Validation of Sea Surface Temperature Derived From Himawari-8 by JAXA

Qianguang Tu and Zengzhou Hao

Abstract—Hourly sea surface temperature (SST) retrieved from Himawari-8 by the Japan Aerospace Exploration Agency (H8-SST/JAXA, latest version 1.2) is becoming an important data source for data merging as well as for resolving diurnal variation (DV). However, the spatial and temporal variation of the errors for the full disk is still unclear. In this article, two years of H8-SSTs/JAXA are validated against *in situ* measurements from iQuam2. In general, H8-SSTs/JAXA shows small biases between -0.11 and -0.03 K with root mean square error (RMSE) between 0.58 and 0.73 K. The spatial distributions of the errors reveal the following patterns: 1) a small median bias close to 0.1 K and RMSE of 0.4–0.6 K comparing to drifters are found at satellite zenith angle (SZA) $0°$–$35°$; 2) negative biases (~ -0.3 K) are detected at SZA s $35°$–$58°$; and 3) larger positive biases exceeding 0.3 K are also found along the viewing boundaries. The temporal variations of the errors show that 1) there is no prominent seasonal variation; 2) the amplitude of the DV of the errors is only ~ 0.1 K for the statistical of all matchups, and 3) the maximum errors appears in the morning rather than in the noon. The statistics will be used in future work for DV analysis and merging purposes.

Index Terms—Errors, Himawari-8, sea surface temperature (SST), spatial variation, temporal variation, validation.

I. INTRODUCTION

SEA surface temperature (SST) is a crucial component in many physical, biological, and chemical processes within the Earth system [1]. For example, SST directly affects the air–sea interactions, seawater composition, and primary productivity. It is a key variable in the fluxes calculation of ocean surface heat and gas. It is also used to drive the numerical weather prediction and ocean forecasting models. High-quality SST data sets are increasingly needed for various applications, including offshore operations (such as oil and gas platforms), marine transportation, maritime safety, emergency response,

military operations, offshore power development and management, ecosystem assessment, marine pollution control, wave and surf models, fisheries support and tourism, and so on [2].

SST is routinely observed through *in situ* platforms (such as ships of opportunity, drifting buoys, moorings, and Argo floats) and satellite remote sensing (polar-orbiting and geostationary satellites). However, the distribution of *in situ* measurements is sparse and nonuniform in space and time. The advantages of satellite data are obvious. In particular, it has the ability to measure large areas of the ocean surface in near real time. Geostationary satellites provide high temporal resolution observations and polar-orbiting satellites provide high spatial resolution observations. One of the major shortcomings of the SST retrievals from the infrared sensors onboard satellites is their inability to estimate SST under cloudy skies. *In situ* measurements, on the other hand, may provide some sparse information over cloudy areas. The microwave sensors can also penetrate through cloud cover under most weather conditions, albeit with lower spatial and temporal resolutions. Thus, SST measurements from infrared sensors, *in situ* platforms, and microwave sensors are commonly blended together to increase the spatial and temporal resolutions of SST data and to generate daily or weekly gap-free SST maps. The derived continuous SST data are also function as input to various models that enable researchers to investigate long-term and large-scale variations at the daily, seasonal, decadal, and climatic timescales. However, SST measurements are usually from different platforms that are managed by different countries and/or agencies. The lack of an established standard on measurement protocols and retrieved methods among the different countries and/or agencies leads to uncertainly in data quality. It is thus necessary to first carefully assess the error and bias of SST data from different sources before data blending [3].

Many research works have been devoted to the investigation of the spatial and temporal coverage, bias, effective depth of different types of SST measurements. The standard deviation of error on *in situ* measurements, Advanced Along-Track Scanning Radiometer (AATSR) and Advanced Microwave Scanning Radiometer for EOS (AMSR-E) are derived using a three-way error analysis [4]. A similar method is applied to evaluate the three sources of the *in situ* SST measurements for satellite Cal/Val [5]. A comparison among different types of *in situ* measurements and Advanced Very High Resolution Radiometer (AVHRR) SST in the western Mediterranean Sea shows that the error assessment varies with the sensor type, the depth of the *in situ* measurements, and the database used [6]. The magnitude and characteristics

of uncertainties in Suomi National Polar-orbiting Partnership, the Visible Infrared Imaging Radiometer Suite (S-NPP VIIRS) SST produced by the Naval Oceanographic Office (NAVO) are also investigated [7]. Recently, the pixel-to-pixel uncertainty in the VIIRS and AVHRR SST fields over the Sargasso Sea are evaluated [8]. In addition, there are several studies on blending different sources of polar-orbiting SST retrievals and *in situ* measurements [9]–[12]. All of the above studies emphasize that the difficulty of validating SST retrieved from polar-orbiting satellites against *in situ* measurements from different platforms due to their specific characteristics. It is thus essential to gain knowledge of the error characteristics of each measurement type.

It should be noted that diurnal variation (DV) is one of the dominant variations in SST due to the solar radiation and the Earth's rotation. SST with the high temporal resolution, such as hourly timescale, is very important for the air–sea interaction studies and coastal ocean modeling [13]. Long-term buoy observations showed that the DV of SST exists at about 0.2 m, with a mean climatic state of about 0.2–0.6 K [14]. Large diurnal warming signals of the SST are found on low winds and high insolation conditions, especially for the skin layer SST (SST_{skin}) that represents the first few millimeters of the upper ocean [15]. This layer is also the observed depth from the infrared radiometers on the satellites. The use of daily mean SST would smooth out the DV signals. In [16], the DV of SST_{skin} could lead to an estimated error of about 4.5 W/m^2 for air–sea flux calculations in the global ocean, and even up to ∼10 W/m^2 in tropical oceans. The DV of SST_{skin} might also affect the mixing of the near-surface ocean, the energy and gaseous exchange at the air–sea interface, the formation and change of clouds, the Madden–Julian oscillation and the inter-annual variation of the atmosphere [17]. Therefore, an accurate SST_{skin} with the high temporal-spatial resolution is of scientific and practical interest for meteorology, oceanography, and remote sensing communities.

Nowadays, geostationary satellites are the only practical way to obtain data with sufficient frequency across the extensive oceans to resolve the DV. For example, a number of geostationary satellites show the capacity for monitoring of high-temporal SST images, such as the Geostationary Operational Environmental Satellites of the National Oceanic and Atmospheric Administration (NOAA), the Meteosat Second Generation (MSG) of European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), the Communication, Ocean, and Meteorological Satellite of the Korean Meteorological Administration, and a series of FengYun geostationary satellite of China Meteorological Administration (CMA) [18]–[21]. They are increasingly used in the study of the oceanic diurnal cycle. In principle, the uncertainties of SSTs retrieved from geostationary satellites would be larger than those from polar-orbiters, due to the coarse channel digitization levels, poor spatial resolution, and sensor acquisition schedules. The SST retrieved from the Multi-functional Transport SATellite-1R (MTSAT-1R) by Australian Bureau of Meteorology are validated against the drifting buoy and AATSR SST data from January to April 2010 in the tropical warm pool region and the results show that the MTSAT-1R SSTs are overestimated at cold SSTs and

underestimated at warm SSTs [22]. MSG spinning enhanced visible and infrared imager derived hourly SSTs are compared against the drifting buoy [23] and unpumped Argo SST measurements [24], drawing the users' attention on systematically investigating the similarity of geostationary satellite SSTs measurements and *in situ* SSTs measurements during the daytime. Therefore, the SSTs retrieved from geostationary satellites are in an urgent need of a strict validation program before they can be used for further applications. Accurate and well-documented *in situ* measurements provide a means to test satellite instrument performance, verify the atmospheric correction strategies, and geophysical algorithms used to derive SST estimates from top of the atmosphere radiances, and quantify the accuracy (bias and standard deviation) of the SST products.

Himawari-8 (H8) is the first next-generation geostationary meteorological satellite operated by the Japan Meteorological Agency. It was successfully launched on October 7, 2014 and entered operational service on July 7, 2015. The satellite is located at 140.7°E, 36 500 km above the equator. The primary sensor on board H8 is the Advanced Himawari Imager (AHI), a 16 multispectral bands imager that capture visible light and infrared images ranging from 60°S to 60°N between 80°E to 160°W, covering the western Pacific Ocean and the eastern Indian Ocean. There are two operational AHI SST products produced by the NOAA Advanced Clear-Sky Processor for Ocean SST system (H8-SST/ACSPO) and Japan Aerospace Exploration Agency (H8-SST/JAXA), respectively. The H8-SSTs/ACSPO and MTSAT-2 SSTs are compared with the Tropical Atmosphere Ocean (TAO) array and the Great Barrier Reef SST measurements during a three-month overlap period, and the H8-SST/ACSPO shows a mean improvement of ∼0.15 K [25]. H8-SSTs/JAXA are validated against drifting and tropical moored buoy data from June to September 2015, and it shows a root-mean-square error of ∼0.59 K with a bias of ∼−0.16 K [26]. These validations are limited in some special regions for only a few months, which failed to provide the seasonal and regional validation of the bias for the full disk.

The aim of this article is to provide a comprehensive validation for the full domain of the H8-SSTs/JAXA (latest version 1.2) and to investigate the temporal and spatial variations of the errors between various types of *in situ* measurements based on data from an extensive time period. Section II introduces the H8-SSTs/JAXA and *in situ* data set, as well as the methods used in this article, followed by Section III that illustrates the validation results and the spatial and temporal characteristics of the difference between the H8-SST/JAXA and *in situ* measurements. Dependence of the differences is discussed in Section IV. The conclusion is presented in Section V.

II. DATA AND METHODS

A. *Himawari-8 SST*

The H8 AHI includes five infrared (3.9, 8.6, 10.4, 11.2, and 12.4 μm) bands that can be used to retrieve the SST. As a successor of the Multi-functional Transport Satellite 2 (MTSAT-2), AHI shows better performances. The AHI infrared bands have a spatial resolution of 2 km at nadir, which doubles that of the

MTSAT-2. Additionally, the time sampling interval of full-disk observations is 10 and 2.5 min for the area adjacent to Japan for H8 as opposed to 30 min of the MTSAT-2.

These significant improvements bring a great opportunity to derive an hourly cloud-free SST that can be used to resolve the DV. The level 3 Himawari-8 hourly SSTs are produced by JAXA based on a quasi-physical algorithm [26]. The algorithm was developed to reduce those systematic errors by taking into account the physical processes of radiative transfer. Thus, it produces the SST in the skin layer (about 20 μm from the surface) in contrast to the ACSPO that converts to the buoy measure depth. The version 1.0 of JAXA was first released to the public in August 2015 and large biases were corrected in version 1.1 in December 2015. The latest version 1.2 improves the cloud screening based on the Bayesian inference method that was used to detect cloudy pixels. It consists of a series of tests that consider various quantities such as the local values of the gradient, temperature, probability of ice, etc. Himawari-8 SST retrievals are classified using a quality flag index of 0 (unprocessed), 1 (erroneous), 2 (bad), 3 (acceptable), 4 (good), and 5 (excellent). In this article, only H8-SSTs/JAXA flagged with quality 3 or higher were used. To investigate the temporal and spatial variation of the errors for these skin H8-SSTs/JAXA, a long period data from July 2015 to June 2017 are used in this article. Version 1.2 H8-SST/JAXA can be accessed from the website ftp.ptree.jaxa.jp.

B. In situ *Observations*

In situ SSTs are usually considered to be the "ground truth" to validate the satellite-retrieved SST. In this article, the *in situ* SSTs are accessed from the second version of *in situ* SST Quality monitor (iQuam2) (http://www.star.nesdis.noaa.gov/sod/sst/iquam/.) The iQuam2 was developed in National Environmental Satellite Data and Information Services/Centre for Satellite Applications and Research (STAR) and the data have a uniform quality control procedure. The quality control algorithm implemented in iQuam2 mainly includes five steps: prescreening, plausibility, internal consistency, mutual consistency, and external consistency. It is designed to remove the gross errors while minimizing the disturbances to the intrinsic statistical characteristics of *in situ* measurement error [27]. Only the best quality (level 5) of the *in situ* data is taken as a reference for validation. Since each type of the *in situ* SST has limited time/space coverage during the two-year period, three major *in situ* data types reported in iQuam2 are combined for our validation: ships, drifters, and Argo floats. Ship reports are mainly from the merchant ships participating in the World Meteorological Organization Voluntary Observing Ship program and to a lesser extent from the research vessels. Ships are grouped along the shipping lines even in coastal regions that are neither sampled by drifters nor Argo floats. Note that the SSTs derived from H8 are skin SSTs, whereas the *in situ* measurements mentioned above are at a certain depth. Thus, a number of skin SSTs measured from an Infrared Autonomous Radiometer model 5D (ISAR) that are equipped on the Australian Integrated Marine Observing System ships (IMOS) are also used for additional validation. The ISAR is a self-calibrating instrument capable of measuring skin SST to an accuracy of 0.1 K [28]. Wind speed from some *in situ* measurements are also used in this study.

C. Collocated Criteria

To validate the satellite retrieved SSTs with *in situ* SSTs, we produced pairs of collocated Himawari-8 SSTs and *in situ* SST observations within certain spatial (Δx) and temporal (Δt) windows. The collocation includes two major processes: matchup of collected data and additional quality control. To speed up the collocation, all the *in situ* measurements located at the validated satellite image pixel (both distance of the *in situ* measurement to the center latitude and longitude of the image pixel less than 1 km) are searched first. Then the *in situ* measurement nearest to the center of the pixel is selected to generate the matchup pair. Temporally, the difference between the two values should not exceed 30 min. Only H8 SST with quality flags of 3 or higher and *in situ* SST with quality flags of 5 are used. Since different platforms employ a large variety of sensors that operate in a wide range of often hostile environments and use different measurement protocols, it is still necessary to make additional quality control tests to ensure the quality of the comparisons. The outliers are conventionally identified with the mean and standard deviation, which may be contaminated by outliers. To overcome this problem, robust statistics is applied to remove the extreme residual outliers [29]. Reynolds optimal interpolation 0.25° daily analysis SST (OISST) is used as a reference. The robust counterparts (median, robust standard deviation, or RSD) are calculated and used to flag outlier. The matchup pairs will be excluded when the differences between SST measurements and reference are greater than "median+ 4×RSD" or less than "median- 4×RSD." Here, RSD is defined as: (3rd quartile – 1st quartile)/1.38. The *in situ* measurements within these spatial and temporal windows were averaged. A total of 2 781 224 data points were used for the validation, and a total of 538 250 data points, about 19% were rejected as outliers. All subsequent results are based on the exploitation of the matchup pairs. The spatial distribution of the *in situ* data is shown in Fig. 1. The retrieval error is estimated by calculating bias and root mean square error (RMSE) using the following formula:

$$\text{Bias} = \sum_{i=1}^{N} \frac{\text{SST}_{\text{satellie},i} - \text{SST}_{\text{in situ},i}}{N} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(\text{SST}_{\text{satellite},i} - \text{SST}_{\text{in situ},i}\right)^2}{N}} \tag{2}$$

where i is the number of collocated data points and N is the total number of collocated data points.

III. RESULTS

A. Characteristics of the Matchups

The validation is performed for the full disk of the H8-SSTs/JAXA for a two-year period from July 2015 to June 2017. This section shows some basic statistics for the spatial and temporal characteristics of the matchups. Representative
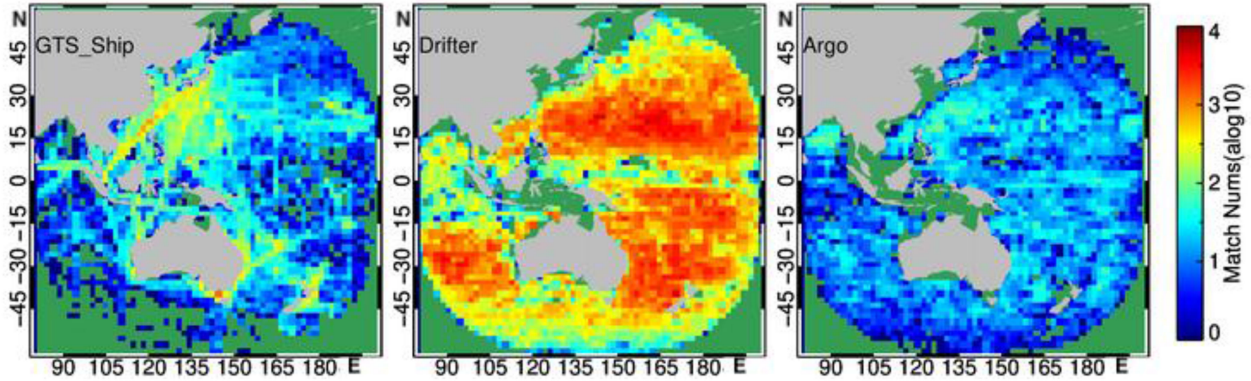
Fig. 1. Spatial distribution of the numbers of the matchups between H8-SST/JAXA retrievals and *in situ* SST measurements. (a) Ship, (b) Drifter, and (c) Argo in $2° \times 2°$ boxes for a 2-year period from July 2015 to June 2017. (Note that the color bar is in log scale).
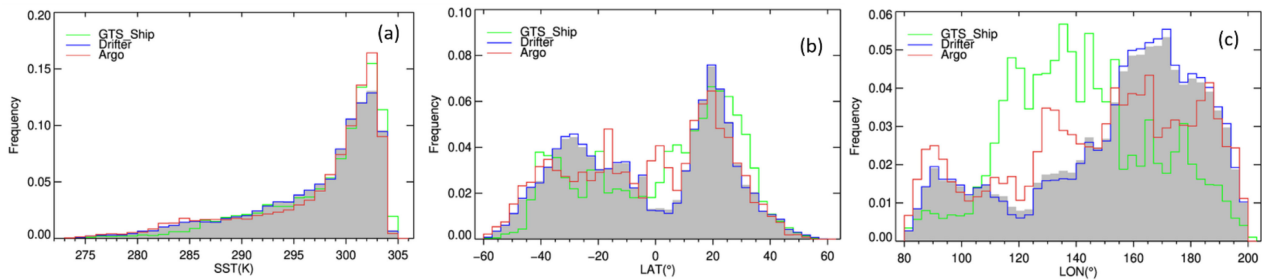

Fig. 2. Histograms (integral normalized to 1) as a function of (a) SST in 1 K bin, (b) zonal, latitude in 3° bin and (c) meridional, longitude in 3° bin for different types of *in situ* data for 2 years matchups (July 2015–June 2017). Grey histograms show corresponding histogram of all matchup pixels.

coverage of *in situ* measurements in spatial and temporal is crucial for the satellite retrievals validation [30]. Fig. 1 shows the spatial distribution of the numbers of the matchups between H8-SSTs/JAXA and *in situ* SST measurements (a) Ship, (b) Drifter, and (c) Argo in $2° \times 2°$ boxes, respectively. Note that the color bar is in log scale. All of these three types of matchups almost cover the entire H8 domain, but the spatial density distribution of matchups is not uniform. Ships well sample the northwestern Pacific (with increasing data density around major ship routes) hundreds of times in each box but are very sparse in the tropics, especially in the Indian Ocean and the Southern Ocean. Drifters provide the densest (hundreds to thousands) and most complete coverage over the open ocean, whereas there are no matchups in the Marginal Seas, such as the East China Sea and Malay Archipelago. Fewer matchups are found in the tropical oceans. On the one hand, the equatorial upwelling and surface current divergence tend to remove the drifters from the tropics. On the other hand, numbers of the H8-SST are masked due to the con- tamination of clouds or precipitation. Argo floats also provide complete and uniform coverage over the open ocean. While the density of the matchups is at the magnitude of tens, the two orders are smaller than those from the drifters. The spatial distribution of the matchup data was also affected by the topography, the presence of clouds, and the distribution of currents.

To further analyze the full disk representativeness of each type of the *in situ* data, their SST and zonal/meridional densities are plotted. Fig. 2 shows the histograms (integral normalized to

1) as a function of (a) SST in 1 K bin, (b) latitude in 3° bin and (c) longitude in 3° bin for ships, drifters, and Argo observations in green, blue, and red line, respectively. The grey histograms corresponding to the histogram of all matchup pixels are analyzed first. The *in situ* SSTs of the matchups range from 275 to 305 K, and the number of the maximum value around 302 K. The deficit of low temperatures is mainly because the full-disk observations of H8 are limited in $\pm 60°$ latitude and the data are excluded as outliers with water freezing. The zonal coverage of the matchups shows an obvious bimodal distribution. They are mostly distributed around the middle latitude ($\sim 20°$) and less in the tropical oceans and high latitudes above 40°. The meridional coverage of the matchups was mostly distributed around 170°E in the open oceans and less in the marginal sea around 120°E. If the *in situ* data cover the disk ocean fully and uniformly, their SST and zonal/meridional densities should closely match all the corresponding histogram of the matchups (in grey color). Drifters cover the zonal areas mostly uniformly, whereas they are slightly underrepresented in the meridional regions between 115°E and 140°E and slightly overrepresented between 155°E and 200°E. The zonal/meridional coverage of ship data is very nonuniform. In particular, the northwestern Pacific Ocean (0–40°N, 110–150°E) is heavily overrepresented, whereas the other areas are strongly underrepresented, such as the Indian Ocean and Southern Ocean. The meridional coverage of Argo data is slightly overrepresented from 80 to 155°E and underrepresented from 155 to 185°E. Results from Figs. 1 and 2
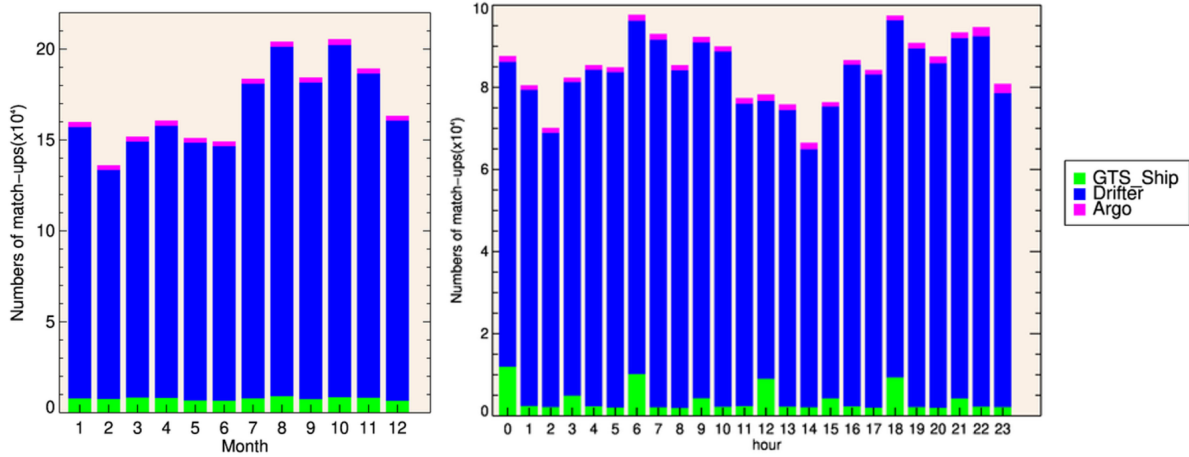
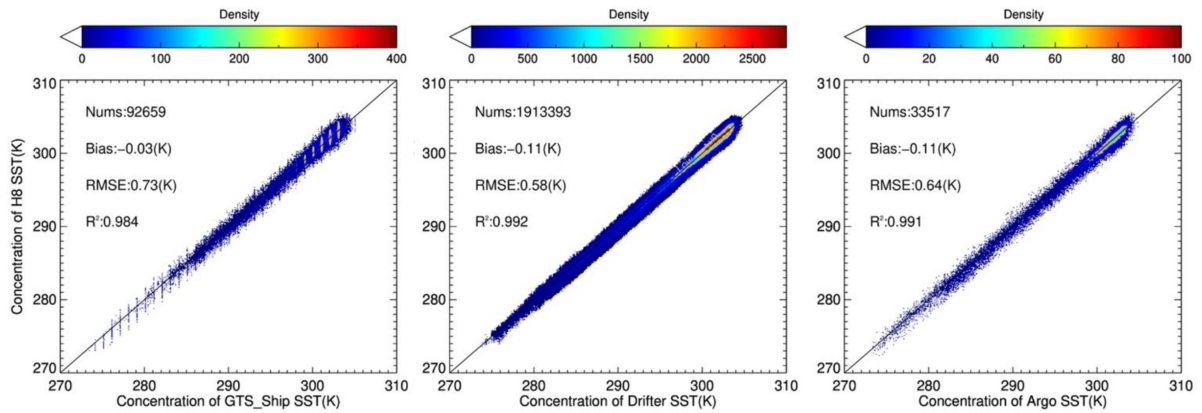Fig. 3.   Histogram of matchups for 12 months and 24 h.

Fig. 4.   Density plots of the H8-SST/JAXA validation against the *in situ* SST measurements from (a) GTS_Ship, (b) drifters, and (c) Argo floats. The black dash line is 1:1 reference line.

suggest that only one type of *in situ* observations do not represent full disk spatial distribution of SST but can be combined with all of them to improve the representation of the matchup data sets.

The temporal distribution of matchups is examined. Fig. 3(a) shows the number of each type of matchups for 12 months. The main contribution of the matchups is drifter, which accounted for more than 92% each month. The matchups for ships and Argo are homogeneous, with a number of about 7700 and 2800 per month, respectively. The number of matchups for drifters' measurements is much larger and little heterogeneous, about 145 000 from January to June and 175 000 from July to December. Fig. 3(b) shows the number of each type of matchups for 24 h. The distribution through the hours of Argo matchups is homogeneous, while the ship and drifter matchups show a slightly inhomogeneous. Presumably, a high number of ships have an automated procedure for the recording of surface temperature data, and it is possible that this procedure is more often established at precise hours four times (00:00, 06:00, 12:00, 18:00) a day.

B. General Error Statistics

After all the necessary quality control and collocation steps, Fig. 4 shows a 2-D density plots for H8 SST validation against

TABLE I
STATISTICS OF H8-SST/JAXA –*IN SITU* TEMPERATURES (K)

in situ type	Nums	Bias (K)	RMSE (K)	R^2
GTS_Ship	92659	-0.03	0.73	0.984
Drifter	1913393	-0.11	0.58	0.992
Argo	33517	-0.11	0.64	0.991

ships, drifters and Argo floats measurements from July 2015 to June 2017. The overall statistics, such as matchup numbers, mean bias, RMSE, and coefficient of determination are listed in Fig. 4 and also summarized in Table I. The number of matchups with respect to Argo floats is smallest and is about 33 517. Because it takes about ten days to get one SST profile report. The number of ships and drifters matchups is ∼2.7 times (∼92 659) and ∼57 times (∼1 913 393) larger than those of Argo, respectively. The H8-SSTs/JAXA having a small negative bias range from −0.03 to −0.11 K against all the three types of *in situ* measurements for the full domain. The RMSE is slightly higher in the case of H8-SST/JAXA validation against the ship data set (0.73 K) as compared to the drifter SST data set (0.58 K) and Argo (0.64 K). The coefficient value of determination ($R^2 = 0.98$–0.99) indicates a very strong agreement between the
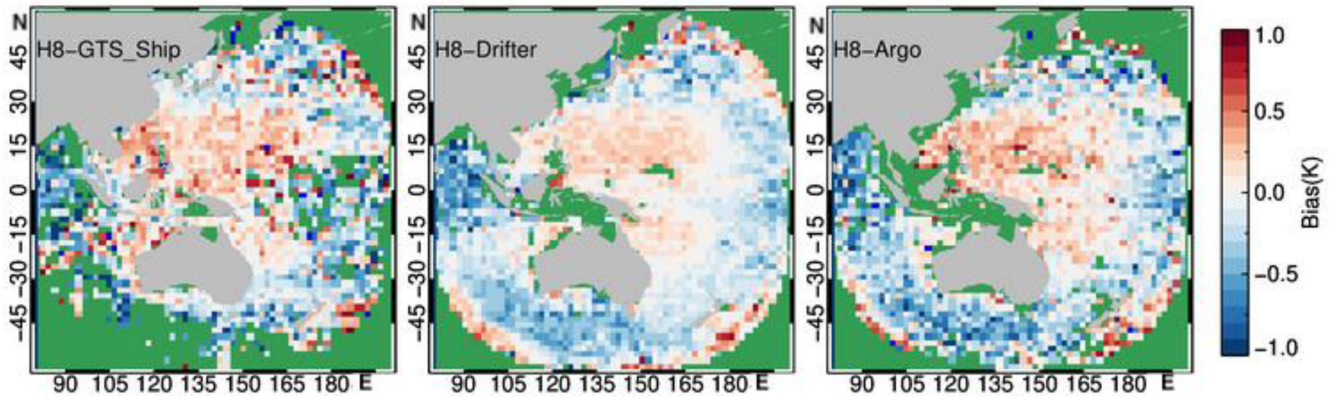
Fig. 5.   Spatial distribution of mean biases between H8-SST/JAXA and *in situ* SST measurements from (a) Ships, (b) Drifter, and (c) Argo in bins of 2° × 2° for the period from July 2015 to June 2017.
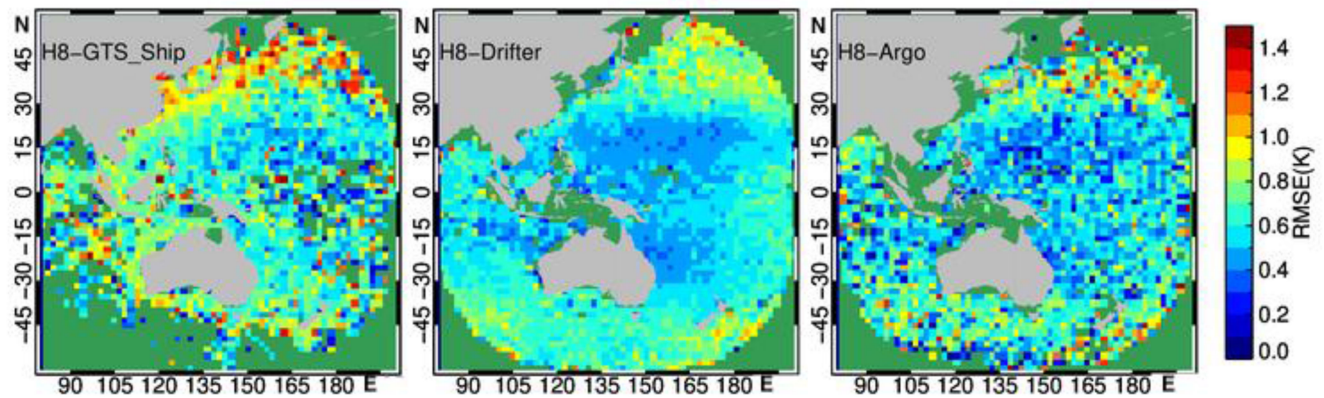

Fig. 6.   Same as in Fig. 5 but for RMSE.

H8 SSTs and *in situ* SSTs. The reasons for this difference will be further discussed in Section III-C.

C. Error Characteristics

Although the density plots of all the data show an overall accuracy of the measurements, more details about the spatial and temporal variations of the errors need to be investigated.

1) Spatial Distribution of the Errors: Fig. 5 illustrates the spatial distribution of differences between H8-SSTs/JAXA and *in situ* temperatures for 2° × 2° boxes. Since the drifters have a much larger number of matchups and uniform distribution (see Fig. 1), the distribution of the mean bias is examined first. In some of the coastal regions, the mean bias was not calculated due to the deficit of drifters. Overall, there are biases within ±0.3 K throughout the full-disk area except when negative biases exceeding −0.3 K are identified in the Indian Ocean. There are small positive biases about 0.1 K in the low- and middle-latitude Pacific Oceans at SZAs about 0°–35°. Larger positive biases exceeding 0.3 K are also found along the viewing boundaries, where SZAs near 60°. The satellite-derived SSTs tend to be slightly underestimated in the Southern Ocean. The comparison between the H8-SSTs/JAXA and ship measurements shows a generally positive and slightly higher in the

low- and middle-latitude of the Northern Hemisphere, especially for the low-latitude northwestern Pacific, such as the South China Sea. Although there are sparse calculated data in the Southern Ocean and some open ocean due to the deficit ship measurements, it filled the gaps in the coastal seas where there is a lack of drifter or Argo observations. The mean bias is about 0 K in the East China Seas and the Japan Sea. The main pattern of spatial distribution for mean bias from Argo is similar to that of the Ship.

Fig. 6 shows the spatial distribution of RMSEs in 2° × 2° bins from H8-SSTs/JAXA with respect to the ships, drifters, and Argo SST. Most of the bins from drifters show relatively small RMSE values, i.e., about 0.4–0.6 K, except for the pixels near the boundary of the full-disk region with relatively high SZA. This trend was also confirmed in the case of ships and Argo measurements comparison. However, the amplitudes of the RMSE from ships are much larger than the others. This is because the uncertainty is much larger for the ship measurements as mentioned before.

2) Temporal Distribution of the Errors: We then compute the monthly and hourly bias of H8-SSTs/JAXA against each type of *in situ* measurements to explore if there are any temporal variations. Fig. 7 shows the monthly bias of the full disk between the H8-SSTs/JAXA and each type of the *in situ* SST in the
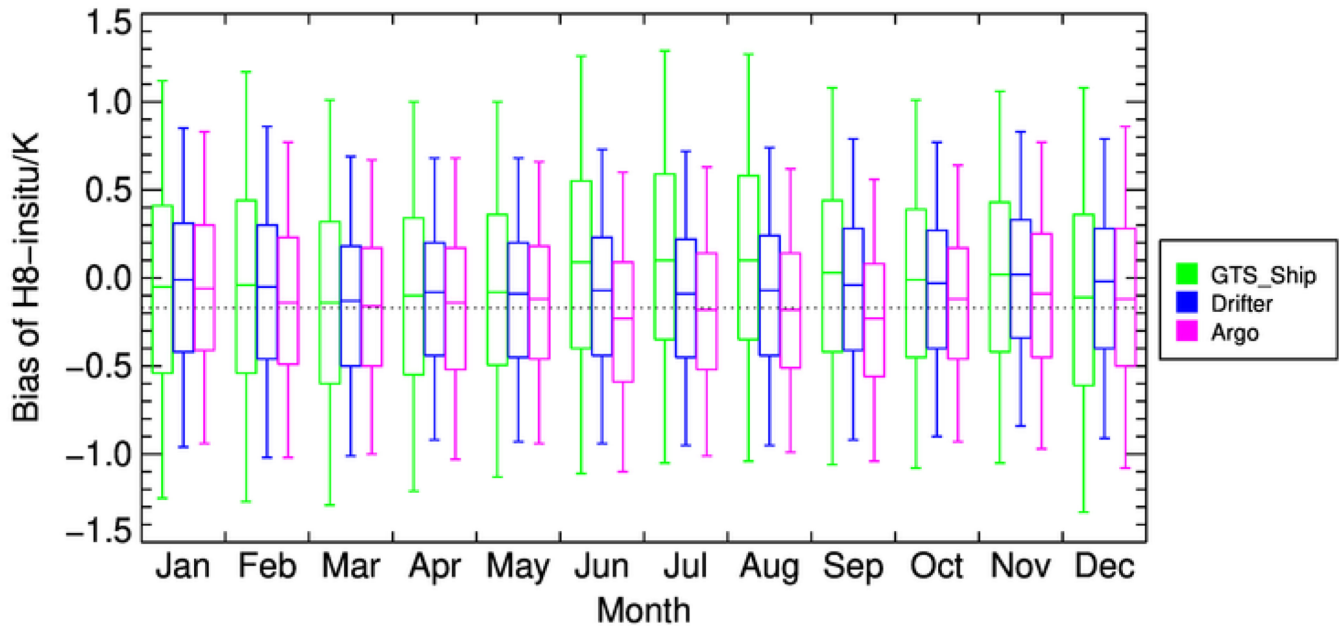
Fig. 7.   Monthly variations of differences between the H8 SSTs and the *in situ* SST from July 2015 to June 2017. Lower and upper box boundaries denote the first and third quartiles, and solid line inside box is median value.

box-whisker plots. The central bar in the box indicates the median value, and the lower and upper borders of the box represent the first and third quartiles of the distribution of values. The whiskers extend out to the maximum or minimum value of the data, or to the 1.5 times of either the first and third quartiles. There is no prominent seasonal variation during the period analyzed. When it comes to median bias, the monthly bias of drifter and Argo are more stable than the ships as expected. Overall, the median bias of drifter and Argo are of the order of -0.1 K, while the median bias of ship fluctuated from -0.1 to 0.2 K. The monthly inter quartile range (IQR) of biases for ships was 0.83–0.98 K, 0.64–0.75 K for drifter and 0.63–0.78 K for Argo floats. The larger positive biases of ships appear in boreal summer times (from June to September) is likely related to their matchups are major located in the Northern Hemisphere. The diurnal warming is significant for the skin layer when compared to the $\sim$10 m depth of the ship measurements at high insolation and low wind speed conditions in these areas.

The most notable accuracy concern becomes apparent when the statistics are computed separately for each time of day at which the satellite measurements are obtained. The box-whisker of the biases for the full disk domain of H8 skin SST relative to the *in situ* measurements are plotted as a function of local solar time in Fig. 8. H8 skin SST and *in situ* SSTs are subject to daytime warming and nighttime cooling, whereas OISST represents the foundation SST, which do not resolve the diurnal cycle [12]. Thus, the DV of the H8-SSTs/JAXA [see Fig. 8(a)] and *in situ* SSTs [see Fig. 8(b)] with respect to the OISST is investigated first. The amplitude of the DV is only $\sim$0.1 K for the statistical of all matchups. In the morning, H8 SSTs start to increase immediately after sunrise, whereas there is about a 3-h lag for the drifters. Both of the peak values appear at $\sim$13:00–15:00 LT. The afternoon cooling is almost simultaneous for

both measurements. These features are similar to what has been observed between 0 and 20 cm depth on *in situ* measurements, for instance in the Mutsu Bay experiment [31]. The morning time lag is however larger in the present comparison; this unexpected characteristic will be the subject of further studies. Therefore, the maximum difference between the H8 skin SST and *in situ* SSTs appears in the morning rather than in the noon when considering the offset of the cool skin constant about 0.17 K [see Fig. 8(c)].

IV. DISCUSSION

The comparison of Himawari-8 SSTs with *in situ* data can be used to identify systematic errors or serious regional anomalies. This is necessary before their further analysis or blended with other sources SSTs. The validation of H8-SSTs/JAXA (version 1.2) has been carried out against the *in situ* measurements from ships, drifter buoys and Argo in this article. Statistical results show that the H8-SST/JAXA (version 1.2) has a biases and RMSE of -0.03 and 0.73 K, -0.11 and 0.58 K, -0.11 and 0.64 K when comparing against ships, drifter, and Argo, respectively. It shows little improvement when comparing to previous validations for the H8-SST/JAXA and H8-SST/ACSPO. The H8-SST/JAXA (version 1.1) has been compared with iQuam SST and statistics show a bias of -0.3 K with the standard deviation of 0.7 K [32]. The preliminary validation statistics for H8-SST/JAXA (version 1.2) using three months (from June to September, 2015) of matchups with drifting and tropical moored buoy data shows a bias of ~-0.16 K and root-mean-square difference of $\sim$0.56 K [26]. Another validation for H8-SST/ACSPO shows a mean bias within $\pm$0.2 K and standard deviation of 0.45 K [32]. The accuracy of H8-SST/ACSPO in the tropical
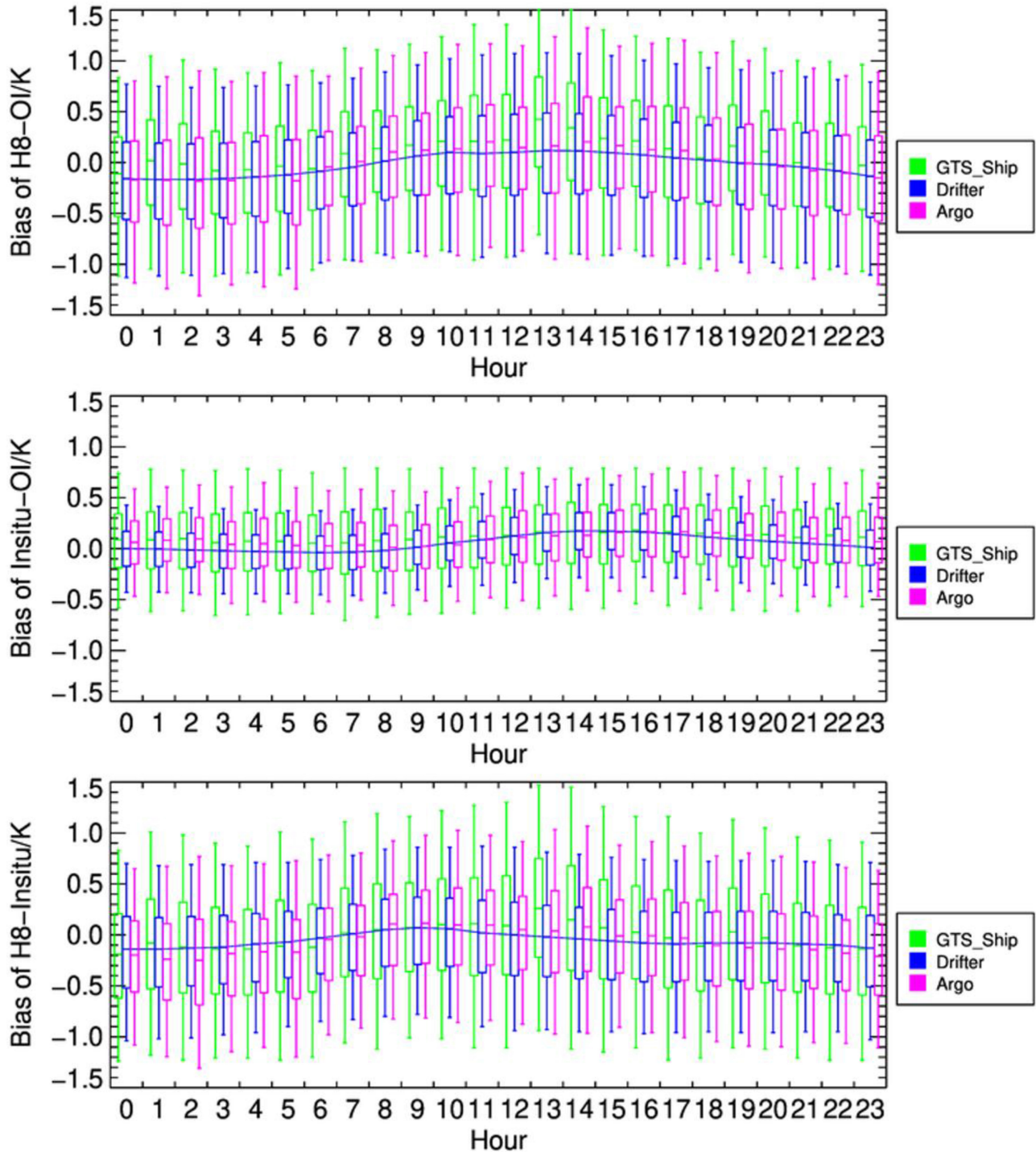
Fig. 8. Box plots of the hourly mean differences between the (a) H8-SSTs/JAXA and OISST, (b) *in situ* SST and OISST, and (c) H8-SSTs/JAXA and *in situ* SST from July 2015 to June 2017.

western Pacific Ocean show a bias of ∼0.16 K and standard deviation of ∼0.53 K based on a comparison against three month *in situ* measurements from the TAO array and self-recording thermometers at the depths of corals of the Great Barrier Reef [25].

Four main terms can contribute to the difference between the satellite observations and *in situ* measurements: the satellite uncertainty, the *in situ* uncertainty, geophysical uncertainty, and cloud contamination [1]. A critical improvement on cloud detection and SST retrieval algorithm was applied to the H8-SST/JAXA (version 1.2) by JAXA Team to reduce the uncertainty of SST. The cloudy pixels are screened by the Bayesian inference method and then the SSTs are retrieved by a physical process of radiative transfer. It would be the main improvement
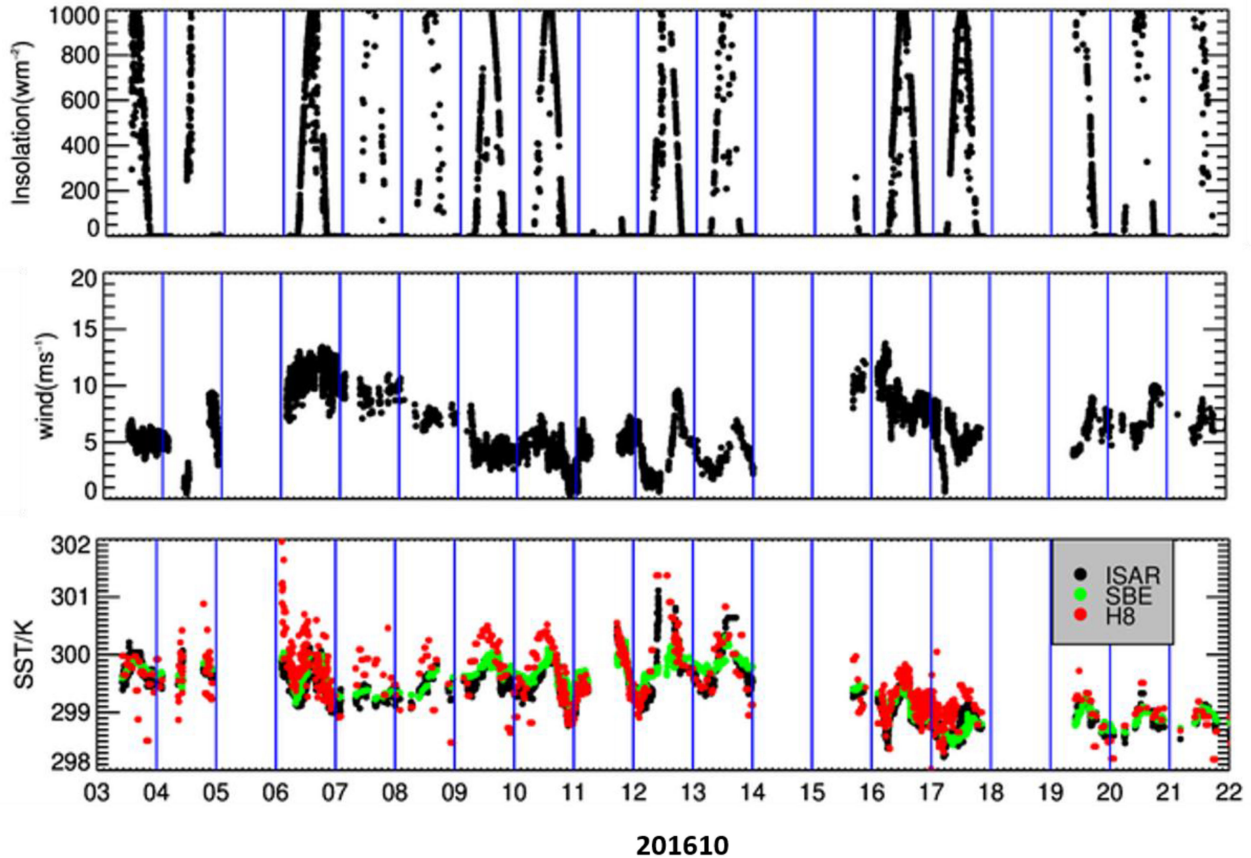
Fig. 9.   Example time series (from top) of insolation (W·m^{-2}), wind speed (ms^{-1}), SBE 38 SST$_{depth}$ (green, /K), ISAR SST$_{skin}$ (black, /K) from IMOS and H8-SST/JAXA (red, /K) during October 3–21, 2016.

to reduce the bias from -0.3 K of version 1.1 to -0.1 K of this version 1.2. The *in situ* uncertainties estimated using three-way error analyses are $\sim$0.75 K for ships, $\sim$0.21–0.22 K for drifters and Argo floats, respectively [33]. It would be the main reason for the difference of RMSE for the H8-SST/JAXA (version 1.2) compared with three types of the *in situ* measurements. In this section, we focused on geophysical uncertainty: depth, spatial, and temporal.

The geophysical depth uncertainty is the main factor that results in the difference between the satellite and *in situ* comparisons [34]. The vertical structure of water temperature in the upper ocean is affected by solar radiation, wind speed, ocean current system, and other variables. Especially under low wind speeds and high insolation conditions, the diurnal thermocline of the upper ocean can develop significantly. The satellite infrared sensors observe the skin temperature at a depth of about 10 μm, while the typical measurement depth of $\sim$0.2 m for drifters, several meters for the ships and 4–5 m for Argo. This application would be limited by the thermal stratification of the top few meters of the ocean, particularly in regions of high solar insolation and low wind speeds. The best satellite skin SST validation measurements are contemporaneous satellite and *in situ* skin SST measurement in theory. In this study, the skin SSTs received from a shipboard radiometer systems ISAR from IMOS are used for an additional comparison against

H8-SST/JAXA. Example time series of ISAR data set and H8-SST/JAXA (red) during October 3–21, 2016 are plotted in Fig. 9. Gaps in the time series have resulted from cloud or precipitation occurs. It illustrates the close agreement between the ISAR SST$_{skin}$ values (green line) and the SBE 38 SST depth values on the vessel, except during periods of diurnal warming of the surface ocean that are associated with low wind speeds ($<$ 6 ms^{-1}) and at high shortwave solar radiation, such as October 12, 2016.

Therefore, the temporal errors of satellite-derived SSTs with respect to the *in situ* temperature may represent different characteristics mainly depending on the wind speed. In Fig. 10, each panel displays the biases between H8 SST and *in situ* measurements as a function of wind speed during daytime and nighttime. It is obvious that most of the matchups are sampled when the wind speeds are less than 15 ms^{-1} because satellite-derived SSTs cannot be obtained in bad weather. For the daytime, the biases show an exponential dependence on wind speed. It decreases as wind speed increases when the wind speed below 6 m/s and then shows a constant bias about 0.1 K. During the day, in clear sky, calm conditions, thermal stratification of the top few meters of the ocean will occur [35]. At high wind speeds, the ocean is sufficiently well mixed that H8 skin SST should be equal to the *in situ* measurements. For the nighttime, there is no insolation and thus the diurnal warming disappears. It shows a small negative
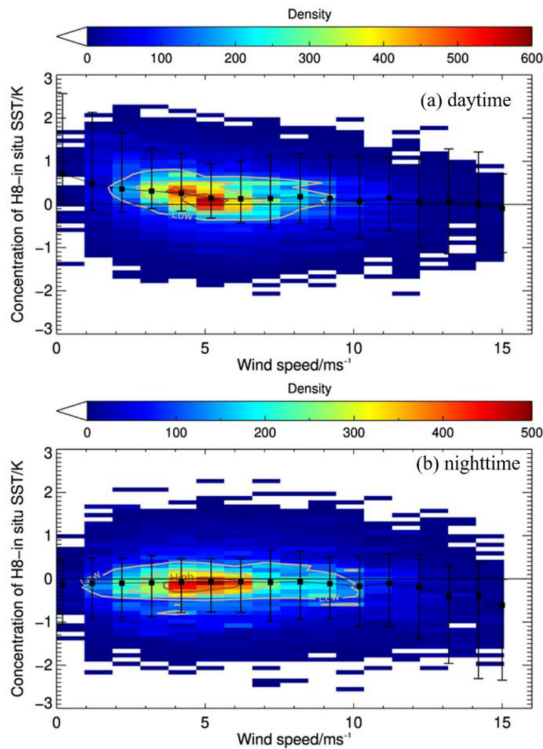
Fig. 10. Biases between H8 SST and *in situ* measurements as a function of wind speed for (a) daytime and (b) nighttime.
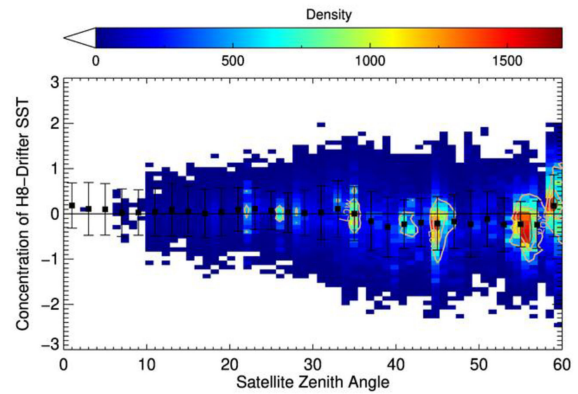

Fig. 11. Density plot as well as error bars of biases as a function of the SZA. The black squares show the median bias and the top whisker is 1.5 times either the IQR75 and the bottom whisker is 1.5 times either the IQR25.

bias ~ -0.1 K at the wind speed range from 0 to 11 m/s. This is mainly due to the cool skin effect. However, larger negative biases ~ 0.5 K appear when wind speed higher than 13 m/s. For these reasons, the errors of H8-SST/JAXA appear to be sensitive to the very low and very high wind speed. This implies that much attention should be paid to the effect of wind when using the H8-SST/JAXA.

However, there is a logistical and financial limit to the number of *in situ* infrared radiometer systems that can be deployed to provide a direct SST_{skin} validation network. The indirect validation using SST_{depth} measurements during nighttime or wind speeds greater than 6 ms^{-1} at daytime becomes a primary data source. The diurnal warming disappears and only the cool skin temperature bias needs to be corrected under these conditions. Additionally, the vertical and horizontal SST gradients will be limited within a satellite pixel due to wind mixing of the upper layers homogenizing the temperature structure, so minimizing discrepancies between point (from a buoy) or line source (from a ship) and spatially averaged measurements (from a satellite).

The spatial coverage of geostationary satellite images is extensive, with latitudinal and longitudinal limits of about $\pm 60°$, and the satellite zenith angle (SZA) of each pixel of H8 varies over a wide range. Therefore, it is necessary to investigate if there is any influence of the SZA on the accuracy of derived SSTs. Fig. 11 plots the density of the matchups as well as error bars of biases as a function of the SZA. To reduce the influence of other factors, only the best quality data and wind speed greater than 6 m/s are used, considering those reasons as mentioned above. It shows that negligible median bias errors close to 0 K are found

between SZAs 0° and 35°. In contrast to this, as SZA becomes larger, negative biases (~ -0.3 K) were detected between SZAs 35° and 58°. Larger negative biases could have been caused by cloud contamination because of incomplete cloud screening due to the lack of visible (3.9 μm) data. Additionally, at larger SZA, the larger atmospheric path length leads to greater attenuation of surface infrared emissions. And there is a greater chance of cloud contamination as SZA increases due to the field of view increases with SZA [36]. While for the highest SZA values, i.e., 58°–60°, a larger positive bias ~ 0.3 K appeared. Those can be also found along the viewing boundaries of the full disk as shown in Fig. 5.

V. CONCLUSION

The full disk of the H8-SSTs/JAXA have been validated against the *in situ* measurements from the iQuam2 for a period from July 2015 to June 2017. It is imperative to determine the statistics of satellite SST validation against different types of platforms. On the one hand, the measurements of the ship provide the validations in the Marginal Seas, where usually lack drifter or Argo observations. On the other hand, the drifter and Argo provide a complete coverage over the open ocean, whereas the ships are very sparse in the tropics, and especially in the Indian Ocean and the Southern Ocean. General error statistics show that the H8-SSTs/JAXA (version 1.2) have small biases of −0.03, −0.11, and −0.11 K with RMSE of 0.73, 0.58, and 0.64 K when comparing against ships, drifter, and Argo over the full disk, respectively.

The spatial distribution of errors shows that a small median bias close to 0.1 K and RMSE of 0.4–0.6 K comparing to drifters measurements were found between SZAs 0° and 35°, corresponding to the low-and middle-latitude of the west of 180°E over the Pacific Ocean. In contrast to this, as SZA becomes larger, negative biases (~ -0.3 K) were detected between SZAs 35° and 58°, corresponding to the northern Indian Ocean, Southern Ocean, and high-latitude of the Pacific Ocean between 180°E and 160°W. Larger positive biases exceeding 0.3 K are also found along the viewing boundaries, where SZAs near 60°. So it should be careful to use H8-SST/JAXA when SZA larger than 35°.

The temporal variations of the errors are explored by computing the difference between the H8-SSTs/JAXA and ships, drifter and Argo at monthly and hourly timescales. Overall, the median bias of drifter and Argo are of the order of −0.1 K, while the ships' fluctuated between −0.1 and 0.2 K. The monthly IQR of biases for ships was 0.83–0.98 K, 0.64–0.75 K for drifter and 0.63–0.78 K for Argo floats. There is no prominent seasonal variation during the period analyzed. The amplitude of the DV of the difference between H8-SST/JAXA and *in situ* measurements is only ∼0.1 K for the statistical of all matchups. The maximum errors appear in the morning rather than in the noon when considering the offset of the cool skin constant about 0.17 K.

It needs to be noted that the *in situ* measurements are not designed with the original purpose of complementing satellite data, therefore they do not necessarily represent the same temperature. The *in situ* collecting surface temperature data are very heterogeneous in depth, and the measurements are very heterogeneous at low wind speeds in clear sky. The diurnal and depth variation correction models are therefore needed in the future work for satellite SST_{skin} validation by using the *in situ* measurements at a certain depth. The statistics obtained from this study will be used in future work for DV analysis and merging purposes.
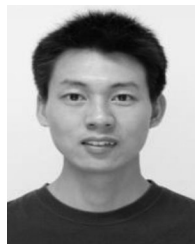
REFERENCES

[1] C. J. Merchant *et al.*, "2–Global sea surface temperature," in *Taking the Temperature of the Earth*, G. C. Hulley and D. Ghent, Eds. New York, NY, USA: Elsevier, 2019, pp. 5–55.

[2] C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, "The operational sea surface temperature and sea ice analysis (OSTIA) system," *Remote Sens. Environ.*, vol. 116, pp. 140–158, 2012.

[3] S. L. Castro, G. A. Wick, D. L. Jackson, and W. J. Emery, "Error characterization of infrared and microwave satellite sea surface temperature products for merging and analysis," *J. Geophys. Res. Oceans*, vol. 113, pp. 1–17, 2008.

[4] A. G. O Carroll, J. R. Eyre, and R. W. Saunders, "Three-Way error analysis between AATSR, AMSR-E, and In Situ sea surface temperature observations," *J. Atmos. Ocean. Technol.*, vol. 25, pp. 1197–1207, 2008.

[5] F. Xu and A. Ignatov, "Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals," *J. Geophys. Res., Oceans*, vol. 115, 2010.

[6] A. Alvera-Azcárate, C. Troupin, A. Barth, and J. Beckers, "Comparison between satellite and in situ sea surface temperature data in the Western Mediterranean sea," *Ocean Dyn.*, vol. 61, pp. 767–778, 2011.

[7] Q. Tu, D. Pan, and Z. Hao, "Validation of S-NPP VIIRS sea surface temperature retrieved from NAVO," *Remote Sens.*, vol. 7, pp. 17234–17245, 2015.

[8] F. Wu, P. Cornillon, B. Boussidi, and L. Guan, "Determining the pixel-to-pixel uncertainty in satellite-derived SST fields," *Remote Sens.*, vol. 9, p. 877, 2017.

[9] Y. Chao, Z. Li, J. D. Farrara, and P. Hung, "Blending sea surface temperatures from multiple satellites and in situ observations for coastal oceans," *J. Atmos. Ocean. Technol.*, vol. 26, pp. 1415–1426, 2009.

[10] W. Wang and P. Xie, "A multiplatform-merged (MPM) SST analysis," *J. Climate*, vol. 20, pp. 1662–1679, 2007.

[11] L. Guan and H. Kawamura, "Merging satellite infrared and microwave SSTs: Methodology and evaluation of the new SST," *J. Oceanogr.*, vol. 60, pp. 905–912, 2004.

[12] R. W. Reynolds and T. M. Smith, "Improved global sea surface temperature analyses using optimum interpolation," *J. Climate*, vol. 7, pp. 929–948, 1994.

[13] I. S. Robinson, *Measuring the Oceans From Space: The Principles and Methods of Satellite Oceanography*. New York, NY, USA: Springer, 2004.

[14] S. Morak-Bozzo, C. J. Merchant, E. C. Kent, D. I. Berry, and G. Carella, "Climatological diurnal variability in sea surface temperature characterized from drifting buoy data," *Geosci. Data J.*, vol. 3, pp. 20–28, 2016.

[15] C. Donlon *et al.*, "The global ocean data assimilation experiment high-resolution sea surface temperature pilot project," *Bull. Amer. Meteorol. Soc.*, vol. 88, pp. 1197–1213, 2007.

[16] C. A. Clayson and A. S. Bogdanoff, "The effect of diurnal sea surface temperature warming on climatological Air–Sea fluxes," *J. Clim.*, vol. 26, pp. 2546–2556, 2013.

[17] Y. Kawai and A. Wada, "Diurnal sea surface temperature variation and its impact on the atmosphere and ocean: A review," *J. Oceanogr.*, vol. 63, pp. 721–744, 1990-10-01 2007.

[18] E. Maturi *et al.*, "NOAA's sea surface temperature products from operational geostationary satellites," *Bull. Amer. Meteorol. Soc.*, vol. 89, pp. 1877–1887, 2008.

[19] J. Schmetz *et al.*, "An introduction to Meteosat second generation (MSG)," *Bull. Amer. Meteorol. Soc.*, vol. 83, pp. 977–992, 2002.

[20] H. Woo, K. Park, X. Li, and E. Lee, "Sea surface temperature retrieval from the first Korean geostationary satellite COMS data: Validation and error assessment," *Remote Sens.*, vol. 10, 2018, Art. no. 1916.

[21] J. Yang, Z. Zhang, C. Wei, F. Lu, and Q. Guo, "Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4," *Bull. Amer. Meteorol. Soc.*, vol. 98, pp. 1637–1658, 2017.

[22] H. Zhang, H. Beggs, L. Majewski, X. H. Wang, and A. Kiss, "Investigating sea surface temperature diurnal variation over the tropical warm pool using MTSAT-1R data," *Remote Sens. Environ.*, vol. 183, pp. 1–12, 2016.

[23] P. Le Borgne, G. Legendre, and S. Péré, "Comparison of MSG/SEVIRI and drifting buoy derived diurnal warming estimates," *Remote Sens. Environ.*, vol. 124, pp. 622–626, 2012.

[24] S. L. Castro, G. A. Wick, and J. J. H. Buck, "Comparison of diurnal warming estimates from unpumped Argo data and SEVIRI satellite observations," *Remote Sens. Environ.*, vol. 140, pp. 789–799, 2014.

[25] A. Ditri, P. Minnett, Y. Liu, K. Kilpatrick, and A. Kumar, "The accuracies of Himawari-8 and MTSAT-2 sea-surface temperatures in the tropical western Pacific Ocean," *Remote Sens.*, vol. 10, p. 212, 2018.

[26] Y. Kurihara, H. Murakami, and M. Kachi, "Sea surface temperature from the new Japanese geostationary meteorological Himawari-8 satellite," *Geophys. Res. Lett.*, vol. 43, pp. 1234–1240, 2016.

[27] F. Xu and A. Ignatov, "In situ SST quality monitor (iQuam)," *J. Atmos. Ocean. Technol.*, vol. 31, pp. 164–180, 2014.

[28] C. Donlon, I. S. Robinson, W. Wimmer, G. Fisher, M. Reynolds, R. Edwards, and T. J. Nightingale, "An infrared sea surface temperature autonomous radiometer (ISAR) for deployment aboard volunteer observing ships (VOS)," *J. Atmos. Ocean. Technol.*, vol. 25, pp. 93–113, 2008.

[29] P. Dash, A. Ignatov, Y. Kihai, and J. Sapper, "The SST quality monitor (SQUAM)," *J. Atmos. Ocean. Technol.*, vol. 27, pp. 1899–1917, 2010.

[30] W. J. Emery, D. J. Baldwin, P. Schlüssel, and R. W. Reynolds, "Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements," *J. Geophys. Res. Atmos.*, vol. 106, pp. 2387–2405, 2001.

[31] R. Yokoyama, S. Tanba, and T. Souma, "Sea surface effects on the sea surface temperature estimation by remote sensing," *Int. J. Remote Sens.*, vol. 16, pp. 227–238, 1995.

[32] M. Kramar, A. Ignatov, B. Petrenko, Y. Kihai, and P. Dash, "Near real time SST retrievals from Himawari-8 at NOAA using ACSPO system," in *Proc. Ocean Sens. Monit. VIII*, 2016, Paper 98270L.

[33] F. Xu and A. Ignatov, "Error characterization in iQuam SSTs using triple collocations with satellite measurements," *Geophys. Res. Lett.*, vol. 43, pp. 10826–10834, 2016.

[34] G. K. Corlett, C. J. Merchant, P. J. Minnett, and C. J. Donlon, "Chapter 6.2 - assessment of long-term satellite derived sea surface temperature records," in *Experimental Methods in the Physical Sciences*, vol. 47, G. Zibordi, C. J. Donlon, and A. C. Parr, Eds., New York, NY, USA: Academic, 2014, pp. 639–677.

[35] J. F. Price, R. A. Weller, and R. Pinkel, "Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing," *J. Geophys. Res.*, vol. 91, pp. 8411–8427, 1986.

[36] X. Li, W. Pichel, P. Clemente-Colon, V. Krasnopolsky, and J. Sapper, "Validation of coastal sea and lake surface temperature measurements derived from NOAA/AVHRR data," *Int. J. Remote Sens.*, vol. 22, pp. 1285–1303, 2001.

Qianguang Tu was born in Fujian, China in 1985. He received the B.S. degree in remote sensing science and technology from Nanjing University of Information Science and Technology, Nanjing, China, in 2009, and M.S. degree in physical oceanography from Second Institute of Oceanography SOA, Hangzhou, China, in 2012, and the Ph.D. degree in geographic information system from Zhejiang University, Hangzhou, China, in 2016.

Since 2016, he has been an Assistant Professor with the School of Surveying and Municipal Engineering, Zhejiang University of Water Resources and Electric Power. He has authored about ten articles. His research interests include remote sensing of short-term coastal ocean processes, remote sensing of water quality, and data assimilation for satellite products.

Zengzhou Hao received the B.S. degree in mathematics and the Ph.D. degree in meteorology from the Nanjing University of Information Science and Technology, Nanjing, China, in 2002 and 2007, respectively.

He is currently with the Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. His research interests include the atmospheric and oceanic environment remote sensing from optical and thermal observations, satellite data analysis, quality control, and data assimilation.