

Validation of GOES-R Satellite Land Surface Temperature Algorithm Using SURFRAD Ground Measurements and Statistical Estimates of Error Properties

Yunyue Yu, Dan Tarpley, Jeffrey L. Privette, Lawrence E. Flynn, Hui Xu, Ming Chen, Konstantin Y. Vinnikov, Donglian Sun, and Yuhong Tian

Abstract—Validation of satellite land surface temperature (LST) is a challenge because of spectral, spatial, and temporal variabilities of land surface emissivity. Highly accurate *in situ* LST measurements are required for validating satellite LST products but are very hard to obtain, except at discrete points or for very short time periods (e.g., during field campaigns). To compare these field-measured point data with moderate-resolution (~ 1 km) satellite products requires a scaling process that can introduce errors that ultimately exceed those in the satellite-derived LST products whose validation is sought. This paper presents a new method of validating the Geostationary Operational Environmental Satellite (GOES) R-Series (GOES-R) Advanced Baseline Imager (ABI) LST algorithm. It considers the error structures of both ground and satellite data sets. The method applies a linear fitting model to the satellite data and coregistered “match-up” ground data for estimating the precisions of both data sets. In this paper, GOES-8 Imager data were used as a proxy of the GOES-R ABI data for the satellite LST derivation. The *in situ* data set was obtained from the National Oceanic and Atmospheric Administration’s SURFACE RADIATION (SURFRAD) budget network using a stringent match-up process. The data cover one year of GOES-8 Imager observations over six SURFRAD sites. For each site, more than 1000 cloud-free match-up data pairs were obtained for day and night to ensure statistical significance. The average precision over all six sites was found to be 1.58 K, as compared to the GOES-R LST required precision of 2.3 K. The least precise comparison at an individual SURFRAD site was 1.8 K. The conclusion is that, for

these ground truth sites, the GOES-R LST algorithm meets the specifications and that an upper boundary on the precision of the satellite LSTs can be determined.

Index Terms—Algorithm evaluation, land surface temperature (LST), satellite measurement, SURFACE RADIATION (SURFRAD).

I. INTRODUCTION

IN THE DEVELOPMENT and use of satellite land surface temperature (LST) retrieval algorithms, validation is crucial yet difficult. Validation provides the quantitative uncertainty information required for the proper use and application of the product. No algorithm or product would be widely accepted without performing thorough calibration and validation. Traditionally, satellite LST validation is performed by comparing the satellite-derived LST to ground, aircraft, or other satellite LST estimates; both real and simulated satellite and ground data have been used. For instance, Wan *et al.* [1] performed direct and indirect validations of the LST product retrieved from Earth Observation System MODerate-resolution Imaging Spectroradiometer (MODIS) data using ground data collected from several field campaigns. Coll *et al.* [2] conducted a field campaign over a large, flat, and homogeneous rice crop area for validation of LST products derived from MODIS and European Space Agency Environmental Satellite Advanced Along-Track Scanning Radiometer data. Yu *et al.* [3] applied their evaluation results to the LST algorithms for Visible and Infrared Image Radiometer Suite of the National Polar-orbiting Operational Environmental Satellite System using a comprehensive simulation data set and MODIS data. Pinheiro *et al.* [4] validated nonnadir Advanced Very High Resolution Radiometer LST estimates over Africa using field measurements combined with an angular emission model. Vinnikov *et al.* [5] evaluated the satellite LSTs from the LST diurnal variation feature derived from Geostationary Operational Environmental Satellite (GOES) Imager data.

There are many challenges in such direct comparisons of LST algorithms and products, including the following.

- 1) The land surface is typically heterogeneous (both in temperature and emissivity) over satellite pixel areas (e.g., ~ 1 km), while *in situ* LST measurements are usually collected over significantly smaller and more homogeneous areas (e.g., ~ 0.01 km).

Manuscript received December 9, 2010; revised March 11, 2011 and June 20, 2011; accepted June 27, 2011. Date of publication August 22, 2011; date of current version February 24, 2012. This work was supported by the Application Working Group, Geostationary Operational Environmental Satellite R-Series, National Oceanic and Atmospheric Administration.

Y. Yu and L. E. Flynn are with the Center for Satellite Applications and Research, National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Camp Springs, MD 20746 USA (e-mail: Yunyue.Yu@noaa.gov; Lawrence.E.Flynn@noaa.gov).

D. Tarpley is with Short and Associates, Camp Springs, MD 20746 USA (e-mail: Dan.Tarpley@noaa.gov).

J. L. Privette is with the National Climatic Data Center, National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Asheville, NC 28801 USA (e-mail: jeff.privette@noaa.gov).

H. Xu, M. Chen, and Y. Tian are with I. M. Systems Group, Inc., Camp Springs, MD 20746 USA (e-mail: Hui.Xu@noaa.gov; Ming.Chen@noaa.gov; Yuhong.Tian@noaa.gov).

K. Y. Vinnikov is with the Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742 USA (e-mail: kostya@atmos.umd.edu).

D. Sun is with the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030 USA (e-mail: dsun@gmu.edu).

Digital Object Identifier 10.1109/TGRS.2011.2162338

- 2) Navigation errors cause the ground truth site to move from place to place within the coincident pixel, and in 1%–2% of pixels, the ground site may be outside the “coincident” pixel. A navigation uncertainty is a significant source of imprecision.
- 3) Accurate fine-resolution land surface emissivity data are needed but hard to obtain.
- 4) The rate of LST change is usually high, so the time differences between satellite LST and ground measurement must be relatively small.
- 5) There are few field sites where *in situ* LST data are routinely or episodically measured.
- 6) Cloud contamination in satellite data may have significant negative impacts on the validation process.
- 7) Angular anisotropy (directional variability) of apparent surface emissivity and temperature has significant impact on the LST retrieval.

For these reasons, collecting and processing highly accurate ground measurements that match the satellite LST measurements can be a tedious and costly task.

Given these challenges, it is important to quantify the uncertainty (accuracy and precision) of the properly scaled independent “validation data set” as part of the validation process. Flynn [6] described a statistical method for validating satellite sounding products. Instead of directly comparing the satellite-derived and *in situ* measurement data (e.g., scatterplots with a reference line and histogram of difference), he explored procedures for simultaneously estimating possible errors from both the satellite data and ground measurements. His approach can help quantify validation results and improve their interpretation.

In this paper, we applied Flynn’s method to estimate errors in LST derived from the U.S. GOES Imager data. Our goal is to evaluate the baseline LST algorithm for the new Advanced Baseline Imager (ABI) instrument that will fly on a new generation of GOES satellites, i.e., the GOES R-Series (GOES-R) [7]. In the following section, we provide details of the data sets used in this study. Section III gives the theoretical fundamentals and derived equations. We then show the results in Section IV, followed by a discussion of results in Section V. Finally, we provide some concluding remarks in Section VI.

II. DATA SETS

Two data sets were used in this study: SURFACE RADiation (SURFRAD) ground measurements and GOES-8 Imager data. We created a set of coregistered “match-up” LST data derived from GOES-8 Imager and ground measurements from the SURFRAD budget network stations. We then used Flynn’s method to evaluate satellite measurements in relation to *in situ* measurements and to estimate the errors and consistency of the satellite LSTs under a variety of scenarios.

A. SURFRAD Data

The SURFRAD network has been operational in the U.S. since 1995. It provides high-quality *in situ* measurements of upwelling and downwelling radiative fluxes, along with other meteorological parameters [8]. In this paper, we used one year

TABLE I
LOCATION OF THE SIX SURFRAD SITES

Site No.	Site Location	LAT, LONG	Surface Type [#]
1	Pennsylvania State University, PA	40.72N, -77.93W	Mixed Forest
2	Bondville, IL	40.05N, -88.37W	Crop Land
3	Goodwin Creek, MS	34.25N, -89.87W	Evergreen Needle Leaf Forest
4	Fort Peck, MT	48.31N, -105.10W	Grass Land
5	Boulder, CO	40.13N, -105.24W	Crop Land
6	Desert Rock, NV	36.63N, -116.02 W	Open Shrub Land

[#]: UMD land surface type.

(2001) of SURFRAD data over six sites, as described in Table I. Surface-type information for the sites from the University of Maryland land classification data set [9] is also provided in the table.

The flux data were measured by a radiometer 10 m above ground level. It collects a sample every 3 min in a spectral window from 3 to 50 μm . These data are used to estimate broad-band flux using predetermined laboratory radiometer calibration factors. A detailed description of the SURFRAD network and associated instrumentation can be found in [8] and [10].

The SURFRAD LST estimates were determined using the Stefan–Boltzmann law. The broad-band surface emissivity used for such estimation for each SURFRAD site was based on a linear regression form provided by Wang *et al.* [11] as

$$\varepsilon_w = 0.2122\varepsilon_{29} + 0.3859\varepsilon_{31} + 0.4029\varepsilon_{32} \quad (1)$$

where ε_w is the broad-band emissivity; ε_{29} , ε_{31} , and ε_{32} are the narrow-band emissivities of MODIS band 29 (8.3 μm), 31 (10.8 μm), and 32 (12.1 μm), respectively, which are available through the MODIS monthly emissivity data set.

B. GOES-8 Imager Data

We use GOES-8 Imager [12], [13] data as a proxy for ABI data, because the Imager has two “split window” channels in the thermal infrared spectrum, similar to those of the ABI. The GOES-8 Imager data were obtained from the proxy data team of GOES-R Algorithm Working Group (AWG). The GOES-8 Imager LST values were calculated using a split window algorithm [15], which has recently been developed by the land team of the GOES-R AWG. This study was designed to validate this algorithm. Table II gives the algorithm formula and the coefficients used for the GOES-8 LST calculation.

C. Data Match-Up Process

A data coregistration (“Match-up”) procedure for both time and location must be performed for the satellite and ground data before they can be compared. The GOES-8 Imager data have a spatial resolution of 4 km at nadir and a temporal resolution of 15 min. The LST product in this study was generated at a 30-min temporal interval. The SURFRAD data are effectively “spot” measurements with a temporal resolution of 3 min. We selected the imager pixels whose centroids were spatially nearest to the SURFRAD locations. In the time domain, we used

TABLE II
GOES-R LST ALGORITHM COEFFICIENTS APPLIED
ON THE GOES-8 IMAGER DATA

	C	A1	A2	A3	D
Daytime dry	35.022546	1.018212	1.263787	-39.387858	0.609744
Daytime moist	27.913362	1.026320	1.990878	-35.758536	0.421895
Nighttime dry	36.160667	1.012895	1.022203	-38.909505	0.669541
Nighttime moist	45.100015	0.962238	2.444521	-34.555664	0.453345

Note :

- 1) Algorithm formula: $T_s = C + A_1 T_{11} + T_2(T_{11} - T_{12}) + A_3 \epsilon + D(T_{11} - T_{12})(\sec \theta - 1)$ where T_{11} and T_{12} are the channel 4 and 5 brightness temperatures (deg K), respectively; θ is the view zenith angle, T_s is the retrieved LST (deg K).
- 2) daytime/nighttime is when solar zenith angle is less/greater than 85° than.
- 3) dry/moist refers atmospheric condition when the total column water is less/greater than 2.0 g/cm^2 .

only the SURFRAD values that were the closest to the GOES-8 measurements; therefore, the maximum temporal difference between the SURFRAD and the satellite measurements was less than 2 min.

Next, we applied a more stringent cloud filtering process compared to our previous study [14] to remove all cloudy data from the match-up data sets. We performed a traditional cloud filtering method [15], [16] that detects cloudy conditions using the following: 1) threshold values of visible and infrared channels; 2) spatial variation of channel data; and 3) temporal variation of the infrared channel data in a short period. We also used the SURFRAD irradiance measurements for additional cloud filtering. For most cloud-free conditions during daytime, the solar irradiance temporal profile varies smoothly (except when thin cirrus clouds occur, which have very little effect on the variation of solar irradiance). Therefore, we excluded LST data during periods when solar irradiance profiles showed high variability or discontinuities indicative of clouds. Further processing details of the GOES-8 Imager and SURFRAD match-up data sets can be found in Yu *et al.* [14].

Note that some cloud-free data may be eliminated through this cloud filtering procedure. Our goal, however, was to ensure that the remaining match-up data are reliably cloud free so that the cloud contamination effect is minimized in this evaluation study. Therefore, a conservative cloud filter is appropriate and was used.

III. METHOD

The examples presented in Flynn's work [6] deal with two sets of independent and simultaneous observations and one set of "truth" data. While Flynn used the method to evaluate cases of satellite sounding data with a single ground station, we apply the method to our satellite LST algorithm evaluation using the SURFRAD ground measurements.

Let LST_g represent the LST derived from the GOES-8 Imager data and LST represent the "true" LST. We assume that LST_g is linearly related to the true LST , as

$$LST_g = \mu_g LST + b_g + \varepsilon_g. \quad (2)$$

Physically, if we are able to plot the LST_g and LST in an x - y coordinate system, the μ_g in (2) represents the slope of the best fit line between the two variables, b_g denotes the bias of the line fitting, and ε_g denotes the random error with zero mean to the fitting line. In other words, ε_g represents the error instance of the LST_g to the true LST .

Similarly, let LST_s represent the SURFRAD ground LST measurement, assuming that LST_s is linearly related to the true LST, as

$$LST_s = \mu_s LST + b_s + \varepsilon_s \quad (3)$$

where μ_s represents the slope of the best fit line between LST_s and LST , b_s represents the bias of the fitting, and ε_s represents the error of the LST_s relative to the true LST . Note that μ_g and μ_s will be close to unity if LST_g and LST_s approximate the LST well, respectively.

In this study, we analyze a set of data pairs $\{LST_g, LST_s\}$, where LST_g and LST_s are the satellite and SURFRAD match-up data pair values described in the previous section. Variances of the two data sets can be expressed using (2) and (3) as

$$\text{VAR}(LST_g) = \mu_g^2 \text{VAR}(LST) + 2\mu_g \text{COV}(LST, \varepsilon_g) + \text{VAR}(\varepsilon_g) \quad (4)$$

$$\text{VAR}(LST_s) = \mu_s^2 \text{VAR}(LST) + 2\mu_s \text{COV}(LST, \varepsilon_s) + \text{VAR}(\varepsilon_s) \quad (5)$$

where VAR and COV are notations of variance and covariance, respectively. Note that $\text{VAR}(\varepsilon_g) = \sigma_g^2$ and $\text{VAR}(\varepsilon_s) = \sigma_s^2$, where σ_g and σ_s represent the precision of LST_g and that of LST_s , respectively. We follow the standard statistical definition of precision as a measure of the random variability or repeatability of an estimate. Note also that the bias terms were eliminated in the aforementioned variance calculation. Furthermore, to estimate the true LST term, the covariances of LST_g and LST_s can be determined as

$$\text{COV}(LST_g, LST_s) = \mu_g \mu_s \text{VAR}(LST) + \mu_s \text{COV}(LST, \varepsilon_g) + \mu_g \text{COV}(LST, \varepsilon_s) + \text{COV}(\varepsilon_g, \varepsilon_s). \quad (6)$$

The purpose of the aforementioned derivation is to estimate, or at least bound, the precision (σ_g) of the satellite measurement LST_g . Some primary assumptions are required for this estimation. First, we assume that LST_g and LST_s are independent measurements so that their error distributions are uncorrelated. Therefore, $\text{COV}(\varepsilon_g, \varepsilon_s)$ is near zero. Second, we assume that the errors in LST_g and LST_s , namely, ε_g and ε_s , are uncorrelated with the true LST, so that both the $\text{COV}(LST, \varepsilon_g)$ and $\text{COV}(LST, \varepsilon_s)$ can be ignored. Therefore, we have the following estimations:

$$\text{VAR}(LST) \approx \frac{\text{COV}(LST_g, LST_s)}{\mu_g \mu_s} \quad (7)$$

$$\sigma_g^2 \approx \text{VAR}(LST_g) - \mu \text{COV}(LST_g, LST_s) \quad (8)$$

$$\sigma_s^2 \approx \text{VAR}(LST_s) - \frac{1}{\mu} \text{COV}(LST_g, LST_s) \quad (9)$$

where $\mu = \mu_g / \mu_s$.

The above equations show that the precision of the two LST approaches (σ_g and σ_s) can be calculated from the variance of the LST estimates and the covariance between the respective estimates. Equations (8) and (9) provide an estimate of the precision of the GOES LST data, given the knowledge of the slope ratio μ .

TABLE III
NUMBER OF MATCH-UP DATA PAIRS OF GOES-8 IMAGER LSTs AND SURFRAD GROUND LST OBSERVATIONS,
FOR EACH SURFRAD SITE FOR EACH MONTH OF THE YEAR 2001

Month	Site 1		Site 2		Site 3		Site 4		Site 5		Site 6	
	Day	Night										
1	47	15	56	65	105	114	63	51	102	64	93	51
2	43	26	42	46	72	57	57	36	44	54	65	48
3	0	0	52	71	94	62	95	80	42	63	125	34
4	94	15	42	80	80	34	60	61	81	57	140	35
5	72	20	27	59	126	65	74	83	81	47	168	75
6	57	24	82	102	80	52	63	50	84	64	188	60
7	6	3	90	77	39	10	25	38	35	43	171	74
8	46	28	127	115	38	42	68	128	54	45	104	52
9	116	47	96	112	123	54	75	111	111	86	179	98
10	104	32	63	96	184	50	91	79	73	58	114	61
11	112	28	50	119	138	70	60	55	110	54	127	86
12	40	35	68	73	126	83	81	81	106	75	113	56
Total	737	273	795	1015	1205	693	812	853	923	710	1587	730

In addition, the slope ratio μ is constrained by (8) and (9), since the precisions σ_g and σ_s must be greater than zero. That is

$$\mu = \frac{\mu_g}{\mu_s} \in \left[m_{gs}, \frac{1}{m_{sg}} \right] \quad (10)$$

where

$$m_{gs} = \frac{\text{COV}(LST_g, LST_s)}{\text{VAR}(LST_s)} \quad (11)$$

$$m_{sg} = \frac{\text{COV}(LST_g, LST_s)}{\text{VAR}(LST_g)} \quad (12)$$

respectively, which define the boundary values of the slope. Therefore, the precision range of the satellite data (LST_g), as well as the precision range of the SURFRAD data (LST_s), can be estimated if the slope ratio is known. Alternatively, the slope ratio can be estimated if the relative precisions are known.

For completeness, there are three additional statistical parameters that may be useful to evaluate the satellite LST algorithm and product. First, the Pearson linear correlation coefficient which indicates how well the LST_g and LST_s are correlated

$$\rho = \sqrt{m_{sg}m_{gs}} = \frac{\text{COV}(LST_g, LST_s)}{\sqrt{\text{VAR}(LST_g)\text{VAR}(LST_s)}}. \quad (13)$$

This parameter can be used as first check of the “goodness” of the SURFRAD LST data for the satellite LST data evaluation. Also, the mean difference and standard deviation of the differences between the satellite and SURFRAD LST estimates are useful for comparisons.

IV. RESULTS

Results of the data match-up process are presented in Table III. It lists the match-up data pair count for the satellite LSTs and the SURFRAD LSTs over each SURFRAD ground station for each month of 2001. The data pairs were sepa-

rated into daytime and nighttime categories since atmospheric conditions may be significantly different from day to night, and the LST estimation from both the satellite and the ground observations varies as well. The numbers of match-up data pairs are statistically significant over each SURFRAD site for each month, except for March and July of site 1 (Pennsylvania State University, University Park). Note that the number of the match-up data collected during the daytime is usually smaller than that for nighttime, indicating that either cloudy conditions occurred more often in daytime or the daytime cloud filtering process is more stringent.

Table IV presents the precision calculation results for all the sites and for the data pairs collected day and night. A range of precision estimates for GOES-8 LSTs (σ_g) and SURFRAD LSTs (σ_s) were calculated using (8) and (9), respectively, along with the possible slope ratio μ determined in (10). That is, the covariance between the two measurements is allocated between the two “systems.” In the table, we evenly divided the range of the ratio μ into ten intervals and calculated the corresponding precision at each of the endpoints of the intervals (we call them N steps). The values of μ and its range are different from site to site, indicating that the SURFRAD LST estimates approaching the true LST [e.g., (3)] vary from site to site if we assume that the uncertainty of the GOES-8 LST estimates [e.g., (2)] to the true LST is the same at different sites.

Interestingly, the precision of the GOES-8 LST (σ_g) is at its worst when the slope ratio μ is at the low boundary value (m_{gs}), as shown in Table IV. It then improves in an almost linear fashion as μ increases. The opposite occurs for the precision of the SURFRAD LSTs (σ_s). Near the middle of the μ steps, the precisions σ_g and σ_s are very close to each other. Thus, the noise levels of the two LST measurements can be determined if they are about the same; otherwise, they are reciprocally related. This is a mathematical consequence. If we assume that all of the random error is from one estimate (e.g., LST_s), then the proper slope to use is the one obtained with it as the dependent variable—one end of the slope range. Note that slopes may have up to 5% difference between sites because of the site difference among SURFRAD stations. The satellite pixel area will be different at each SURFRAD station, introducing different errors. Some studies have been conducted

TABLE IV

PRECISION (IN DEGREES KELVIN) OF THE SATELLITE LST MEASUREMENTS (σ_g) AND SURFRAD LST MEASUREMENTS (σ_s), FROM THE SIX SURFRAD SITE MATCH-UP DATA SETS. A RANGE OF SLOPES μ IS DETERMINED BY USING (9) FOR EACH SITE, AND IT WAS EVENLY DIVIDED INTO TEN INTERVALS ($N = 1, \dots, 11$) FOR THE PRECISION CALCULATIONS. THE RESULTS WERE DERIVED FROM ALL THE DATA SETS

N	Site 1			Site 2			Site 3			Site 4			Site 5			Site 6		
	μ	σ_g	σ_s															
1	0.9756	1.6	0.0	1.0104	1.7	0.0	0.9763	1.3	0.0	0.9948	1.8	0.0	1.0130	1.6	0.0	1.0291	1.5	0.0
2	0.9783	1.6	0.5	1.0126	1.6	0.5	0.9780	1.2	0.4	0.9968	1.7	0.6	1.0150	1.5	0.5	1.0303	1.4	0.5
3	0.9810	1.5	0.8	1.0147	1.5	0.7	0.9797	1.2	0.6	0.9988	1.6	0.8	1.0170	1.4	0.7	1.0314	1.4	0.7
4	0.9838	1.4	0.9	1.0169	1.4	0.9	0.9814	1.1	0.7	1.0008	1.5	1.0	1.0190	1.3	0.9	1.0325	1.3	0.8
5	0.9865	1.3	1.1	1.0190	1.3	1.0	0.9831	1.0	0.8	1.0028	1.4	1.1	1.0210	1.2	1.0	1.0337	1.2	0.9
6	0.9893	1.2	1.2	1.0211	1.2	1.2	0.9847	0.9	0.9	1.0048	1.3	1.3	1.0230	1.1	1.1	1.0348	1.1	1.0
7	0.9902	1.0	1.3	1.0233	1.0	1.3	0.9864	0.8	1.0	1.0068	1.1	1.4	1.0250	1.0	1.2	1.0360	1.0	1.1
8	0.9947	0.9	1.4	1.0254	0.9	1.4	0.9881	0.7	1.1	1.0088	1.0	1.5	1.0270	0.9	1.3	1.0371	0.8	1.2
9	0.9975	0.7	1.5	1.0275	0.7	1.4	0.9898	0.6	1.2	1.0108	0.8	1.6	1.0290	0.7	1.4	1.0383	0.7	1.3
10	1.0002	0.5	1.6	1.0297	0.5	1.5	0.9915	0.4	1.3	1.0128	0.6	1.7	1.0311	0.5	1.5	1.0394	0.5	1.4
11	1.0030	0.0	1.7	1.0318	0.0	1.6	0.9932	0.0	1.3	1.0148	0.0	1.8	1.0331	0.0	1.6	1.0406	0.0	1.5

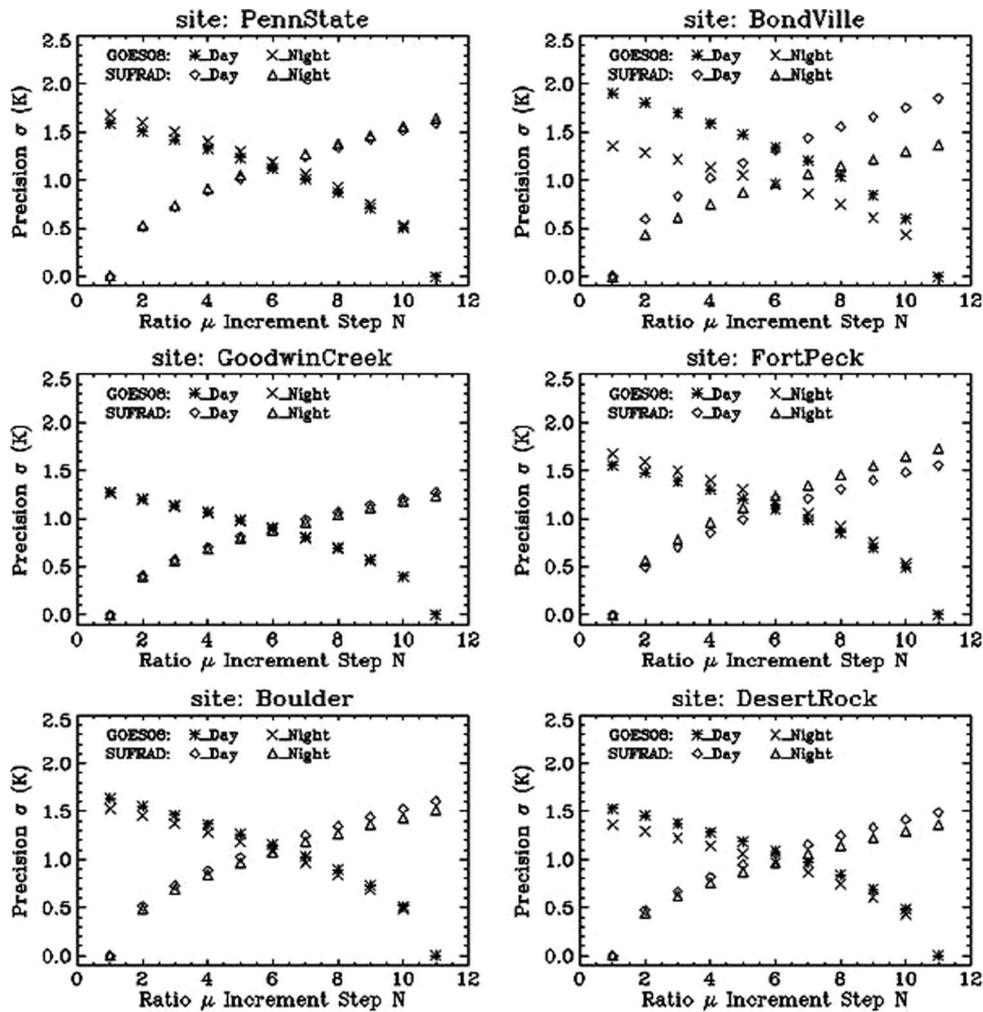


Fig. 1. Plots of measurement precision versus ratio μ values (in the possible range).

for estimating such errors by generating a site-to-pixel characterization model (e.g., Hale *et al.* [17] and Román *et al.* [18]). Within the μ range, the worst precision of both the GOES satellite LSTs and the SURFRAD LSTs occurred at site 4, where the σ value is 1.8 K.

To further investigate this feature, we calculated the precisions σ_g and σ_s for the daytime and nighttime data sets, respectively, for each site and plotted them against the step (N)

of the slope ratio increment. The results are shown in Fig. 1. Again, for each site, the precisions σ_g and σ_s vary reciprocally: σ_g increases with a decrease of the slope ratio μ and vice versa for precision σ_s . Note also that, at daytime, σ_g and σ_s at each μ step generally agree well with their counterparts at nighttime, except for site 2 (Bondville). We suspect that this is because of significant water vapor difference between daytime and nighttime.

TABLE V
 VARIANCES (VAR) AND COVARIANCES (COV) OF THE GOES-8 SATELLITE (LST_g) AND SURFRAD (LST_s) MEASUREMENT MATCH-UP DATA SETS OVER THE SIX SURFRAD GROUND SITES. THE TEMPERATURE UNIT IS IN DEGREES KELVIN

Site	VAR(LST_g)		VAR(LST_s)		COV(LST_g, LST_s)		Summary	
	Daytime	Nighttime	Daytime	Nighttime	Daytime	Nighttime	Daytime	Nighttime
1	85.24	77.74	85.50	73.44	84.09	74.17	$COV < V_g < V_s$	$V_s < COV < V_g$
2	129.26	91.05	122.49	91.71	124.06	90.45	$V_s < COV < V_g$	$COV < V_g < V_s$
3	68.44	75.25	68.25	72.95	67.54	73.31	$COV < V_s < V_g$	$V_s < COV < V_g$
4	178.86	98.42	178.95	105.39	177.69	100.38	$COV < V_g < V_s$	$V_g < COV < V_s$
5	104.69	69.06	101.40	67.30	101.72	67.01	$V_s < COV < V_g$	$V_s < COV < V_g$
6	175.27	79.27	164.82	79.19	168.82	78.29	$V_s < COV < V_g$	$COV < V_s < V_g$

TABLE VI
 CORRELATION COEFFICIENTS, MEAN VALUES OF THE DIFFERENCE (Satellite LST Minus SURFRAD LST), AND STANDARD DEVIATIONS OF THE DIFFERENCE FOR EACH SITE. THE RESULTS WERE DERIVED FROM ALL THE DATA SETS. THE NUMBER OF THE DATA PAIRS USED FOR THE CALCULATION IS ALSO LISTED

Site	ρ	Mean of Difference	Standard Deviation	Data Pairs
1	0.9862	-0.10	1.66	1010
2	0.9896	0.32	1.65	1810
3	0.9915	-0.11	1.32	1898
4	0.9901	-0.40	1.76	1665
5	0.9902	-0.33	1.61	1633
6	0.9945	-1.52	1.56	2317

Another study that should be performed is of the variances and covariances of the match-up data sets. Flynn [6] pointed out that the variances and covariances may indicate which data set more closely reflects the variation of the real area-averaged LST. Table V gives the calculation results for the daytime and nighttime cases of the match-up data sets over the six SURFRAD sites. Comparison summaries are listed in the most right column of the table, where V_s and V_g represent the variances $VAR(LST_s)$ and $VAR(LST_g)$, respectively. It is shown that, during daytime, the variance of the GOES-8 LST_g (V_g) is a little bit larger than the variance of the SURFRAD LSTs (V_s) for the sites 2, 3, 5, and 6, indicating that the satellite measurements may be a little bit noisier than the SURFRAD measurements in these sites; this is true for the nighttime cases also, except site 2. For site 4, V_s is always larger than V_g , for both the daytime and nighttime cases, indicating the noisier SURFRAD measurements. Note that the difference between V_g and V_s for all the sites, for daytime and nighttime, is relatively small, which is expected since both the data sets represent the same true LST variation.

Finally, we list in Table VI the results for the correlation coefficients, mean values of the difference, and standard deviations of the difference between the two measurements. The corresponding scatterplots in Fig. 2 show a direct comparison between the satellite and ground LST data. Over all sites, the correlation coefficients are fairly high, which indicates that, if the SURFRAD LST measurement [i.e., (3)] is a good estimation of the true LST, the precision evaluation for the GOES satellite LST would be reasonable. The mean difference and standard deviation of the difference indicate the relative accuracy of the two data sets. Interestingly, the standard deviation values for site 4 are the largest, which is coincident with the largest random error level shown in Table IV for this site. In

fact, the standard deviation values are the same (or very close) to the precisions of GOES LSTs in Table IV at step $N = 1$, when the SURFRAD LST precision is perfect (zero precision error). For any nonzero random error of the *in situ* data, this implies that the precision of the satellite LST is better than the traditionally calculated standard deviation between the satellite data and the *in situ* data.

V. DISCUSSION

There are two fundamental assumptions in this validation study. First, we assumed that both the GOES satellite and the SURFRAD LST measurements are linearly related to the true LST as defined in (2) and (3). Note that the linear fitting model separates the measurement errors into a constant (bias) term and a random error term, which allowed us to perform the error estimation through the linear algebra mechanism. Second, we assumed that the errors of the two measurements are independent since they were obtained from different platforms, which implies that the covariance of the two measurements represents the true LST variance [as is shown in (6)].

A slope applied in the linear fitting model means that, in addition to the bias and the random error, an LST-dependent error will occur if the slope is not unity. This is more likely with the SURFRAD LST approach than with the satellite LST approach since the former estimates the true LST from a point measurement that may differ significantly from the true LST over a satellite pixel from a different season. If we assume that μ_g equals unity, then $\mu = 1/\mu_s$. Therefore, μ may indicate features of the SURFRAD site-to-pixel estimation error: The SURFRAD LST is likely underestimated/overestimated for the measurements below/above the crossover point of the SURFRAD LST fitting line and the true LST line if μ is greater than one (e.g., sites 2, 5, and 6 in Table IV) and vice versa if μ is less than one (e.g., sites 1, 3, and 4 in Table IV). We observed both cases in Table IV.

As noted in the Introduction, the purpose of this study was to validate the satellite LST product. Using the linear fitting model, we obtained precision ranges over the six validation sites. A better estimation of the satellite LST precision relies on better knowledge of precision of the SURFRAD LST measurement, which is possible by further analyzing the SURFRAD data site by site. There are three possible scenarios to further narrow down the satellite LST precision estimation. If the SURFRAD LSTs are less noisy/noisier than the GOES LSTs, then we may safely reduce the GOES LST precision range into the values of first/last four ratio steps (Table IV). Alternatively, the GOES LST precision can be determined with that at the fifth to sixth ratio steps if the random error levels of the two data

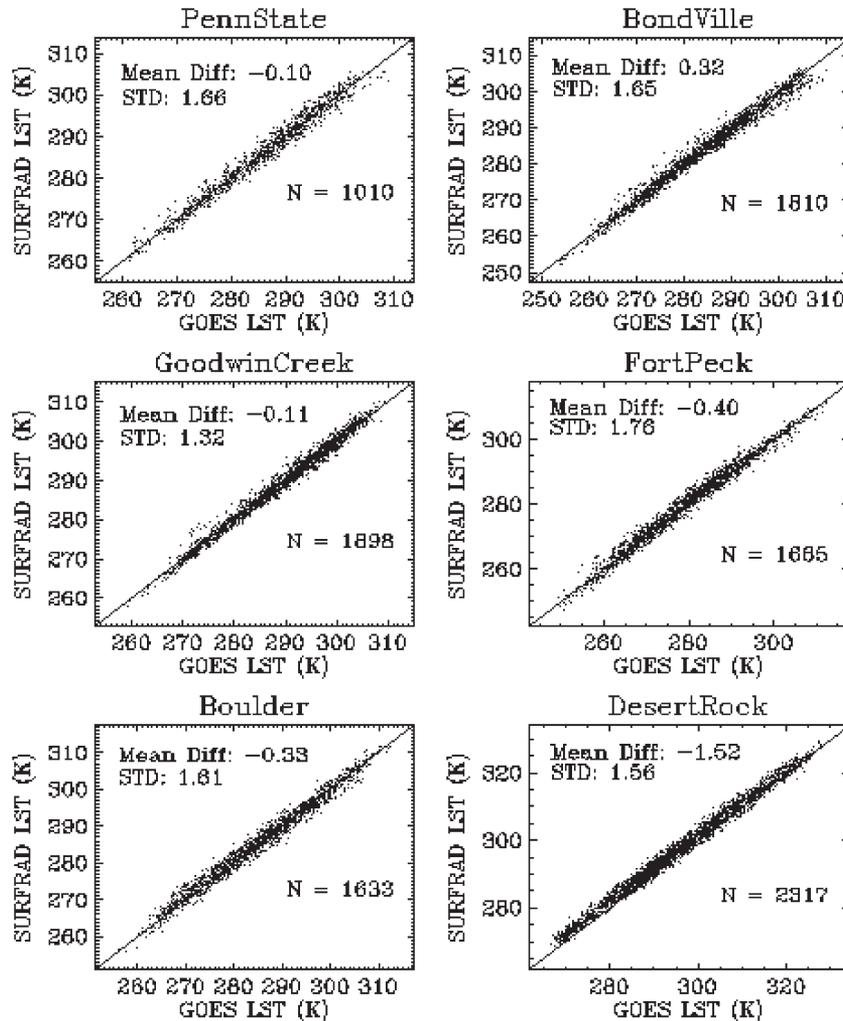


Fig. 2. Scatterplots of the satellite LSTs and SURFRAD LSTs.

sets are similar. Nevertheless, in the cases shown in Table IV, the worst precision of the GOES LST_g (σ_g) is 1.8 K at site 4; the average precision of the worst σ_g (mean of σ_g over the first row of Table IV) is 1.58 K, which is significantly better than the precision requirement (2.3 K) of the GOES-R mission [20], [21]. Interestingly, this result is also better than our previous GOES-R LST evaluation [14], which implies that the traditional LST validation process may count the *in situ* data noise into the satellite data since the *in situ* data are considered as the truth.

As mentioned earlier, a random error in the SURFRAD LST is likely site dependent. Characterizing the spot-to-pixel difference between the SURFRAD and the satellite measurements may provide meaningful information in determining the SURFRAD LST precision.

Note also that the method described in this study does not provide estimation of the bias and root-mean-square errors. The precision estimate is important and difficult to obtain. The mean difference between the satellite LSTs and the SURFRAD LSTs shown in Table VI may be the best estimate of the bias. Obviously, accurate determination of the bias relies on accurate *in situ* data. Such *in situ* data may be collected from specially designed field campaigns with very limited match-up data pairs. Wan *et al.* described examples on their MODIS LST product

evaluation [1], [21]. The bias may also be determined through the satellite LST applications.

Importantly, the linear approach model defined in (2) and (3) “wrapped up” all the differences between the satellite and *in situ* measurements. This is particularly useful for the satellite LST validation. Traditionally, the satellite LST validation is a difficult task because a satellite measurement usually represents the mean value over a pixel area (e.g., $\sim 4 \times 4$ km for GOES); due to the heterogeneity in land surface features, LST may vary significantly within a satellite pixel area. While *in situ* observations are usually collected at point scale ($\sim m$), a challenge remains in scaling these measurements to satellite pixel representation. In addition, the LST (and emissivity) anisotropic property both in the satellite and *in situ* measurements is a big concern. By using the linear approach model, all those differences are counted in the bias and random errors.

VI. CONCLUDING REMARKS

In this paper, we have presented a new method for evaluating satellite LST products. The method is applied to the GOES-R ABI LST products using GOES-8 Imager data as proxy and *in situ* data collected from the SURFRAD program. It is based

on a linear fitting model first described by Flynn [6]. It assumes that the match-up satellite data and *in situ* data are independent, which means that the error distributions of the two data sets are uncorrelated. We have demonstrated that a precision range of the GOES LST can be estimated by calculating the variances and the covariances of the match-up GOES LST and SURFRAD LST data. Such precision range can be further narrowed if the precision of the ground validation data set is known. Our experimental results over six SURFRAD sites in 2001 demonstrate that the precision of the GOES LST is, on average, about 1.58 K.

Determining the bias error is more difficult using this method. However, we consider bias estimation to be less important in prelaunch LST algorithm development. An estimation of the bias error may be obtained from calculating the mean difference between the satellite LSTs and *in situ* LSTs. However, errors may be introduced if the *in situ* data are biased and/or the slope of its fitting to the truth is not unity.

Further determination of the satellite LST precision depends on better knowledge of the *in situ* data errors. Therefore, we recommend further investigations on how to estimate random error levels of *in situ* data such as the SURFRAD data used in this study.

ACKNOWLEDGMENT

The authors would like to thank F. Weng and T. Zhu of the Geostationary Operational Environmental Satellite (GOES) R-Series Algorithm Working Group Proxy Data Team for providing the GOES-8 data and J. Augustine of the SURFRAD program for providing support in using the SURFRAD data.

The manuscript contents are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of the National Oceanic and Atmospheric Administration or the U.S. Government.

REFERENCES

[1] Z. Wan, Y. Zhang, Q. Zhang, and Z.-L. Li, "Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 163–180, Nov. 2002.

[2] C. Coll, V. Caselles, J. M. Galve, E. Valor, R. Niclos, J. M. Sánchez, and R. Rivas, "Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 288–300, Aug. 2005.

[3] Y. Yu, J. L. Privette, and A. C. Pinheiro, "Analysis of the NPOESS VIIRS land surface temperature algorithm using MODIS data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2340–2350, Oct. 2005.

[4] A. C. T. Pinheiro, R. Mahoney, J. L. Privette, and C. J. Tucker, "Development of a daily long term record of NOAA-14 AVHRR land surface temperature over Africa," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 153–164, Jul. 2006.

[5] K. Y. Vinnikov, Y. Yu, M. K. Rama Varma Raja, D. Tarpley, and M. D. Goldberg, "Diurnal-seasonal and weather-related variations of land surface temperature observed from geostationary satellites," *Geophys. Res. Lett.*, vol. 35, no. 22, p. L22 708, 2008. DOI:10.1029/2008GL035759.

[6] L. E. Flynn, "Comparisons two sets of noisy measurements," U.S. Dept. Commerce, Washington, DC, NOAA Tech. Rep., NESDIS 123, Apr. 2007.

[7] T. J. Schmit, W. P. Menzel, J. Gurka, and M. Gunshor, "The ABI on GOES-R," in *Proc. 3rd Annu. Symp. Future Nat. Oper. Environ. Satellite Syst.*, San Antonio, TX, Jan. 16, 2007.

[8] J. A. Augustine, J. J. DeLuisi, and C. N. Long, "SURFRAD—A national surface radiation budget network for atmospheric research," *Bull. Amer. Meteorol. Soc.*, vol. 81, no. 10, pp. 2341–2357, Oct. 2000.

[9] M. Hansen and B. Reed, "A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1365–1373, 2000.

[10] J. A. Augustine, G. B. Hodges, C. R. Cornwall, J. J. Michalsky, and C. I. Medina, "An update on SURFRAD—The GCOS surface radiation budget network for the continental United States," *J. Atmos. Ocean. Technol.*, vol. 22, no. 10, pp. 1460–1472, Oct. 2005.

[11] K. Wang, Z. Wan, P. Wang, M. Sparrow, J. Liu, X. Zhou, and S. Haginoya. (2005). Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products. *J. Geophys. Res.-Atmos.* [Online]. 110(12), p. D11109. Available: 10.1029/2004JD005566

[12] D. Jones, EOM, vol. 13, no. 5, Aug./Sep. 2004. [Online]. Available: http://www.eomonline.com/Common/Archives/2004augsep/04augsep_GOES-R.html

[13] History of GOES, "GOES-1 Through GOES-7." [Online]. Available: [http://www2010.atmos.uiuc.edu/\(Gh\)/guides/rs/sat/goes/oldg.rxml](http://www2010.atmos.uiuc.edu/(Gh)/guides/rs/sat/goes/oldg.rxml)

[14] Y. Yu, D. Tarpley, J. L. Privette, M. D. Goldberg, M. K. Rama Varma Raja, K. Vinnikov, and H. Xu, "Developing algorithm for operational GOES-R land surface, temperature product," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 936–951, Mar. 2009.

[15] L. L. Stowe, P. A. Davis, and E. P. McClain, "Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer," *J. Atmos. Ocean. Technol.*, vol. 16, no. 6, pp. 656–681, Jun. 1999.

[16] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophys. Res.*, vol. 103, no. D24, pp. 32 141–32 157, 1998.

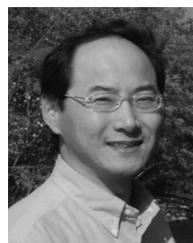
[17] R. C. Hale, K. P. Gallo, D. Tarpley, and Y. Yu, "Characterization of variability at *in situ* locations for calibration/validation of satellite-derived land surface temperature data," *Remote Sens. Lett.*, vol. 2, no. 1, pp. 41–50, 2011.

[18] M. O. Román, C. B. Schaaf, C. E. Woodcock, A. H. Strahler, X. Yang, R. H. Braswell, P. Curtis, K. J. Davis, D. Dragoni, M. L. Goulden, L. Gu, D. Y. Hollinger, T. E. Kolb, T. P. Meyer, J. W. Munger, J. L. Privette, A. D. Richardson, T. B. Wilson, and S. C. Wofsy, "The MODIS (Collection V005) BRDF/Albedo product: Assessment of spatial representativeness over forested landscapes," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2476–2498, Nov. 2009.

[19] P417-R-MRD-0070, 2007 GOES-R Series Mission Requirements Document (MRD), P417-R-MRD-0070, 2007.

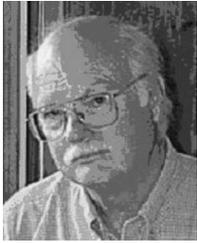
[20] GOES-R Series Ground Segment (GS) Project Functional and Performance Specification (F&PS) Version 1.10, May 8, 2009. [Online]. Available: http://www.star.nesdis.noaa.gov/star/goesr/MRD/FPS_1.10.pdf

[21] Z. Wan, Y. Zhang, Q. Zhang, and Z.-L. Li, "Quality assessment and validation of the MODIS global land surface temperature," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 261–274, 2004.



Yunyue Yu received the B.Sc. degree in physics from the Ocean University of Qingdao (OUQ), Qingdao, China, in 1982, the Diploma of M.Sc. degree-equivalent in advanced physics from Peking University, Beijing, China, in 1986, and the Ph.D. degree in aerospace engineering sciences from the University of Colorado (CU), Boulder, in 1996.

During his tenure with OUQ (in 1982–1993), he served as a Lecturer and an Associate Professor and held a leadership role in multiple international corporation projects. From 1987 to 1992, he was a Visiting Scientist at the University of Dundee, Dundee, U.K.; the Division of Atmospheric Research, Australian Commonwealth Scientific and Industrial Research Organization; and the Colorado Center for Astrodynamic Research, CU Boulder. In 1996, he joined the Earth Observation System Satellites program through Raytheon Information Technology and Satellite System (ITSS) and George Mason University, Fairfax, VA, and worked at the Goddard Space Flight Center, National Aeronautics and Space Administration. He has accomplished a variety of projects in ocean and land surface remote sensing and applications. He is currently a Physical Scientist with the Center for Satellite Applications and Research, National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Camp Springs, MD, where he is the Chairman of Land Surface Algorithms Working Group, Geostationary Operational Environmental Satellite R-Series satellite mission, and the Government Lead of the Joint Polar-orbiting Satellite System/Visible and Infrared Image Radiometer Suite land surface temperature and Albedo product development.



Dan Tarpley received the B.S. degree in physics from Texas Tech University, Lubbock, and the Ph.D. degree in atmospheric physics from the University of Colorado, Boulder.

He is currently with Short and Associates, Camp Springs, MD, where he is a Consultant for the Center for Satellite Applications and Research, National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Camp Springs, working primarily on Geostationary Operational Environmental Satellite R-Series algorithm development. His interests include the development and use of remotely sensed snow cover, vegetation conditions, land surface temperature, surface radiation budget, and precipitation products for validation and boundary conditions in NWP models.



Hui Xu received the B.Sc. degree in geography from Peking University, Beijing, China, the M.Sc. degree in land and water management from Cranfield Institute of Technology, Cranfield, U.K., and the Ph.D. degree in geography from the University of Edinburgh, Edinburgh, U.K.

She has worked in remote sensing data application, research and product development, as well as integration of remote sensing and geographical information system (GIS). She was a Project Scientist for snow mapping and snow depth monitoring with the University of Bristol, Bristol, U.K. She was with The University of Nottingham, Nottingham, U.K., where her work on monitoring the leaf area of sugar beet using ERS-1 SAR data contributed to a yield prediction project for the British sugar. She was an Assistant Professor with Frostburg State University, Frostburg, MD, where she taught remote sensing and GIS courses, before joining I. M. Systems Group, Inc., Camp Springs, MD, in 2001. Since then, she has worked at the National Oceanic and Atmospheric Administration (NOAA) Science Center on various projects, including rehosting, documentation, and validation of the automatic snow mapping system, on-orbit verification and intersatellite calibration of NOAA polar-orbiting radiometers, and her current participation in the Geostationary Operational Environmental Satellite R-Series algorithm development and testing of Normalized Difference Vegetation Index and land surface temperature products in the Land Application Team.



Jeffrey L. Privette received the B.S. degree from the University of Michigan, Ann Arbor, the B.S. degree from The College of Wooster, Wooster, OH, and the M.S. and Ph.D. degrees from the University of Colorado, Boulder, in 1994.

During his tenure with the National Aeronautics and Space Administration (NASA; in 1996–2006), he held leadership positions in SAFARI 2000, the MODerate-resolution Imaging Spectroradiometer (MODIS) Land Validation Program, and the Committee on Earth Observation Satellites Working Group for Calibration and Validation. He was NASA's Deputy Project Scientist for the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project from 2002 to 2006. In 2006, he joined the National Climatic Data Center, National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration (NOAA), Asheville, NC, to serve as the Project Manager of Scientific Data Stewardship (SDS). The SDS Project coordinates and executes NOAA's activities in climate data records. He also serves as the Land Validation Lead and as a Visible and Infrared Image Radiometer Suite Operational Algorithm Team Member for the NPOESS program. His research has focused on the retrieval and validation of land biophysical parameters from wide field-of-view imagers (e.g., Advanced Very High Resolution Radiometer and MODIS), with a special emphasis on directional effects.



Ming Chen received the B.S. degree in atmospheric physics from Nanjing University, Nanjing, China, the M.S. degree in atmospheric physics from the Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in geophysics from the University of Turin, Turin, Italy.

He was a Visiting Researcher at the University of Turin from 1993 to 1995, with grants provided by the International Centre for Scientific Culture—World Laboratory, Geneva, Switzerland. From 1996 through 1999, he was a Visiting Scientist at the ENEL-CRAM Laboratory, Milan, Italy, with fellowships provided by the International Center for Theoretical Physics, Trieste, Italy. From 1999 to 2001, he was a Postdoctoral Fellow with the Department of Civil and Environmental Engineering, University of California, Berkeley. Since 2003, he has been with the I. M. Systems Group, Inc., Camp Springs, MD, to support various research and application projects for the National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration Science Center. His research interests are in temporal–spatial filtering and modeling as they apply to remote sensing and data assimilation. He is particularly interested in physics-based modeling of remotely sensed scenes and the inference of Earth surface properties, e.g., various vegetation canopy parameters, from satellite observations through invertible models.

Lawrence E. Flynn received the B.A. degree in mathematics from the University of Maryland, College Park, in 1978 and the M.A. degree in mathematics and the Ph.D. degree in applied mathematics from the University of California, Davis, in 1981 and 1987, respectively.

For the last 15 years, he has been a Research Scientist with the National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Camp Springs, MD. His duties include research and analysis for validation, algorithm development, and calibration of existing and next-generation satellite ozone sensors.

Dr. Flynn is a member of the American Geophysical Union. His awards include an individual U.S. Department of Commerce (DOC) Bronze Medal in 2001, shared U.S. DOC Bronze Medals in 2003, 2007, 2009, and 2010, the William T. Pecora Award (Total Ozone Mapping Spectrometer Team) in 2006, and the EPA Stratospheric Ozone Protection Award (Ozone Science Tiger Team) in 2005.



Konstantin Y. Vinnikov received the Ph.D. degree from Voeikov Main Geophysical Observatory, St. Petersburg, Russia, in 1966 and the D.Sc. degree from the Higher Certifying Commission of the Council of Ministers of the Former USSR, Moscow, Russia, in 1983.

He is currently a Senior Research Scientist with the Department of Atmospheric and Oceanic Science, University of Maryland, College Park. His research interests include climate change and remote sensing.



Donglian Sun received the Ph.D. degree in remote sensing from the University of Maryland, College Park, in 2003.

She is currently a Faculty Member with the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA. Her research interests focus on information retrieval from remotely sensed data, particularly focusing on land surface temperature retrieval and flood/drought detection from satellite observations, satellite data assimilation, numerical simulations of tropical cy-

clones, and remote sensing applications to coastal studies.

Yuhong Tian received the B.S. degree in meteorology from Nanjing Institute of Meteorology, Nanjing, China, in 1992, the M.S. degree in meteorology from the Chinese Academy of Meteorological Sciences, Beijing, China, in 1995, and the Ph.D. degree in geography from Boston University, Boston, MA, in 2002.

From 2002 to 2007, she was a Research Scientist with Georgia Institute of Technology, Atlanta. She joined National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, through a contract with I. M. Systems Group, Inc., Camp Springs, MD, in August 2007. Her research interests include remote sensing of the biosphere, radiative transfer in canopy and atmosphere, application of remote sensing data in land surface modeling, and data assimilation in climate and weather forecast models.