

Article

Assessment of Extreme Diurnal Warming in Operational Geosynchronous Satellite Sea Surface Temperature Products

Gary A. Wick ^{1,*} and Sandra L. Castro ²

¹ NOAA Physical Sciences Laboratory, R/PSD2, 325 Broadway, Boulder, CO 80305, USA

² Colorado Center for Astroynamics Research, University of Colorado, 431 UCB, Boulder, CO 80309, USA; sandrac@colorado.edu

* Correspondence: gary.a.wick@noaa.gov; Tel.: +1-(303)-497-6322

Received: 21 October 2020; Accepted: 13 November 2020; Published: 17 November 2020



Abstract: We evaluate the reliability and basic characteristics of observations of extreme DW events from current operational geostationary satellite sea surface temperature (SST) products through examination of three months of diurnal warming (DW) estimates derived by different methodologies from the Spinning Enhanced Visible and Infrared Imager on Meteosat-11, Advanced Himawari Imager on Himawari-8, and Advanced Baseline Imager on the Geostationary Operational Environmental Satellite (GOES)-16. This work primarily focuses on the following research questions: (1) Can these operational SST products accurately characterize extreme DW events? (2) What are the amplitudes and frequencies of these events? To answer these, we compute distributions of DW and DW exceedance and compare them amongst the different methods and geostationary sensors. Examination of the DW estimates demonstrates several challenges in accurately deriving distributions of the DW amplitude, particularly associated with estimating the “foundation” temperature and uncertainties in cloud screening. Overall, the results suggest that current geostationary sensors can reliably assess extreme DW, but the estimates are sensitive to the computational methods applied. We thus suggest careful processing/screening of the SST retrievals. We find a value of 3 K, corresponding to the 99th percentile, provides a potential practical threshold for extreme warming, but events of at least 6 K were reliably observed. Warming in excess of 6 K occurred somewhere an average of 47% of the time, and its probability at a given location was of $O(10^{-6})$.

Keywords: diurnal warming; sea surface temperature; satellite product reliability; extremes; exceedance probability

1. Introduction

The daily cycle of solar radiation leads to periodic heating of the near-surface layer of the ocean. At low wind speeds, turbulent mixing is reduced and a warm layer and diurnal thermocline can form near the ocean surface during the day. At night, mixing typically erodes this layer. While the amplitude of diurnal warming is relatively small on average ($O(0.5$ K), e.g., [1,2]), under conditions of very low wind speed and sufficient insolation, the warming at the surface sensed by satellites can be highly significant (e.g., [3,4]). In situ observations have shown warming in excess of 5 K at depths of 0.3–0.6 m [5]. Satellite observations from multiple sensors have observed extreme warming events up to 7 K in magnitude at the surface, and it has been suggested that events exceeding 5 K are not infrequent [3,4].

Large diurnal variations of the sea surface temperature (SST) are important to observe and understand for multiple reasons, including the production and validation of satellite-based SST

analyses and accurate characterization of air–sea interaction processes. Present SST analyses (level 4 products) attempt to blend data from multiple satellite sensors with different measurement times and different effective measurement depths. To create a blended product representative of either a specific time and depth or a daily value representative of a subsurface depth free from any diurnal warming (the “foundation” temperature; e.g., see [6]), it is necessary to compensate for the different amount of diurnal warming present in each individual retrieval.

Additionally, the effect of diurnal temperature variations on lower frequency weather and climate variability is increasingly being explored and demonstrated. Yang and Slingo [7] suggest that the diurnal cycle is one of the most fundamental modes of variability in the climate system. Perhaps most directly, diurnal SST variations have been shown to directly impact calculations of the air–sea heat flux (e.g., [8–12]). Clayson and Bogdanoff [13] estimated that neglecting diurnal SST variations could lead to globally averaged flux differences of $\sim 4.5 \text{ W m}^{-2}$ and up to 200 W m^{-2} at specific times and locations. Several studies have further shown that inclusion of diurnal forcing has an impact on mixed layer dynamics and the intraseasonal evolution of the SST (e.g., [14–17]). Other studies have suggested the important role of diurnal variability in predictability of the Madden–Julian oscillation (e.g., [18,19]) and on characteristics of the El Niño–Southern Oscillation (ENSO) (e.g., [20,21]).

The current generation of geostationary satellites with improved radiometric capabilities and increased spatial and temporal resolution provide the opportunity for significantly enhanced characterization of diurnal warming (DW) at the ocean surface (e.g., [22,23]). These capabilities began with the launch of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on the European Meteosat Second Generation (MSG) satellites. Increased regional coverage of the observations has become available through the Japanese Advanced Himawari Imager (AHI) on the Himawari satellites and, most recently, the Advanced Baseline Imager (ABI) on the U.S. Geostationary Operational Environmental Satellite R (GOES-R) series. Polar-orbiting satellites have had the required radiometric capabilities for some time, but with repeat sampling at a given location constrained to at most a couple of times per day at lower latitudes, they are unable to fully resolve the diurnal temperature variability.

The general ability to accurately estimate the amplitude of diurnal warming using SEVIRI has been broadly demonstrated (e.g., [3,4,22–24]). Estimation of a DW amplitude requires computation of a day–night SST difference. The most common approaches have utilized initial determination of the foundation temperature (e.g., [23,24]) and then taken the DW as the difference between instantaneous daytime SST retrievals and that foundation temperature. Operational SST retrievals from Himawari-8 and GOES-16 are newer, and while the initial capability to resolve diurnal variability has been documented [25], analysis of the observed warming has been more limited thus far.

Despite these successful results, the ability of operational SST products to accurately resolve the most extreme diurnal events ($> \sim 3 \text{ K}$) deserves additional attention, particularly on an individual satellite and product basis. With the exception of [3,4], previous studies have largely focused on obtaining estimates of average DW values. Gentemann et al. [3] demonstrated observations of large DW events from SEVIRI were consistent with coincident observations from other satellites, but this analysis primarily examined distinct, coherent, larger-scale events. Merchant et al. [4] derived estimates of the frequency and scale of large DW events from SEVIRI but, while noting the potential for spurious warming from noisy pixels, did not fully document just how likely such spurious warming might be. LeBorgne et al. [22], through comparisons against drifting buoys, found that the uncertainty in SEVIRI DW estimates increased from $\sim 0.4 \text{ K}$ for an average of all events to nearly 0.7 K for events with amplitudes greater than 2.5 K (their threshold for large DW events), but the number of available events was relatively small. While previous research evaluating different methodologies for estimating the foundation temperature from SEVIRI found the derived DW to be largely independent of the specific foundation approach [23], at least for more typical DW amplitudes, additional evaluation seems warranted for the most extreme events. As will be shown here, direct examination of derived diurnal warming maps suggests the existence of a broader distribution of large-amplitude events that could be potentially misinterpreted if not studied in more detail.

The purpose of this paper is to evaluate the reliability and basic characteristics of observations of the largest DW events from current operational geostationary satellite SST products. Fundamentally, we seek to answer two overarching questions: (1) Can operational SST products from current geostationary sensors accurately characterize the most extreme DW events? (2) What are the amplitudes and frequencies of these events (or, in other words, how large is extreme)? To answer these questions, we must assess the quality of the observations and the details of the methods employed in the DW calculations. Specifically, we must consider whether peak retrieved DW estimates are representative of physical DW maxima as opposed to being inordinately influenced by noisy or otherwise contaminated observations, and we must also consider to what extent the specific methods for deriving DW influence the estimates of the most extreme warming (i.e., how the computational method impacts the tail of the probability distribution corresponding to the extremely high DW).

The work supports two primary research and operational priorities: the development of climatologies of peak DW amplitude and the validation of numerical model predictions of large DW. Previous efforts have generated and examined climatologies of the average DW (e.g., [1,2]) using data from multiple sources, but a complete characterization of the largest, and rarest, DW events is still lacking. The accuracy of the observations and methods used to estimate the peak amplitudes must be firmly established. Physically based models of DW are key tools for incorporating DW information into SST analyses and conducting air–sea interaction process studies as described above. As one example, the National Oceanic and Atmospheric Administration (NOAA) is presently integrating physically based model estimates of the DW into the computation of their new blended SST analysis [26]. Estimates of the DW present in individual satellite retrievals are subtracted prior to blending into a daily measure of the foundation temperature. These corrections are most important for the largest DW amplitudes, and yet the accuracy of the models in these extreme cases is poorly understood.

To address these objectives, we examine three months of DW estimates derived by different methodologies from three current operational geostationary sensors. The sensors and specific operational SST products are described in Section 2. The approaches for estimating the DW amplitudes, including different methods for deriving a foundation temperature, are introduced in Section 3. The results, expressed largely in terms of frequency distributions of the derived DW, are presented and compared in Section 4. The implications of the results are then discussed in Section 5 relative to the questions posed above, followed by overall conclusions in Section 6.

2. Satellite-Based SST Input Products

Diurnal warming estimates were derived from operational SST retrievals from the SEVIRI on Meteosat-11, AHI on Himawari-8, and ABI on GOES-16 over the period of June–August 2019. The spatial extent of these data is shown in Figure 1 (note the region of overlap between Meteosat-11 and GOES-16). While an ABI was also flying on GOES-17 during this period, sensor cooling issues degraded the quality of the data, and we did not consider the corresponding SST product. These products represent the current operational SST retrievals from the latest generation of geostationary satellites in most common use at NOAA.

SEVIRI SST retrievals from Meteosat-11 were produced by the Meteo-France/Centre de Météorologie Spatiale (CMS, Lannion, France) within the framework of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI-SAF) and were obtained directly from IFREMER (The French Research Institute for Exploitation of the Sea). The hourly, 5-km (0.05°) gridded data acquired and employed here were originally derived by averaging data obtained at full resolution every 15 min [27]. The SST product was generated using a nonlinear split-window algorithm (e.g., [28]) processed through the OSI SAF geostationary SST chain [27] and incorporates a numerical weather prediction (NWP) model-derived adjustment following Le Borgne et al. [29]. The native observations are representative of the skin temperature, but the product incorporates an adjustment to estimate the temperature at a depth

of roughly 20 cm, representative of what would be sampled by a drifting buoy. Centered over 0° longitude, SEVIRI provides spatial sampling over a grid from about 60° N to 60° S and 60° W to 60° E.

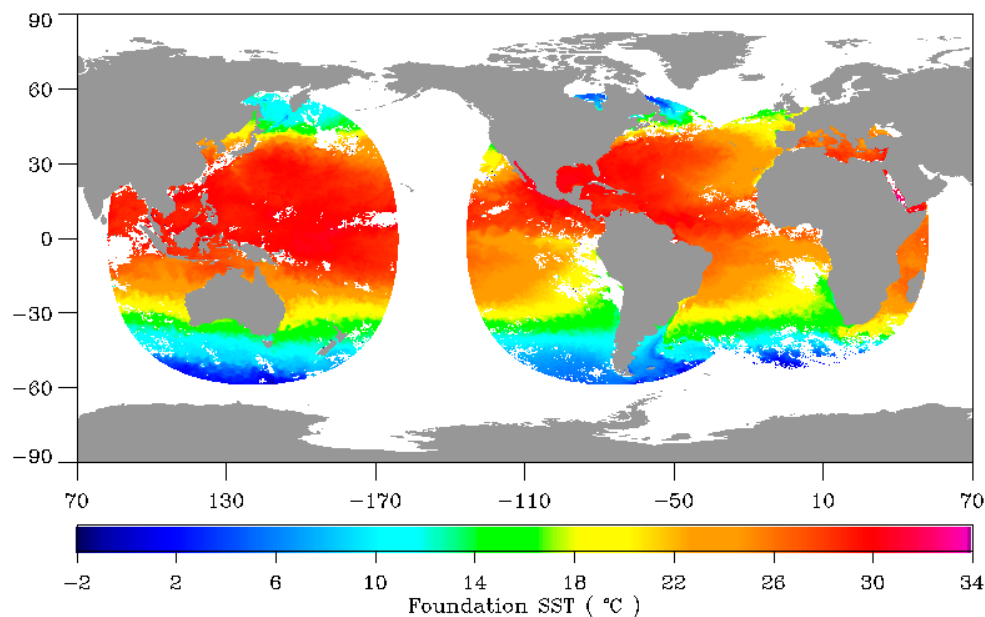


Figure 1. Derived foundation sea surface temperature (SST) values for 10 August 2018 illustrating the spatial extent of operational SST retrievals from the Himawari-8 AHI (left), GOES-16 ABI (center), and Meteosat-11 SEVIRI (right). The methodology for computation of the daily foundation SST is described in Section 3.

Diurnal warming estimates for the Himawari-8 AHI and GOES-16 ABI were derived from the operational SST retrievals produced by the Center for Satellite Applications and Research (STAR) at the NOAA National Environmental Satellite and Data Information Service (NESDIS). The SST retrievals were generated through the NOAA Advanced Clear-Sky Processor for Oceans (ACSPO) system adapted to geostationary data as described by Petrenko et al. [25]. The algorithm is based on regression to an analyzed SST product, which is representative of a subsurface temperature, but only nighttime observations were used in the derivation so that any detected diurnal variability would be representative of the skin temperature as actually sensed by the satellite. Hourly, 0.02° spatial resolution values were obtained from the Version 2.70, level 3 collated products. Within the collation procedure, hourly values are derived from 15-min full disk scans using a technique that attempts to provide a best fit to the true diurnal cycle while attempting to improve filtering for any residual cloud contamination in the individual observations. The Himawari-8 AHI is centered over 140.7° E longitude, sampling a disk extending from approximately 83° E to 162° W. The GOES-16 ABI is centered over 75.2° W longitude, providing longitudinal coverage from approximately 134° W to 16° W. The products provide latitudinal coverage from about 59° N to 59° S for both sensors.

All the retrieved SST products were provided with ancillary data consistent with specifications from the Group for High Resolution SST (GHRSSST) [6]. These include coincident wind speeds, uncertainty estimates, and quality level indices indicating the likelihood of potential cloud contamination. The quality indices range from 0 to 5, with 5 representing the highest quality, most confidently clear retrievals. The expected uncertainty of the individual SST retrievals ranges between approximately 0.4 and 0.6 K depending on the specific product and selected quality levels [22,23,25].

3. Estimation of the Diurnal Warming Amplitude

The DW present at a point in time is given by the difference between the instantaneous SST value and the temperature that would have been present had there been no heating from solar radiation.

This concept of an SST in the absence of any diurnal warming is precisely the definition of the foundation temperature, SST_{found} , as adopted within the Group for High Resolution SST (GHRSSST) [6]. As such, we calculate DW as

$$DW_{\text{inst}} = SST_{\text{inst}} - SST_{\text{found}} \quad (1)$$

where DW_{inst} is the instantaneous DW, and SST_{inst} is the instantaneous SST retrieval. The DW_{inst} is computed for every hourly geostationary satellite image at every grid point at which cloud-free estimates of SST_{inst} and SST_{found} are available. The daily DW amplitude, DW_{amp} , at each grid location is then taken as the maximum of the corresponding hourly DW_{inst} values within a daily cycle, usually defined between 6 a.m. local solar time (LST) on the current day and 6 a.m. LST on the following day. Monthly and seasonal maps of peak diurnal warming are also compiled as the maximum of all the DW_{amp} values over the selected period at each grid cell.

The SST_{inst} values were taken directly from each of the operational products presented in Section 2 at the native product resolution. Since we are primarily concerned with the validity and characteristics of the most extreme DW events and peak values as opposed to mean-value statistics, we accepted a greater risk of residual cloud contamination in the instantaneous daytime SST values (by incorporating data with lower quality indices) in exchange for enhanced spatial coverage and inclusion of retrievals that might have been incorrectly flagged as cloudy. Cloud contamination in the daytime SST retrievals tends to reduce the DW amplitude. The data quality flagging criteria vary with data producer, and previous studies [23] have shown a tendency for some operational systems to incorrectly flag pixels affected by DW as of lower quality. Here, we considered all values with quality indices of 3 or higher for the daytime SST_{inst} . The benefit of this approach over limiting the observations to the highest quality level 5 is demonstrated briefly in Section 4.1. Restricting the observations to quality level 5 only notably reduces the number of available observations but does not impact our fundamental conclusions. During nighttime hours, however, when residual cloud contamination would generally increase the DW amplitude, only retrievals with quality level 5 were used in SST_{inst} .

While the value of the instantaneous daytime SST is straightforward, estimation of the foundation temperature is more complex, as no direct measurement of this conceptual temperature exists. The SST_{found} is broadly assumed to be consistent with the SST just prior to dawn, but given the potential for cloud contamination of individual SST retrievals and other uncertainties in the retrievals, multiple approaches to estimating SST_{found} from available observations have been developed and tested (e.g., [23,24]).

In this work, we evaluate and compare estimates of extreme DW amplitudes obtained using several different methods for deriving SST_{found} . The different methodologies, based largely on those previously tested, consider inclusion of data of heterogeneous quality, different time periods for data accumulation, and different compositing methods over various spatial resolutions. Construction of the foundation temperature estimate represents a tradeoff between data coverage and the best representation of the foundation temperature for the given day and location. Whereas residual cloud contamination was less of a concern in the daytime SST_{inst} values, derived peak DW is highly sensitive to any cloud influences on SST_{found} . The cold bias introduced by cloud-contaminated SST retrievals would lower the estimated SST_{found} , leading to overestimation of the DW amplitude. A priority in the development and evaluation of the foundation methodologies was to minimize any potential cloud influence. Overestimates of SST_{found} were viewed as preferable to underestimates for the purpose of obtaining conservative (i.e., not overstated) estimates of peak DW amplitudes.

The first method evaluated for SST_{found} , termed “QC5” for the accepted quality control (QC) levels, is the closest to a single predawn value and most closely follows the approach advocated by Karagali and Hoyer [23] and later adopted by Zhang et al. [30] without further evaluation. This method emphasizes potential data quality and point consistency over coverage extent. At each grid cell, SST_{found} is taken as the mean of all retrievals with quality index 5 between the local solar hours of midnight and 4 a.m. prior to dawn on the day to be analyzed. The mean is used rather than a minimum value to minimize the potential impact of any residual cloud contamination. Limited studies were also

conducted with a modified version of this approach, “QC45”, which considered all retrievals of quality indices 4 or 5.

The second method, termed “multiday”, in contrast, places a greater emphasis on spatial coverage as well as spatial and temporal consistency of the derived SST_{found} fields. This is most consistent with the approach adopted by Castro et al. [24]. Consideration of this methodology was also important because of its potential extension to use with polar-orbiting SST sensors which provide only a single nighttime SST retrieval. Multiple steps are employed in this procedure. Initially, for each satellite scene with observations between midnight and 6 a.m. LST, the median of all quality level 5 SST retrievals within 5×5 spatial subarrays of the image is determined. The pixel-wise mean of these downscaled images between midnight and 6 a.m. is then computed to provide a “nighttime” mean. Finally, the median of the nighttime means (again on a pixel-by-pixel basis) is computed over a running window of five days centered on the desired day. Prior to estimation of DW_{inst} , the downscaled foundation product is bilinearly interpolated back to the full resolution of the SST retrievals. The resulting product is less noisy and minimizes the impact of single cloud-contaminated retrievals. Because of the spatial smoothing incorporated, however, caution must be exercised in regions of strong SST gradients.

A third method attempted to combine the relative strengths of these two approaches. Termed “merge”, the approach first takes the QC5 foundation value where available and then fills in missing locations with the multiday estimate.

The derived DW_{amp} estimates are named by the corresponding approaches to obtaining SST_{found} . An additional DW_{amp} estimate is derived by applying additional filtering to the estimates obtained using the multiday foundation method. These estimates, termed “filter”, were obtained using only DW_{inst} values where the coincident wind speed, as obtained from the ancillary fields provided with the SST products, was less than 3 m/s. Extreme DW should occur only at the lowest wind speeds, and thus this method provided an additional way to examine the potential impact of noisy retrievals on perceived large DW events.

Validation of the derived DW estimates would ideally be conducted using direct in situ observations, but few observations corresponding to the largest DW events observed by satellites exist. While we can compare the consistency and characteristics of the different estimates described above, it is desirable to test the different products against a reference value in which we have greater confidence. To accomplish this, we generated one more DW estimate based on the continuity of the time series and diurnal cycle of SST retrievals at a given location.

Daily time series of the SST at each grid cell were constructed, and instances were extracted where the sequence of cloud-free retrievals from near midnight through mid-afternoon of the following day were largely continuous and smooth, exhibiting a realistic diurnal cycle. While the previous methods required no explicit consistency of the observations at different times throughout the day, the presence of a consistent, continuous diurnal cycle gives increased confidence in each of the individual retrievals comprising that cycle. For these instances, a DW reference estimate, termed “profile”, was computed as a peak-minus-trough difference from a sinusoidal function fit through the hourly SST retrievals comprising the identified diurnal cycle. Since the sequence of observations was used in the estimate, allowing for further identification of potential cloud contamination, quality indices of both 4 and 5 were considered during the nighttime hours. While the number of such cases is inevitably smaller in number, the expectation was that extreme DW events would largely correspond to extended periods of clear sky and, thus, be captured with this methodology.

4. Analysis and Assessment of Derived Extreme Diurnal Warming

An illustration comparing the monthly peak DW_{amp} for June 2019 computed using the different methods just described for the GOES-16 ABI is shown in Figure 2. June was selected for coincidence with the summer solstice in the Northern Hemisphere. At first glance, the results generally seem consistent with what might be expected, and the products are broadly similar in appearance. Monthly peak DW amplitudes extend up to values near 5 K and exhibit spatial coherence over a relatively

large portion of the summer hemisphere. Closer examination, however, highlights certain features worthy of additional consideration. Several of the foundation-based DW estimates suggest regions of large diurnal warming in the Southern Hemisphere that are not expected at that time of winter (given the shorter days and reduced insolation) and that are not observed in the reference profile product. Overall monthly peak DW amplitudes also appear larger in the foundation-based products than in the profile product, emphasizing the need to carefully assess these most extreme amplitudes. Similar results are also observed (not shown) for the other months and satellites.

To examine the validity of the large DW_{amp} estimates and address the questions posed in the introduction, we conducted a series of analyses on the DW amplitudes and resulting distributions obtained from the different methods. This section presents the results of these analyses based largely on intercomparison of the different products and examination of their key differences.

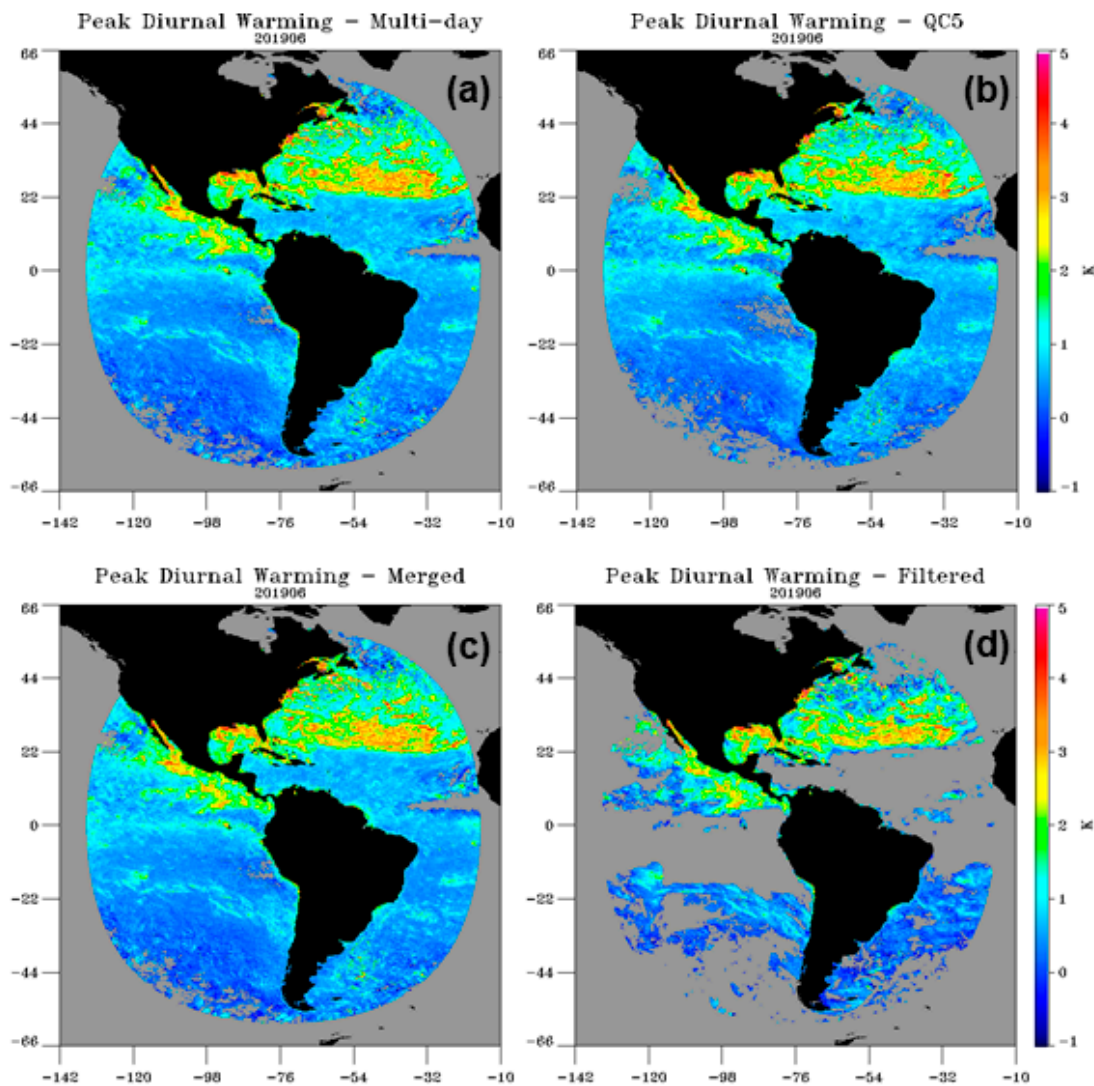


Figure 2. Cont.

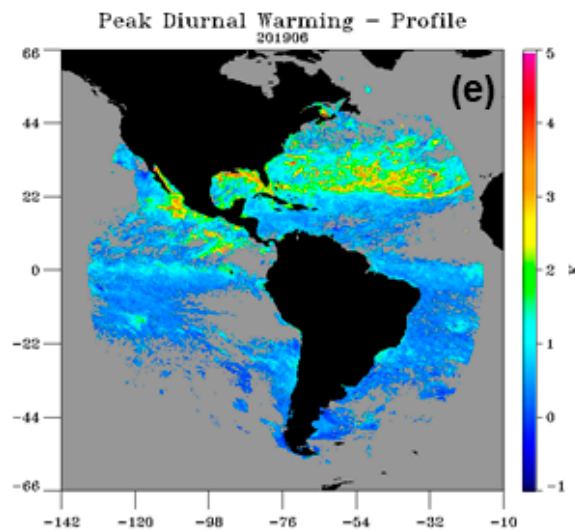


Figure 2. Monthly composited peak DW_{amp} for the month of June 2019 from the GOES-16 ABI as computed using the different methods evaluated: (a) multiday, (b) QC5, (c) merge, (d) filter, and (e) profile. See text for definition of the methods. Note, in particular, different suggestions of whether large diurnal warming (DW) exists in the south Atlantic and different peak DW_{amp} in the north Atlantic.

4.1. Challenges in Identifying Extreme Diurnal Warming

Several results highlight the challenges in estimating the largest DW amplitudes (and their associated distributions) and demonstrate why care must be used in interpretation of the results. The issues are primarily related to the potential for residual cloud contamination and other noise in the retrievals. Since the DW is derived from a difference in values, uncertainties in individual retrievals can be amplified in the result. Nevertheless, some of the apparent uncertainties in the derived magnitudes exceed what would be expected from the SST product accuracies assumed to be on the order of 0.5 K.

A first challenge is attempting to capture all the large DW events in the data while not biasing the results by overinclusion of cloud-contaminated observations. This problem is directly related to the choice of what quality indices to include in the analysis. An example comparing the Meteosat-11 SEVIRI DW_{amp} computed for a single day, 1 July 2017, with selected methodologies is shown in Figure 3. When estimation of the SST_{found} is limited to the use of only the highest quality data, a potential large DW event in the Mediterranean visible using the multiday approach (Figure 3b) is not captured in the QC5 results (Figure 3a). The fact that the profile methodology also detects the event (Figure 3c) lends confidence to the presence of warming on that day. While the natural tendency is to restrict analyses to data with the highest confidence levels, to be able to capture all the significant warming events and their correct frequency of occurrence, it may be appropriate, at least for SEVIRI, to reassess the validity of data that may have been flagged with reduced quality indices. Castro et al. [24] also argued that the quality flags might mask DW events, and LeBorgne et al. [22] utilized quality flags of 3–5 in their study of DW using SEVIRI. This issue appears to be most significant for SEVIRI, as the vast majority of useful NOAA V2.70 retrievals for Himawari-8 and GOES-16 appear to have a quality index of 5.

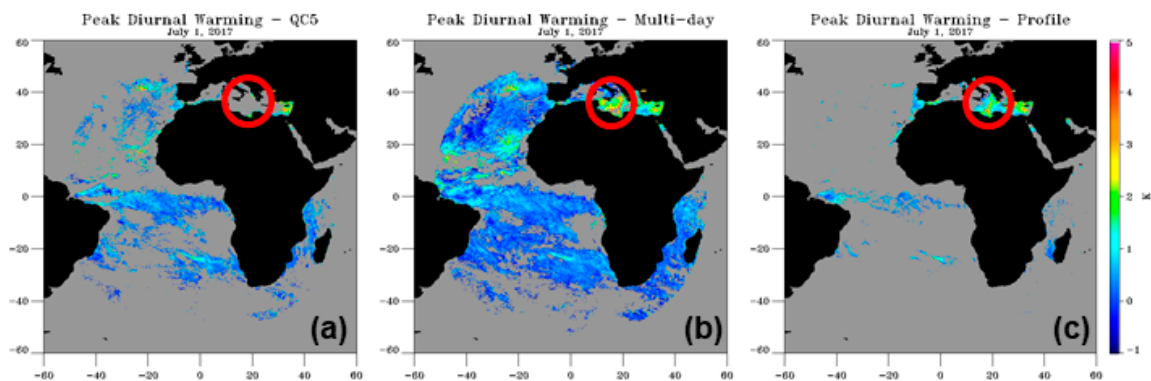


Figure 3. Comparison of the spatial coverage of the computed DW_{amp} for the Meteosat-11 SEVIRI on 1 July 2017 using the (a) QC5, (b) multiday, and (c) profile methods. The red circle, in particular, highlights a region in the Mediterranean Sea where using only retrievals with the highest quality flag misses a large DW event.

While some possibly valid data may be flagged with reduced-quality indices, there is also the potential for undetected cloud contamination in the highest quality-level data. For each of the sensors, the hourly distribution of DW_{inst} on 21 June 2019 derived using the QC5 methodology was examined and is displayed in Figure 4 using box and whisker plots. The results show the spread of DW_{inst} estimates from throughout the satellite's full disk broken down by local solar time (LST). While the mean diurnal signal is consistent with expectations and the envelope of hourly DW values exhibits a reasonable diurnal cycle pattern, many of the individual observations fall outside the range of physically expected values. First, observations of DW_{inst} in the predawn hours show outliers (circles indicating values in excess of the 75th percentile) in excess of 3 K. Such large residual amplitudes just prior to dawn are not realistic and suggest some corresponding uncertainty in the observations extending to 8 K at the peak of the diurnal cycle. Elevated amplitudes can occur as a result of the foundation temperature being underestimated, such as by incorporation of cloud-contaminated observations. The jump in the extent of the negative outliers for Himawari-8 and GOES-16 to include increasingly negative DW estimates in the nighttime hours (see values to -4 K after sunset, $\sim 18:00$ LST, in Figure 4a,b) reflects increased difficulty in cloud screening when visible reflectance data are not available. Since only the highest quality data were used for both the foundation and SST_{inst} values at night, the abrupt jump and large negative DW values in excess of cooling physically expected during a single day suggest that some residual cloud contamination does exist, at least for the Himawari-8 and GOES-16 SST retrievals. Values influenced in this way will then adversely depress the SST_{found} estimates.

Issues directly related to the method of estimating the foundation temperature may also exist. Comparison of the different SST_{found} estimates reveals differences that could influence interpretation of the largest DW_{amp} values. Differences in SST_{found} estimates computed using the QC5 and multiday methodologies were accumulated over the full three-month period for each of the sensors, and corresponding distributions are shown in Figure 5. Extreme differences in magnitude, exceeding 2 K, are observed in each case. Differences of these magnitudes are certainly significant relative to the scale of DW amplitudes and should be considered when assessing the reliability of the derived DW estimates. While these largest differences occur infrequently, the occurrence frequencies are similar to those of the most extreme DW_{amp} values that will be discussed below. For SEVIRI, differences in SST_{found} obtained using the QC5 and QC45 methods show smaller differences resulting solely from the inclusion of different quality levels, but differences in excess of 1.5 K are still observed.

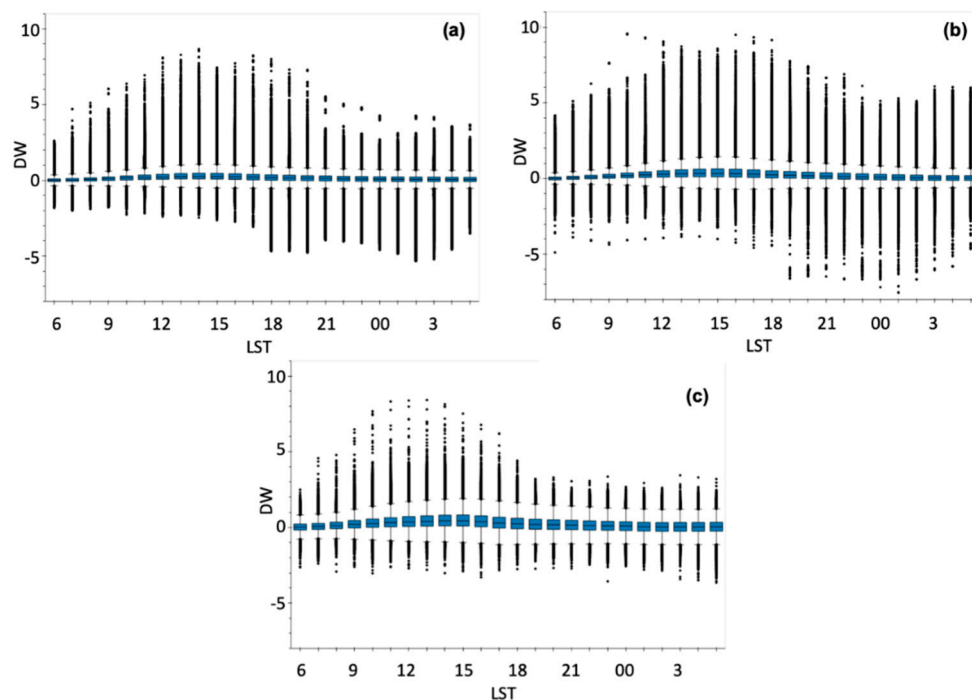


Figure 4. Box and whisker plots showing the distribution of DW_{inst} computed using the QC5 methodology as a function of local solar time (LST) for the (a) Himawari-8 ABI, (b) GOES-16 ABI, and (c) Meteosat-11 SEVIRI for 21 June 2019. Individual estimates were obtained for all available spatial grid points across the respective satellite disks.

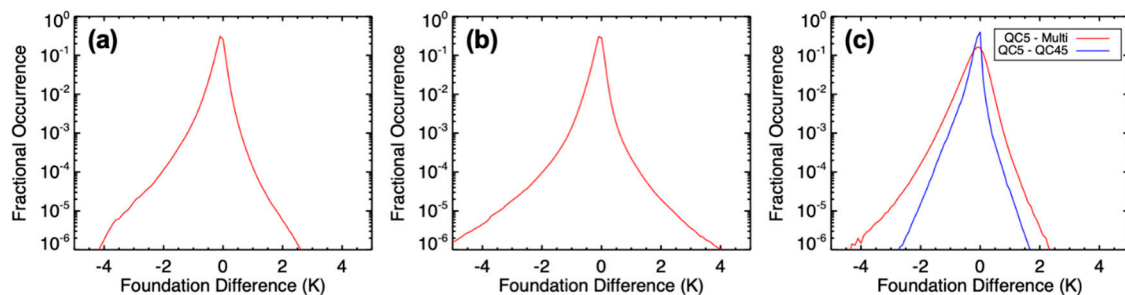


Figure 5. Distribution of differences in the SST_{found} values obtained from different methods accumulated over the full three-month period for the (a) Himawari-8 ABI, (b) GOES-16 ABI, and (c) Meteosat-11 SEVIRI. The red trace in each panel shows the difference between the QC5 and multiday methods, while for Meteosat-11 the additional blue trace represents the difference between the QC5 and QC45 methods.

The impact of differences in the SST_{found} can also be observed in a different context. Differences in the SST_{found} between successive days at each grid point were computed and then averaged over the month of June 2019. The results are presented in Figure 6 for the QC5 and multiday approaches. The results represent the average daily change in the subsurface temperature and can be compared with the DW values to see the impact of DW on the oceanic upper layer heat content. The differences are dramatic, and what is immediately apparent is that the QC5 SST_{found} product is noisier and much more variable on a daily basis. The multiday SST_{found} product exhibits more uniform day-to-day evolution that is more consistent with expected daily mean subsurface warming (or cooling) for that time of year. This follows from the temporal smoothing employed in the methodology. Even so, some localized anomalies exist, suggesting notable variability still exists in the daily values.

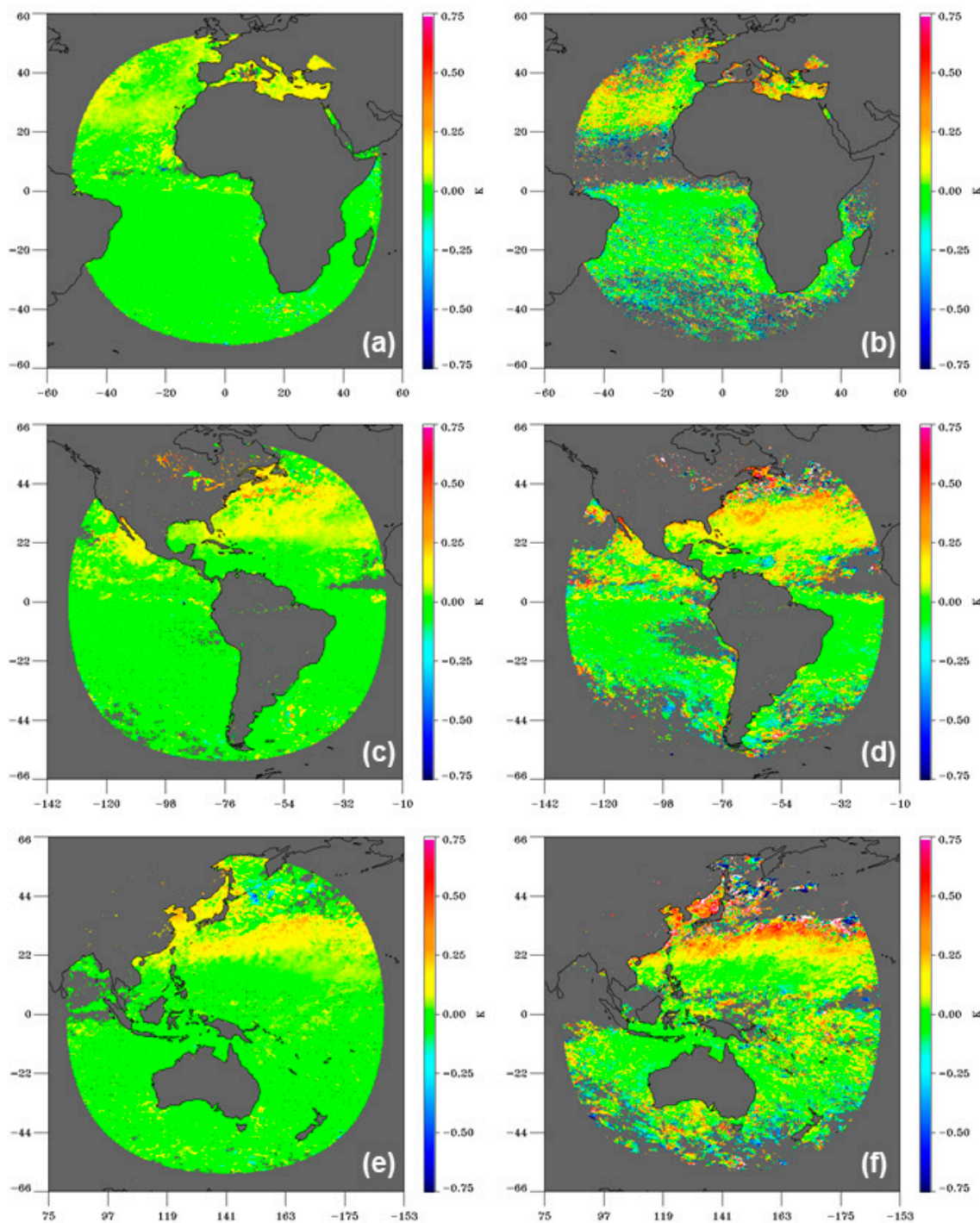


Figure 6. Averaged one-day differences in the SST_{found} computed using the (left: **a,c,d**) multiday and (right: **b,d,f**) QC5 methods for the (top: **a,b**) Meteosat-11 SEVIRI, (middle: **c,d**) GOES-16 ABI, and (bottom: **e,f**) Himawari-8 AHI, for the month of June 2019.

4.2. Comparison of Derived DW Distributions

Recognizing these potential issues influencing the accuracy of the DW estimates, we compare the distribution of DW_{amp} computed with the different methods. Monthly distributions were derived for each sensor over the full extent of the satellite disk. Examples of the total distributions of DW_{amp} are shown in Figure 7 for June 2019. The occurrence frequency is computed relative to the total number of potential DW_{amp} retrievals, which is given by the product of the number of ocean pixels and the number of days analyzed. For this computation, a pixel was defined as oceanic if there was at least

one valid SST retrieval at that point during the three months. Peak amplitudes up to about 8 K are inferred with each method derived using SST_{found} estimates. While the distributions are generally similar for DW amplitudes from about 0.5 K up to 4 K, the probability of occurrence $DW_{amp} > 4$ K diverges depending on the methodology used for SST_{found} . In comparison, the DW_{amp} distribution obtained using the reference profile methodology reaches maximum values well below the other methods for Himawari-8 and GOES-16 (~6 K and 7 K, respectively), and the frequency of the largest events is reduced. The peak occurrence frequencies vary between methods as a result of differences in the spatial density of derived amplitudes. Overall, reduced frequencies for the profile and filtered methodologies are the result of the fact that DW_{amp} estimates are only obtained at a smaller subset of the available locations where the required conditions are met.

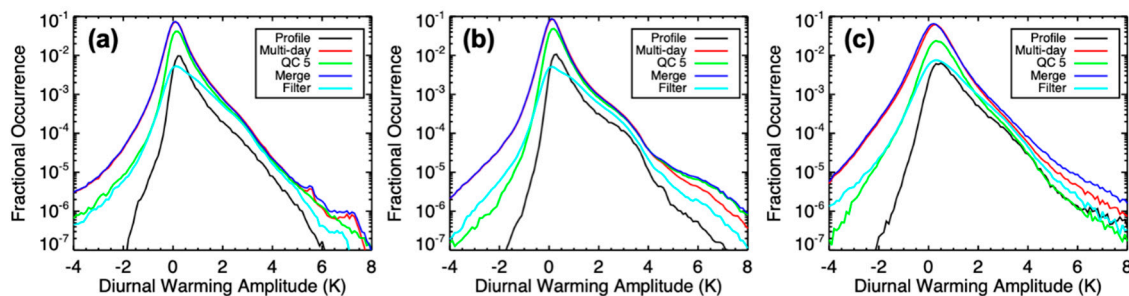


Figure 7. Comparison of the total distributions of the DW_{amp} during June 2019 computed using the different evaluated methodologies (see legend) for the (a) Himawari-8 ABI, (b) GOES-16 ABI, and (c) Meteosat-11 SEVIRI. Frequency is presented as the fractional occurrence computed as the number of observed values divided by the total number of potential open ocean retrievals.

A notable difference in the total distributions is the frequency of derived negative DW_{amp} values. Larger and more frequent negative amplitudes are obtained with the foundation-based methodologies, particularly those incorporating the multiday approach. The reduced negative amplitudes computed using the reference methodology are more consistent with physically expected values, and they reflect the associated, more stringent, data consistency requirements. The enhanced negative amplitudes in the other methods result from designed efforts to minimize any cloud contamination so as to reduce the likelihood of any overestimations of the DW and obtain realistic upper limits on SST_{found} . While the larger negative amplitudes are not physical, we accept their presence in our results based on the assumption that the larger positive DW_{amp} estimates then reflect a conservative, lower bound on the peak values.

To assess the differences in, and veracity of, the largest DW_{amp} values in more detail, cumulative distributions of diurnal warming exceedance values were generated and are presented in Figure 8. The results reflect the percentage of available observations with computed DW greater than the indicated values. The most immediate result is that, while there is good agreement, and thus enhanced confidence, in the frequency of occurrence of DW_{amp} estimates up to about 4 K, there is increased uncertainty in the number of events with greater amplitudes. The existence and commonality of events approaching 8 K are particularly uncertain. While these events, if present, would be very rare, it is still important to understand whether they are physically occurring at all as events of this magnitude would have disproportionately large impacts on the ocean and the atmosphere.

Assuming that the reference profile methodology captures the largest DW events and is our most reliable satellite-based estimate, we can make inferences about the validity of the DW_{amp} estimates from the different methodologies. There is good confidence in the physical occurrence of DW values up to at least 6 K in all domains over the months examined. In general, however, all the methods, without additional filtering, exhibit a tendency to potentially overestimate the magnitude of the largest DW values (consistent with the results discussed with respect to the apparent outliers in Figure 4), particularly for Himawari-8 and GOES-16. The relative differences in the estimated peak

magnitudes and associated frequencies between the methods show variations among sensors and months. The difference among methods is generally reduced for Himawari-8. For SEVIRI, the QC5 approach suggests a reduced occurrence of larger events in part because of fewer quality level 5 values in the screening scheme, consistent with the missed event shown in Figure 3 and the discussion in Section 4.1.

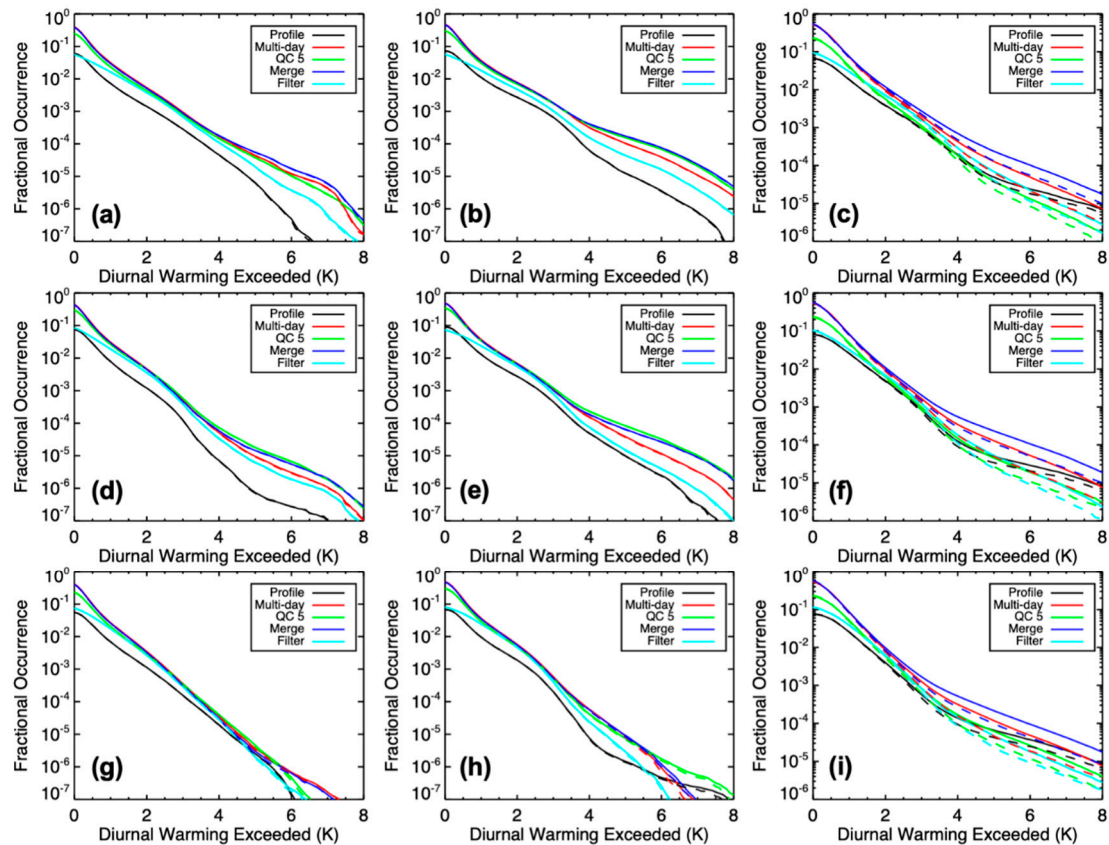


Figure 8. Comparison of distributions of the cumulative diurnal warming exceedance computed for the (left: **a,d,g**) Himawari-8 AHL, (middle: **b,e,h**) GOES-16 ABI, and (right: **c,f,i**) Meteosat-11 SEVIRI during the months of (top: **a–c**) June, (middle: **d–f**) July, and (bottom: **g–i**) August 2019 using the different methods evaluated (see legend for trace color). Solid traces represent values obtained at all grid points at full resolution, while the dotted traces are obtained based on the median of a 3×3 array of grid values centered at each location.

As noted previously, overestimation of the most extreme DW values could result from noise in individual SST retrievals and impacts of any residual cloud contamination on SST_{found} . Maximum value composite procedures, as employed in constructing the distributions of DW_{amp} , are particularly sensitive to observational noise. Various steps were explored to reduce the potential impact of noisy retrievals on the estimation of the largest amplitudes. Alternate DW_{amp} values were computed replacing the individual, full grid resolution SST_{inst} values with the median of the 3×3 array of grid values centered on each location. This approach, consistent with that employed by Merchant et al. [4], reduces the impact of a single noisy retrieval but also reduces the effective spatial resolution of the DW events sampled. These results are reflected by the dashed lines in Figure 8. The median filter has very little effect on the higher resolution products from Himawari-8 and GOES-16 but does impact the tails of the DW_{amp} distributions from SEVIRI, reducing the occurrence frequency of estimates of $DW_{amp} > \sim 4$ K. Not surprisingly, the reduction is smallest for the reference profile methodology, which was designed to be more rigorous and less sensitive to noise.

A more consistent reduction in DW_{amp} values across all sensors was obtained by filtering out any derived amplitudes obtained corresponding to wind speeds greater than 3 m/s (termed the “filter” methodology). Warming in excess of 6 K at higher wind speeds is not expected, and the fact that DW_{amp} estimates of that magnitude were obtained under those conditions strongly suggests the presence and influence of noise. For Himawari-8 and GOES-16, DW estimates filtered by wind speed provide the best overall match to the distribution of large-amplitude events from the profile methodology. Filtering by wind also appears to consistently provide a conservative estimate of the frequency of the largest warming events, and the agreement with the profile methodology provides initial support to the assertion that, with appropriate care, the operational SST retrievals can provide viable estimates of extreme DW . The “optimal” choice of a wind speed threshold or other filtering criteria is open to further investigation, but the use of filtering does appear to be a valuable option.

4.3. Comparison of Overlapping GOES-16 and Meteosat-11 Results

To further assess the veracity of the extreme DW_{amp} estimates in the absence of validation against in situ observations, we can also compare independent estimates from different sensors in regions of overlapping coverage. The GOES-16 ABI and Meteosat-11 SEVIRI retrievals overlap in the Atlantic Ocean as observed in Figure 1, albeit at elevated sensor zenith angles. Unfortunately, no similar overlap with Himawari-8 is currently possible given the status of retrieved products from GOES-17.

Overlapping daily DW_{amp} estimates from GOES-16 and Meteosat-11 derived with the different methodologies were compared over the full three months of the study. The finer-resolution GOES-16 values were spatially averaged onto the grid cells corresponding to the Meteosat-11 data. The results are presented as density scatter plots in Figure 9, and corresponding statistics are summarized in Table 1. At first glance, the scatter plots suggest favorable agreement. Common overlapping DW events up to ~4 K magnitude are observed. Large DW_{amp} estimates from one sensor generally correspond to similar large estimates from the other, particularly for the reference profile methodology. This lends further confidence in the ability to obtain estimates of large DW from the products. The least scatter in the differences is obtained with the reference profile methodology.

Table 1. Meteosat-11–GOES-16 DW_{amp} differences for overlapping coverage disks during June–August 2019. The “All” category includes all the DW_{amp} differences where collocated data exist, while the other categories include differences only where the DW_{amp} exceeds the indicated threshold for either sensor.

Method	All Values		$DW_{amp} > 0.5$ K		$DW_{amp} > 3$ K	
	Mean Difference	Standard Deviation	Mean Difference	Standard Deviation	Mean Difference	Standard Deviation
Profile	0.04 K	0.33 K	−0.02 K	0.33 K	−0.38 K	0.39 K
Multiday	0.13 K	0.37 K	0.06 K	0.38 K	−0.43 K	0.67 K
QC5	0.14 K	0.35 K	0.06 K	0.36 K	−0.37 K	0.53 K
Merge	0.08 K	0.39 K	0.02 K	0.41 K	−0.56 K	0.69 K
Filter	−0.04 K	0.46 K	−0.08 K	0.43 K	−0.40 K	0.68 K

A closer examination, however, reflects some important differences in the absolute derived magnitudes. The scatter plots suggest a tendency for larger DW_{amp} values from GOES-16, and the statistics firmly reflect this. When the results are constrained to cases where a DW_{amp} value from either sensor is estimated to be greater than 3 K, the GOES-16 amplitudes are significantly larger on average, as reflected by the negative biases in Table 1. These results appear consistent with the increased sensitivity of the GOES-16 (and Himawari-8) SST retrieval algorithms noted by Petrenko et al. [25]. The sensitivity refers to the ability of an algorithm to reproduce the true variations of the SST in the retrieved SST in the presence of atmospheric contaminants [31]. The present GOES-16 and Himawari-8 algorithms were found [25] to more accurately detect small variations in the true SST and predict larger diurnal variability than other retrieval approaches. These results suggest that peak DW_{amp} values

within the Meteosat-11 domain may actually be slightly larger than those retrieved operationally from SEVIRI.

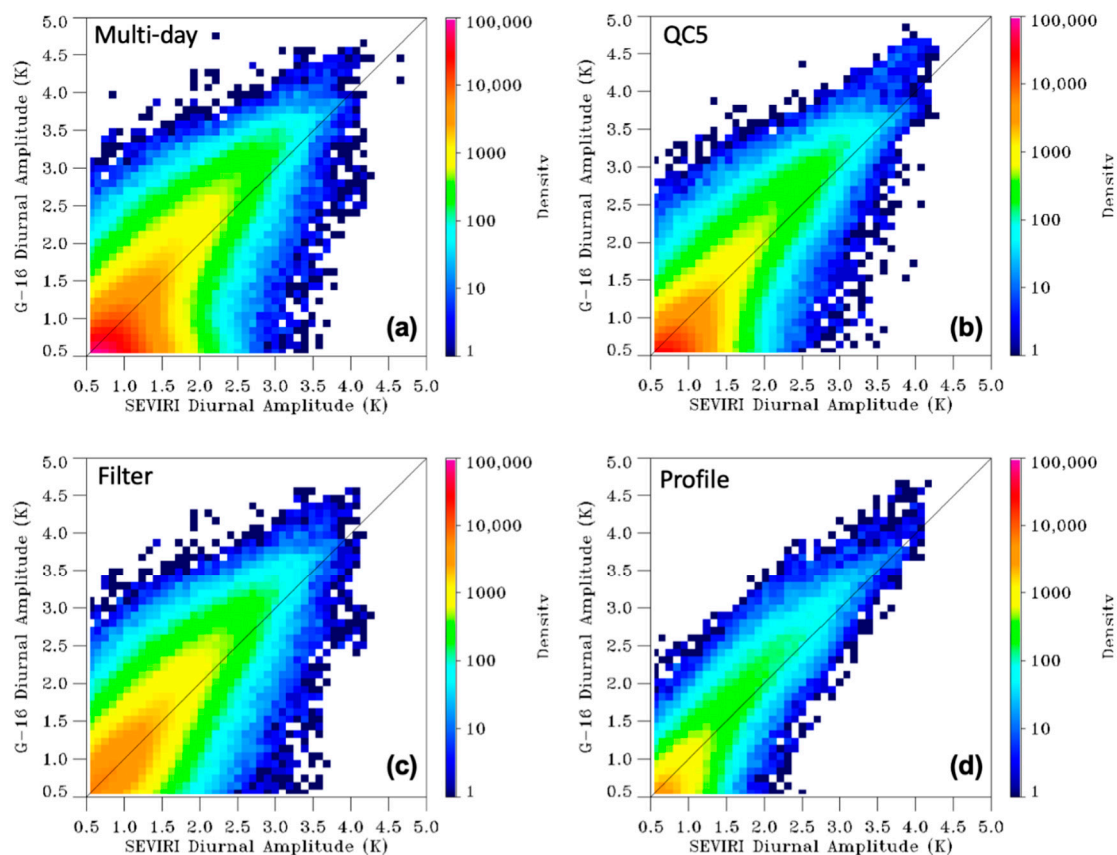


Figure 9. Comparison of coincident daily DW_{amp} estimates from GOES-16 and Meteosat-11 derived using the (a) multiday, (b) QC5, (c) filter, and (d) profile methods compiled over all three months of the study. The point color represents the number of observations as indicated by the color bars to the right of the panels. Note the change in range between the top and bottom panels.

Interestingly, however, this tendency for greater extreme DW_{amp} values from GOES-16 is not uniformly reflected across the full distribution of DW_{amp} retrievals. Additional results derived from the collocated overlapping observations, presented in Figures 10 and 11, demonstrate that while GOES-16 yields larger and more frequent extreme diurnal warming, for smaller DW_{amp} events, the warming is often greater for Meteosat-11. Maps of the monthly peak DW_{amp} values for June 2019 (Figure 10) show larger DW_{amp} values from GOES-16 for peaks in excess of ~ 1.5 K but larger peak warming from Meteosat-11 below that threshold. Exceedance distributions of collocated DW_{amp} values compiled over the full three-month period (Figure 11) demonstrate greater exceedance frequency for Meteosat-11 up to a transition at 1.5 K. Since this paper is focused on extreme DW, these differences for smaller amplitudes will be explored further in a separate analysis.

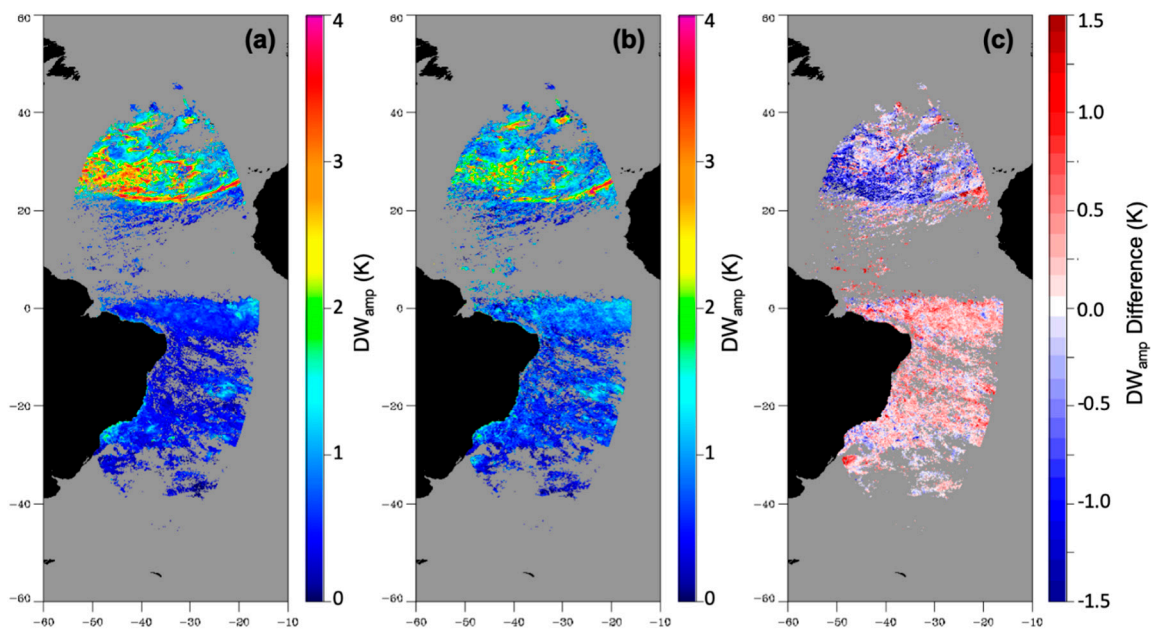


Figure 10. Comparison of monthly peak daily DW_{amp} estimates for coincident measurements from (a) GOES-16 and (b) Meteosat-11 in June 2019. Results were obtained with the profile methodology. The corresponding difference (Meteosat-11–GOES-16) is shown in (c).

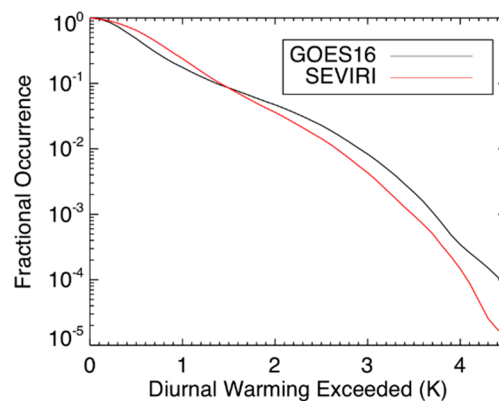


Figure 11. Comparison of distributions of the cumulative diurnal warming exceedance computed for the GOES-16 ABI and Meteosat-11 SEVIRI during the months of June–August 2019 using the profile methodology (see legend for trace color).

4.4. Characterization of Extreme DW Values

Since obtaining viable estimates of the largest DW amplitudes appears feasible with the operational SST products using appropriate care, it is possible to explore what constitutes “extreme” warming and how it varies with geographical region. A representation of the geographical distribution of extreme DW is shown in Figure 12, which illustrates the derived monthly peak DW_{amp} for each sensor. The results obtained using the filter method are shown here to provide broader spatial coverage while not overestimating the peak values. The patterns generally appear reasonable, with the largest warming occurring in the midlatitudes of the summer hemisphere. Very large warming is also observed in the Mediterranean and in other near-coastal and enclosed regions. Coherent areas of large warming in high latitudes of the winter hemisphere, noted in Figure 2 for some methods, are now largely absent. Areas with no valid DW_{amp} estimates are the result of persistent cloudiness and/or the lack of wind speeds less than 3 m/s.

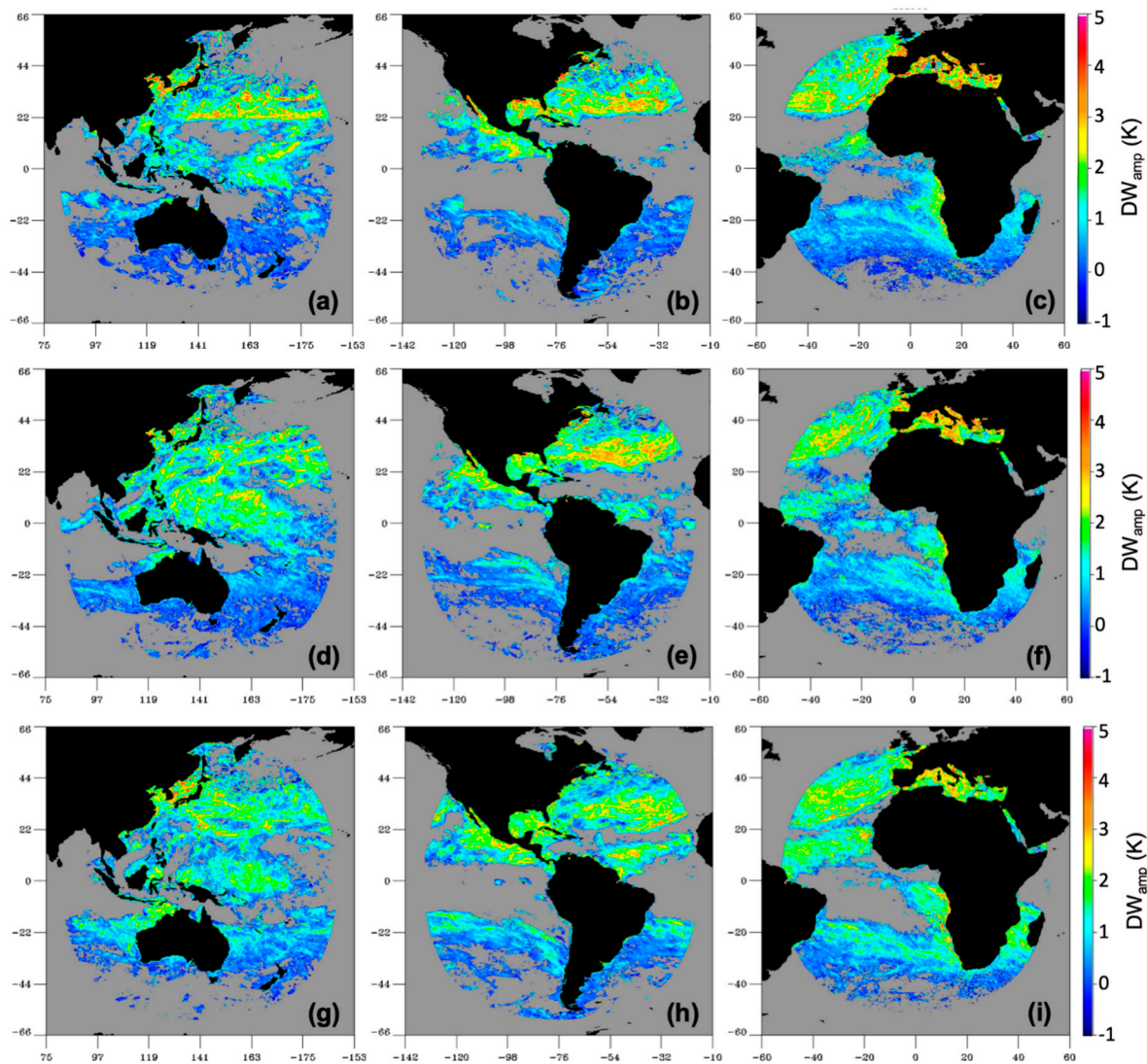


Figure 12. Geographical distribution of the monthly peak DW_{amp} computed using the filter method for the (left: **a,d,g**) Himawari-8 AHL, (middle: **b,e,h**) GOES-16 ABI, and (right: **c,f,i**) Meteosat-11 SEVIRI during the months of (top: **a–c**) June, (middle: **d–f**) July, and (bottom: **g–i**) August 2019. The color bar is the same for each panel and is displayed at the right of each row. Gray colors indicate the lack of any daily DW_{amp} estimate at that location during the month.

As a first attempt at quantifying the frequency of extreme warming, we computed the fraction of days for which coherent warming in excess of 4 and 6 K is observed at least somewhere within the full disk of the individual sensors. Averaging across the sensors and all three months, based on the profile methodology, 47% of the days exhibited warming of at least 6 K. The corresponding number is 91% for a threshold of 4 K. To minimize the potential impact of any noise and ensure some measure of coherence in the warming, days were included if the median DW_{amp} value for a 2×2 subarray of the retrieved values exceeded the selected threshold somewhere within the scene. Additional methods for filtering out anomalous individual estimates such as requiring the warming to be present in at least 5 individual grid points were also tested with generally similar results.

To further explore potential thresholds for extreme warming and the associated likelihood, using the cumulative distributions of DW_{amp} constructed in Figure 8, we extracted the probabilities for exceeding specific values and computed the 95th and 99th percentiles of retrieved DW_{amp} values. Combining the results for the full June–August period, the probabilities of DW_{amp} exceeding 4 and 6 K are summarized in Table 2 for the different methodologies. Results are shown both for the full grid

resolutions and after applying the 3×3 median filter on the spatial grids as discussed in reference to the results of Figure 8.

Table 2. Probability of DW exceedance for thresholds of 4 and 6 K for the full June–August period stratified by sensor and method.

Resolution	Method	Himawari-8		GOES-16		Meteosat-11	
		>4 K	>6 K	>4 K	>6 K	>4 K	>6 K
Full Resolution	Profile	2.0×10^{-5}	2.2×10^{-7}	3.3×10^{-5}	1.9×10^{-6}	1.3×10^{-4}	2.9×10^{-5}
	Multiday	6.2×10^{-5}	4.7×10^{-6}	1.4×10^{-4}	1.5×10^{-5}	3.1×10^{-4}	4.6×10^{-5}
	QC5	7.0×10^{-5}	5.1×10^{-6}	1.9×10^{-4}	3.0×10^{-5}	1.4×10^{-4}	1.8×10^{-5}
	Merge	6.6×10^{-5}	7.0×10^{-6}	1.8×10^{-4}	3.1×10^{-5}	5.2×10^{-4}	9.3×10^{-5}
	Filter	4.7×10^{-5}	1.7×10^{-6}	7.0×10^{-5}	5.9×10^{-6}	1.7×10^{-4}	1.8×10^{-5}
Median Filter	Profile	1.9×10^{-5}	2.0×10^{-7}	3.2×10^{-5}	1.8×10^{-6}	9.9×10^{-5}	2.0×10^{-5}
	Multiday	7.1×10^{-5}	4.6×10^{-6}	1.5×10^{-4}	1.5×10^{-5}	1.8×10^{-4}	1.8×10^{-5}
	QC5	7.2×10^{-5}	5.0×10^{-6}	1.9×10^{-4}	3.0×10^{-5}	9.2×10^{-5}	9.5×10^{-6}
	Merge	7.5×10^{-5}	6.8×10^{-6}	1.9×10^{-4}	3.1×10^{-5}	3.1×10^{-4}	4.6×10^{-5}
	Filter	4.7×10^{-5}	1.7×10^{-6}	7.1×10^{-5}	5.9×10^{-6}	1.1×10^{-4}	9.1×10^{-6}

While DW up to 6 K is frequently observed within the satellite scenes, the overall likelihood of the DW at any grid cell exceeding 4 or 6 K is small given the expanse of the oceans. Based on the reference profile methodology, the probability of DW exceeding 4 K ranges from 1.3×10^{-4} for the Meteosat-11 SEVIRI to 2.0×10^{-5} for the Himawari-8 AHI. The corresponding values for exceeding 6 K are 2.9×10^{-5} and 2.2×10^{-7} , respectively. Within the individual months, the values were largely similar though the probabilities decreased slightly in July and August for Himawari-8 and GOES-16. The differences between sensors reflect both the regions sampled and differences in product resolution. For the finer-resolution GOES-16 and Himawari-8 estimates, higher peak values are possible (e.g., [3]), but the total number of grid cells is much greater and the results show that a smaller fraction of overall grid cells was affected by the largest warming.

The likelihood of large diurnal warming exhibits a strong dependence on the region sampled. Exceedance statistics were also calculated independently for the Northern and Southern Hemispheres (NH and SH, respectively), using the profile methodology for each sensor, and for the Mediterranean Sea (MED) in the case of the Meteosat-11 SEVIRI. The results are shown in Table 3. The probability of large DW is much greater in the Mediterranean Sea region, and the computed values agree in order of magnitude with those found by Merchant et al. [4]. In all other regions, the likelihood of large warming is significantly reduced. Not surprisingly, probabilities are larger in the summer hemisphere. Northern Hemisphere statistics for warming in excess of 4 K agree within the order of magnitude across all sensors. The smallest hemispheric differences are observed for the Meteosat-11 SEVIRI. Overall, the lowest probabilities for extreme warming are observed for the Himawari-8 domain.

Table 3. Probability of DW exceedance for the June–August period stratified by sensor and geographical region based on the profile methodology.

Threshold	Himawari-8		GOES-16		Meteosat-11		
	NH	SH	NH	SH	Med	NH	SH
>4 K	4.3×10^{-5}	7.7×10^{-7}	6.2×10^{-5}	3.4×10^{-6}	1.7×10^{-3}	3.6×10^{-5}	2.1×10^{-5}
>6 K	4.3×10^{-7}	3.3×10^{-8}	3.3×10^{-6}	4.0×10^{-7}	3.3×10^{-4}	7.0×10^{-6}	4.9×10^{-6}

Comparing the results from the different methodologies (Table 2), although the overall probabilities of DW in excess of 4 or 6 K are small, the choice of the method results in differences that vary by more than a factor of two. Clearly, the method employed has an impact on the behavior of the tail of the DW_{amp} distribution, as also seen in Figure 8. Overall, the lowest probabilities and closest

agreement with the reference profile method are obtained with the filter methodology. The other methods (with the exception of QC5 for Meteosat-11) generally exhibit elevated probabilities consistent with overestimation of the warming as discussed with respect to Figure 8. The utility of the filter methodology to provide a conservative estimate of extreme warming was already noted previously in Section 4.2.

The magnitude of extreme warming can also be characterized in terms of percentiles of the estimated DW_{amp} values. The corresponding 95th and 99th percentiles of retrieved DW_{amp} values are shown in Table 4. Two different approaches were utilized to focus on instances where we were confident that significant DW occurred. While DW is typically small on average (e.g., $DW_{amp} < \sim 0.5$ K), there is uncertainty surrounding these smaller amplitudes due to measurement noise and the efforts to be conservative and avoid overestimating any warming. Smaller amplitudes are less than the expected error bars, and it is difficult to state with confidence whether any warming occurred. Additionally, negative DW_{amp} values were occasionally computed as a result of the conservative methodologies employed. The approaches considered attempted to compute percentiles for the conditional case of DW when significant warming did occur. In the first method (consistent with [22]), we simply considered the percentiles of all estimates where DW_{amp} was greater than 0.5 K. In the second, we assumed that the portion of the probability distribution with negative DW_{amp} values represented random variability that could occur in the absence of significant warming. We then subtracted the counts from this negative portion of the distribution from the corresponding bins for positive amplitudes and computed the percentiles for the remaining values.

Table 4. The 95th and 99th percentiles of the DW amplitude stratified by sensor and method.

Month	Method	Himawari-8		GOES-16		Meteosat-11	
		95%	99%	95%	99%	95%	99%
DW > 0.5	Profile	2.0 K	2.8 K	2.2 K	3.0 K	2.3 K	3.2 K
	Multiday	1.9 K	2.7 K	2.0 K	2.9 K	1.8 K	2.7 K
	QC5	1.9 K	2.8 K	2.0 K	3.0 K	1.9 K	2.8 K
	Merge	1.9 K	2.7 K	2.0 K	2.9 K	1.9 K	2.8 K
	Filter	2.3 K	3.0 K	2.4 K	3.2 K	2.4 K	3.3 K
Folded	Profile	1.4 K	2.4 K	1.6 K	2.7 K	2.1 K	3.0 K
	Multiday	1.3 K	2.2 K	1.3 K	2.3 K	1.5 K	2.4 K
	QC5	1.3 K	2.3 K	1.4 K	2.4 K	1.7 K	2.6 K
	Merge	1.3 K	2.2 K	1.3 K	2.3 K	1.6 K	2.5 K
	Filter	2.0 K	2.8 K	2.3 K	3.0 K	2.2 K	3.1 K

The results suggest again that the method employed does impact quantitative conclusions as to what constitutes extreme warming. Considering only events >0.5 K, the profile methodology consistently reflects a value of ~ 2 K for the 95th percentile and ~ 3 K for the 99th percentile of DW. Not surprisingly, the “folded” distribution approach, which attempts to also consider smaller amplitude events, results in smaller values. For the Meteosat-11 SEVIRI, the filter methodology consistently best matches the profile results. The differences between methodologies are smaller for Himawari-8 and GOES-16. Broadly considering all the results, a DW_{amp} value of 3 K appears a good approximation of the 99th percentile of DW values and potentially could serve as a practical threshold for defining “extreme” warming.

5. Discussion

Based on the results, one can return to the questions posed in the introduction. As to the first question, i.e., that considering whether operational SST products from current geostationary sensors can accurately characterize the most extreme DW events, the answer appears to be affirmative, but significant care must be taken in processing data and in selecting the methodology employed.

All the sensors and methods evaluated consistently demonstrate very large warming events up to 6 K magnitude and greater, and where the sensor coverage overlaps, the different products generally agree on the occurrence of extreme warming. While the performance of the sensors and the retrieval algorithms is good, the estimation of extreme values is highly sensitive to noise and the methods employed. Outliers show DW values in excess of what is expected for the given conditions, and uncertainty exists as to the frequency and magnitude of the greatest values. One should not rely solely on retrieved SST values passing all traditional quality screening to infer the largest DW_{amp} values. Some level of filtering or careful processing that broadly takes the consistency and context of the data into account should be employed.

Filtering methods exhibiting some success included the profile methodology, which interpreted warming in the context of the full observational time series, and screening based on the existence of coincident low wind speeds required for significant warming. Other filtering methods could potentially look more closely at spatial coherence and consistency in the retrieved SST and derived foundation temperature values. Lower wind speed thresholds could also be explored when isolating the absolute highest DW values.

The “best” method for estimating large-amplitude DW events is likely dependent on the specific application. The approach must balance data quality and availability. For validation of DW models or other applications where the greatest precision is required and it is acceptable to have estimates at only a limited number of locations, the use of the profile methodology gives the best accuracy. When increased sampling is desirable, the QC5 method yields fairly similar statistics to the profile reference (at least for Meteosat-11) over a greater number of points with a minimum of additional filtering. To more accurately estimate the full distribution of DW_{amp} values, the multiday approach enables more complete sampling, and additional wind filtering helps ensure the extreme values are not overestimated. Of the foundation-based approaches, the wind filtering generally provides the representation of extreme warming most consistent with the profile methodology. If the SST_{found} is to be used as a product of its own, the multiday approach yields the best temporal and spatial consistency. The multiday foundation approach with filtering can also be extended to estimation of DW_{amp} from polar satellites where observations are generally limited to twice per day and are inadequate for application of the other methods.

Addressing the second question, i.e., that of determining the amplitude and frequency of the largest DW events, we see that there clearly is some uncertainty, and the values depend, to some extent, on the methods used. Based on the reference profile methodology and a comparison of the other methods, however, it is possible to provide some useful numerical estimates. Probability values were derived for warming in excess of specific thresholds, and percentiles were found to be another valid way of characterizing what constitutes extreme warming. A value of 3 K was found to provide a good estimate of the 99th percentile of diurnal warming and might serve as a practical threshold for extreme warming. Comparison of the different methods suggested confidence in DW values up to at least 6 K. Warming in excess of 6 K was observed from each sensor on approximately 47% of the days examined, but the overall probability of such values was of $O(10^{-6})$ when considering all regions. The frequency of such large warming is clearly dependent upon region and season, and a probability of 3×10^{-4} was found for the Mediterranean in the summer. Additional work underway will better characterize such regional and seasonal variations.

While the uncertainty in individual DW_{amp} estimates is not inconsistent with documented geostationary SST retrieval uncertainties (order 0.4 K according to [4,25]), the uncertainty in peak warming estimates can be larger than what might be casually assumed based on those retrieval uncertainties. The difference of two quantities with random errors of 0.4 K implies an uncertainty in individual DW_{amp} values of 0.6 K, but yet notably larger differences are observed. Uncertainties are typically expressed, however, in terms of one standard deviation, and larger differences are expected statistically. Peak DW values are based on the extremes of both quantities being differenced and, thus, emphasize the extremes. Estimation of the foundation temperature adds more uncertainty. The results

(e.g., Figure 4) using normally screened SST retrievals clearly showed derived DW_{amp} values more than 1 K different from what might be expected under the corresponding conditions. Composited peak values (over time and/or space) are also dependent on the sampling, and some of the largest DW events were not captured by retrievals of the highest assumed quality levels. Comparisons across the different methods suggested differences in perceived peak amplitudes on the order of 1 K if careful processing and/or filtering of the results is not performed. The potential for uncertainty of this magnitude should be recognized and addressed.

The sensitivity of the SST retrieval algorithm remains an important concern. Differences between the largest DW_{amp} values from overlapping GOES-16 ABI and Meteosat-11 SEVIRI were consistent with [25], where it was noted that retrieval algorithms with reduced sensitivity can underestimate the DW. The largest DW_{amp} values were primarily found to be located in the midlatitudes. It is important to further evaluate whether there is still potentially some underestimation of peak warming at lower latitudes where increased atmospheric attenuation can degrade the sensitivity.

6. Conclusions

Various methods were employed to explore and characterize the occurrence of extreme diurnal warming as observed using operational SST retrievals from multiple current geosynchronous satellites for a three-month period in 2019. The methods differed primarily in their approach to estimating a foundation temperature from which the diurnal warming could be derived. Close examination of the DW estimates demonstrates several challenges in accurately deriving distributions of the DW amplitude. The results consistently suggest the occurrence of large DW events in excess of 6 K but demonstrate uncertainty as to the frequency of these events as well as the absolute peak values depending on the method utilized. While there seems to be little doubt that such large DW events occur and are detectable, the results are sensitive to the method employed, and some level of screening/filtering is required in addition to the operational SST processing.

With appropriate care, current geostationary satellite sensors do appear capable of providing viable estimates of extreme diurnal warming. The results demonstrate this is true for the GOES-16 ABI and Himawari-8 AHI as had been established previously for SEVIRI. Coincident overlapping DW_{amp} estimates from GOES-16 and Meteosat-11 confirmed generally consistent observations of large warming, but amplitudes were somewhat greater for GOES-16 potentially due to increased sensitivity of the SST retrieval algorithm. Application of a reference methodology in which the DW amplitude is derived from continuous and physically believable diurnal time series of the SST retrievals further supported the existence and detectability of extreme DW and provided a means to evaluate simplified DW_{amp} estimates based on a derived foundation temperature. For aspects like validation of DW models and when adequate data exist, the use of this reference profile methodology appears most reliable. Other methods exhibit some potential to overestimate the likelihood of extreme DW but, with appropriate filtering, can be applied to expand coverage and provide a better representation of the complete distribution of actual DW events.

Overall, DW in excess of 6 K was found to occur at least somewhere within the geosynchronous coverage disks on nearly 47% of the days analyzed. While occurrence of such extremes is quite common, the likelihood at any given location is much smaller. Based on the reference profile methodology, the probability of DW in excess of 4 and 6 K at a specific grid point for the regions and period analyzed was found to be on the order of 10^{-5} and 10^{-6} , respectively. Additional percentiles of estimated DW_{amp} values were computed, and based on the 99th percentile, 3 K may potentially be considered a useful threshold for defining extreme warming.

While we have suggested that some level of screening/filtering is required, we do not claim to have identified an “optimum” solution for estimation of extreme DW. Future work could explore additional filtering criteria and methods for estimation of the foundation temperature to determine if further improvements in the representation of extreme DW are possible. Any “best” method would

potentially exhibit dependence on the specific sensor being used. Additional new and latest generation geostationary-deployed sensors also warrant similar investigation.

The results of this study are now being applied to enhanced validation of predictions of large DW by physically based models and a more detailed investigation of the climatology of large DW events. While geostationary sensors enable the most direct estimation of DW_{amp} , global climatologies derived over longer periods also require supplemental information from polar satellites. The detailed evaluation of the foundation-based methods with the more temporally complete geostationary products also enables more confident extension of the methods to the less frequent polar-orbiting data and provides an indication of the expected uncertainties.

Author Contributions: G.A.W. conceived the study, participated in data processing, analyzed final results, and led drafting of the paper; S.L.C. implemented the initial foundation temperature and DW processing for geostationary satellites, contributed to the data processing, and participated in the drafting of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Early components of the project were supported by NASA Grant NNH16AC10I from the Physical Oceanography Program. S.L.C. was supported in part through the National Ocean Partnership Project Multi Sensor Improved Sea Surface Temperature: continuing the GHRSSST partnership and improving Arctic data Contract 80NSSC18K0837. Additional efforts for G.A.W. were supported by NOAA funding to the Physical Sciences Laboratory.

Acknowledgments: The authors gratefully acknowledge the provision of satellite data and discussions regarding its content by Stephane Saux-Picart from Meteo-France CMS and Alexander Ignatov of NOAA NESDIS. The constructive reviews of four anonymous reviewers were helpful and very appreciated.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Stuart-Menteth, A.C.; Robinson, I.S.; Challenor, P.G. A global study of diurnal warming using satellite-derived sea surface temperature. *J. Geophys. Res.* **2003**, *108*, 3155. [[CrossRef](#)]
2. Kennedy, J.J.; Brohan, P.; Tett, S.F.B. A global climatology of the diurnal variations in sea surface temperature and implications for MSU temperature trends. *Geophys. Res. Lett.* **2007**, *34*, L05712. [[CrossRef](#)]
3. Gentemann, C.L.; Minnett, P.J.; Le Borgne, P.; Merchant, C.J. Multi-satellite measurements of large diurnal warming events. *Geophys. Res. Lett.* **2008**, *35*, L22602. [[CrossRef](#)]
4. Merchant, C.J.; Filipiak, M.J.; Le Borgne, P.; Roquet, H.; Autret, E.; Piolle, J.-F.; Lavender, S. Diurnal warm-layer events in the western Mediterranean and European shelf seas. *Geophys. Res. Lett.* **2008**, *35*, L04601. [[CrossRef](#)]
5. Flament, P.; Firing, J.; Sawyer, M.; Trefois, C. Amplitude and horizontal structure of a large sea surface warming event during the Coastal Ocean Dynamics Experiment. *J. Phys. Oceanogr.* **1994**, *24*, 124–139.
6. Donlon, C.; Robinson, I.; Casey, K.S.; Vazquez-Cuervo, J.; Armstrong, E.; Arino, O.; Gentemann, C.; May, D.; Le Borgne, P.; Piollé, J. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bull. Amer. Meteor. Soc.* **2007**, *88*, 1197–1213. [[CrossRef](#)]
7. Yang, G.; Slingo, J. The diurnal cycle in the tropics. *Mon. Weather Rev.* **2001**, *129*, 784–801. [[CrossRef](#)]
8. Price, J.F.; Weller, R.A.; Pinkel, R. Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing. *J. Geophys. Res.* **1986**, *91*, 8411–8427. [[CrossRef](#)]
9. Fairall, C.W.; Bradley, E.F.; Godfrey, J.S.; Wick, G.A.; Edson, J.B.; Young, G.S. Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.* **1996**, *101*, 1295–1308. [[CrossRef](#)]
10. Danabasoglu, G.; Large, W.G.; Tribbia, J.J.; Gent, P.R.; Briegleb, B.P. Diurnal coupling in the tropical oceans of CCSM3. *J. Clim.* **2006**, *19*, 2347–2365. [[CrossRef](#)]
11. Weihs, R.R.; Bourassa, M.A. Modeled diurnally varying sea surface temperatures and their influence on surface heat fluxes. *J. Geophys. Res.* **2014**, *119*, 4101–4123. [[CrossRef](#)]
12. Marullo, S.; Minnett, P.J.; Santoleri, R.; Tonani, M. The diurnal cycle of sea-surface temperature and estimation of the heat budget of the Mediterranean Sea. *J. Geophys. Res.* **2016**, *121*, 8351–8367. [[CrossRef](#)]
13. Clayson, C.A.; Bogdanoff, A.S. The effect of diurnal sea surface temperature warming on climatological air-sea fluxes. *J. Clim.* **2013**, *26*, 2546–2556. [[CrossRef](#)]

14. Shinoda, T.; Hendon, H.H. Mixed layer modeling of intraseasonal variability in the tropical western Pacific and Indian Oceans. *J. Clim.* **1998**, *11*, 2668–2685. [[CrossRef](#)]
15. McCreary, J.P., Jr.; Kohler, K.E.; Hood, R.R.; Smith, S.; Kindle, J.; Fischer, A.S.; Weller, R.A. Influences of diurnal and intraseasonal forcing on mixed-layer and biological variability in the central Arabian Sea. *J. Geophys. Res.* **2001**, *106*, 7139–7155. [[CrossRef](#)]
16. Shinoda, T. Impact of the diurnal cycle of solar radiation on intraseasonal SST variability in the western equatorial Pacific. *J. Clim.* **2005**, *18*, 2628–2636. [[CrossRef](#)]
17. Bernie, D.J.; Woolnough, S.J.; Slingo, J.M.; Guilyardi, E. Modeling diurnal and intraseasonal variability of the ocean mixed layer. *J. Clim.* **2005**, *18*, 1190–1202. [[CrossRef](#)]
18. Woolnough, S.J.; Vitart, F.; Balmaseda, M.A. The role of the ocean in the Madden-Julian Oscillation: Implications for MJO prediction. *Quart. J. Roy. Meteor. Soc.* **2007**, *133*, 117–128. [[CrossRef](#)]
19. Seo, H.; Subramanian, A.C.; Miller, A.J.; Cavanaugh, N.R. Coupled impacts of the diurnal cycle of sea surface temperature on the Madden-Julian oscillation. *J. Clim.* **2014**, *27*, 8422–8443. [[CrossRef](#)]
20. Ham, Y.-G.; Kug, J.-S.; Kang, I.-S.; Jin, F.-F.; Timmermann, A. Impact of diurnal atmosphere-ocean coupling on tropical climate simulations using a coupled GCM. *Clim. Dyn.* **2010**, *34*, 905–917. [[CrossRef](#)]
21. Tian, F.; von Storch, J.-S.; Hertwig, E. Impact of SST diurnal cycle on ENSO asymmetry. *Clim. Dyn.* **2019**, *52*, 2399–2411. [[CrossRef](#)]
22. Le Borgne, P.; Legendre, G.; Péré, S. Comparison of MSG/SEVIRI and drifting buoy derived diurnal warming estimates. *Remote Sens. Environ.* **2012**, *124*, 622–626. [[CrossRef](#)]
23. Karagali, I.; Høyer, J. Characterization and quantification of regional diurnal SST cycles from SEVIRI. *Ocean Sci.* **2014**, *10*, 745–758. [[CrossRef](#)]
24. Castro, S.; Wick, G.; Buck, J.J.H. Comparison of diurnal warming estimates from unpumped ARGO data and SEVIRI satellite observations. *Remote Sens. Environ.* **2014**, *140*, 789–799. [[CrossRef](#)]
25. Petrenko, B.; Ignatov, A.; Kihai, Y.; Pennybacker, M. Optimization of sensitivity of GOES-16 ABI sea surface temperature by matching satellite observations with L4 analysis. *Remote Sens.* **2019**, *11*, 206. [[CrossRef](#)]
26. Maturi, E.; Harris, A.; Mittaz, J.; Sapper, J.; Wick, G.; Zhu, X.; Dash, P.; Koner, P. A new high resolution sea surface temperature blended analysis. *Bull. Amer. Meteor. Soc.* **2017**, *98*, 1015–1026. [[CrossRef](#)]
27. EUMETSAT. Geostationary Sea Surface Temperature Product User Manual. 2011. SAF/OSI/CDOP/M-F/TEC/MA/181. Available online: <http://www.osi-saf.org> (accessed on 3 June 2019).
28. Walton, C.C.; Pichel, W.G.; Sapper, J.F.; May, D.A. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *J. Geophys. Res.* **1998**, *103*, 27999–28012. [[CrossRef](#)]
29. Le Borgne, P.; Roquet, H.; Merchant, C.J. Estimation of sea surface temperature from the Spinning Enhanced Visible and Infrared Imager, improved using numerical weather prediction. *Remote Sens. Environ.* **2011**, *115*, 55–65. [[CrossRef](#)]
30. Zhang, H.; Beggs, H.; Majewski, L.; Wang, X.H.; Kiss, A. Investigating sea surface temperature diurnal variation over the Tropical Warm Pool using MTSAT-1R data. *Remote Sens. Environ.* **2016**, *183*, 1–12. [[CrossRef](#)]
31. Merchant, C.J.; Harris, A.R.; Roquet, H.; Le Borgne, P. Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. *Geophys. Res. Lett.* **2009**, *36*, L17604. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).