, while both K100 and K10 only suffer a more modest decline to 26% and 28%, respectively.

Finally, Figure 12 for the 0° to 70° zenith band shows that about 85% of the data buoy points filtered through at the less aggressive range of $|T_{field} - T_{buoy}|$ for all characterizations of T_{field}. In order for the TfieldSST algorithm to achieve the 0.3°K threshold required with value 0.7°K of $|T_{field} - T_{buoy}|$, only about 45% of the buoy data points for Clim are left to be ingested into the algorithm, while both K100 and K10 only suffer a more modest decline to 68% and 74%, respectively.

Summarizing, for the present 0° to 53° range of zenith angles considered in SST algorithms, a reduction of about only 9% in filtered buoy data points results in a drop in rms error from 0.5 for MCSST to the goal of 0.3°K when $T_{field} = K10$ is used in the TfieldSST algorithm. Similarly, for the 53° to 70° range of zenith angles presently ignored in SST algorithms, a reduction of about only 2% in filtered buoy data points results in a drop in rms error from 0.75 for MCSST to the goal of 0.3°K for SST_{Tfield} using $T_{field} = K10$. Finally, for the full swath 0° to 70° range of zenith angles, a reduction of about only 11% in filtered buoy data points results in a drop in rms error from 0.65 for SST_{MC} to the goal of 0.3°K for SST_{Tfield} using $T_{field} = K10$.

It should be emphasized at this point that only about 55% of the buoy data points are being presently used to achieve a rms error of about 0.5 considering only the 0° to 53° range of zenith angles, while our results now provide about 75% of the buoy data points over the full swath from 0° to 70° range of zenith angles with an overall rms error of 0.3°K, attaining the seemingly elusive goal of SST algorithm development.

Our assumption about the linear weighting of SST_{MC} and T_{field} is also validated by the dashed curves on the rightmost plots in Figures 10 through 12, which represent the values of $100\ c_5^{Tfield}$ as a function of $|\ T_{field} - T_{buoy}|$. Recalling that the linear weighting resulted in the relationship between the coefficient of T_{field} and the rms errors of SST_{MC} and T_{field}:

$$c_5^{Tfield} = \frac{\sigma_{SST_{MC}}^2}{\sigma_{Tfield}^2 + \sigma_{SST_{MC}}^2}\ ,$$

and that 1.0, 0.7 and 0.55 were the rms errors quoted earlier for Clim, K100 and K10, respectively, we consider the least aggressive values of 5.0 for $|\ T_{field} - T_{buoy}|$. From Figure 10, we take the SST_{MC} rms error to be 0.55. Combining with the three characterizations of T_{field}, we arrive at values for $100\,c_5^{Tfield}$ of 23.22%, 38.2% and 50.0% for Clim, K100 and K10, all representing a difference of less 5% compared to values from the plot. From Figure 11, we take the SST_{MC} rms error to be 0.75 when $|\ T_{field} - T_{buoy}| = 5°K$. Combining with the three characterizations of T_{field}, we arrive at values for 100 c_5^{Tfield} of 36.0%, 53.4% and 65.0% for Clim, K100 and K10, all representing a difference of less 15% compared to values from the plot. From Figure 12, we take the SST_{MC} rms error to be 0.67 when $|\ T_{field} - T_{buoy}| = 5.0$. Combining with the three characterizations of T_{field}, we arrive at values for $100\,c_5^{Tfield}$ of 30.9%, 47.8% and 59.7% for Clim, K100 and K10, all representing a difference of less 10% compared to values from the plot.

The gradual trend for $100\,c_5^{Tfield}$ for all zenith bands is to increase to 100% as the aggressiveness of pre-filtering increases. The smoothness of the dashed curves are evidence of the inherent stability of the TfiedSST algorithm. This trend to a value of 100% is revealed by the equation for c_5^{Tfield} above by setting the T_{field} rms error to 0.0, a condition which results from the artificially induced reduction in rms error due to more aggressive pre-filtering. Discrepancies between calculated values of $100\,c_5^{Tfield}$ and those read from the plots in Figures 10 through 12, could have arisen from the fact that all the runs were performed by allowing all of the MCSST coefficients to interplay with the T_{field} coefficient. The MCSST algorithm was not inserted as a single predictor by itself during our investigation, thereby slightly diminishing the dichotomy between SST_{MC} and T_{field} needed to derive the above relationships between variances. It is also probable that some small correlation exists between errors in SST_{MC} and T_{field}, which would slightly affect the predictions of the values for $100\,c_5^{Tfield}$. Overall, however, the interpretation of the TfieldSST algorithm as a simple linear combination of satellite-derived data through SST_{MC} and background observation through T_{field} resulted in predictions for the values of their coefficients that closely mimic both the trends and magnitude of the results of our investigations.

CONCLUSION

A visual side-by-side comparison of the characteristics of two scatterplots was first performed to look for clues that would lead to an overall reduction in SST algorithm rms error. The horizontal axis on both represented the differences between T_{11} and T_{12}. The vertical axis were similar, as one involved the differences between buoy temperatures in Kelvin units and the brightness temperature T_{11}, while the other involved the differences between SST predictions in Kelvin units and the brightness temperature T_{11}. The presence of T_{field} in the NLSST algorithm somehow generated a slight randomness that was present in the scatterplot involving the buoy data, and not in the MCSST-generated scatterplot. This behavior suggested the use of T_{field} as a separate predictor added to the MCSST algorithm, resulting in an algorithm appropriately named TfieldSST.

Analysis of the resulting coefficients in the TfieldSST algorithm showed a remarkable constancy in the sum of c_1^{Tfield} and c_5^{Tfield}, the coefficients of T_{11} and T_{field}, respectively, gravitating around a value of 1.0 regardless of the pre-filtering aggressiveness or zenith and latitude bands. This constancy also surfaced for three characterizations of T_{field} that were used in this study, in order of increasing spatial resolution: Clim, K100 and K10. This was accompanied by substantial reductions in rms error of well over 30% for certain zenith and latitude bands. An additional finding was that a general relationship between coefficients of MCSST and TfieldSST seemed to be present, suggesting the simple relationship: $c_i^{Tfield} = c_1^{Tfield} c_i^{MC}$, where $i = 1, 4$.

When these two relationships were substituted into the TfieldSST algorithm, it took on the form of a simple linear weighted average of two separate predictors: SST_{MC} predictions derived from satellite observations, and a background first guess temperature field, T_{field}. Well-known established relationships between the coefficients and rms errors for this type of equation provided major insight into how to proceed towards the goal of attaining an accuracy of 0.3°K accuracy. The "resistors in parallel" relationship between the variances of SST_{Tfield}, SST_{MC} and T_{field} ensured that the rms error value of SST_{Tfield} will always be smaller than the SST_{MC} and T_{field} rms errors.

In particular, it was shown that the 0.3°K threshold could be attained at the very minor cost of sacrificing a small percentage of data buoy points through more aggressive pre-filtering. However, because our method is valid over the full swath from 0° to 70° zenith angles, over 75% of the original data buoy points are now being used with an accuracy of 0.3°K. This is compared to the present 0° to 53° zenith angles, for which about 50% of the original data buoy points are now being used to obtain a rms error of about 0.5 for daytime scenarios. There is therefore a need for pre-filtering standards for consistent comparison of results because aggressive filtering results in lower rms error in the TfieldSST algorithm, much more so than in the MCSST algorithm.

Future improvements include efforts to reduce rms errors in both predictors in TfieldSST: SST_{MC} and T_{field} accuracy. The latter is expected to result from a NAVOCEANO effort to develop a 2km spatial resolution K2 characterization of T_{field}. Increasing the accuracy of SST_{MC} at large zenith angles beyond 53° should be achievable by including sea surface emissivity and roughness.

The linear weighted average form of TfieldSST guarantees increased accuracy in all existing SST algorithms, daytime and nighttime, by simply adding T_{field} as a predictor, due to the "resistors in parallel" relationship between variances. In effect, existing SST algorithms can be considered as "fine-tuning" the first guess field, T_{field}. It should be pointed out that K100 and K10 are also described by the community as satellite-derived [3] and can be considered as satellite data from an earlier date. The multichannel concept and physics behind present SST algorithms remains intact since they act as separate predictors in TfieldSST. Also of note is that the T_{field} information has no information about zenith angles involved, and daily satellite data is still crucial in obtaining much better overall accuracy.

It should be noted that the predictors SST_{MC} and Tfield can either be input together in °Celsius or °Kelvin units, depending on what units are desired for SST_{Tfield} predictions, avoiding the confusion in temperature units present in NLSST.

Another salient feature of TfieldSST is that it reveals the strong and weak domains of present SST algorithms. The relative values of of c_1^{Tfield} and c_5^{Tfield}, because they have been shown to be related to combinations of the respective rms errors of SST_{MC} and T_{field}, helps to identify zenith and latitude bands where satellite data contribution through SST_{MC} is insufficient by itself right now to obtain the desired 0.3°K accuracy.

A fruitful analogy that parallels our findings is now presented that neatly summarizes our results, which are very similar to the well-known noise reduction achieved by frame averaging imagery captured by a camera. If the noise is assumed to

be uncorrelated from frame to frame and of equal variance, the addition of only two frames reduces the variance by half or, equivalently, reducing the noise rms error by 30%, a substantial result that was shown earlier during our discussion of linear weighting between SST_{MC} and T_{field}. If this "imagery" now represents the global distribution of T_{buoy}, which is the information that we are trying to extract, then SST_{MC} and T_{field} are equivalent to two "noisy" frames of T_{buoy}, but with different variances. As we found, the result of adding them together with linear weighting is to automatically reduce the variance of SST_{Tfield}. The TfieldSST algorithm finds the optimal combination of these two noisy frames that will provide the smallest resulting variance, thereby providing a less noisy image of T_{buoy}.

REFERENCES

[1]. Walton, C., Pichel, W. and Sapper, J., "The development and operational application of nonlinear algorithms for the measurement of sea surface temperatgures with the NOAA polar-orbiting environmental satellites", *J. Geophys. Res.*, Vol. **103**, No. C12, pp. 27,999 – 28,012, (November 15, 1998).

[2]. Cayula, J.F and May, D., *private communications*.

[3]. Martin, M., et al., "Group for High Resolution Sea Surface Temperature (GHSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE)", *Deep-Sea Research* II 77-80, pp. 21-30, (2012).

[4]. http://en.wikipedia.org/wiki/Multicollinearity

[5] Bouttier, F., and Courtier, P., "Data assimilation concepts and methods", http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts6.html#963809