

A Dataset of Hourly Sea Surface Temperature From Drifting Buoys

Shane Elipot^{1,*}, Adam Sykulski², Rick Lumpkin³, Luca Centurioni⁴, and Mayra Pazos³

¹Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA

²Lancaster University, Department of Mathematics and Statistics, Lancaster, LA1 4YW, UK

³NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, FL 33149, USA

⁴Lagrangian Drifter Laboratory, Scripps Institution of Oceanography, University of California San Diego, San Diego, CA 92103, USA

*corresponding author(s): Shane Elipot (selipot@miami.edu)

ABSTRACT

A dataset of sea surface temperature (SST) estimates is generated from the temperature observations of surface drifting buoys of NOAA's Global Drifter Program. Estimates of SST at regular hourly time steps along drifter trajectories are obtained by fitting to observations a mathematical model representing simultaneously SST diurnal variability with three harmonics of the daily frequency, and SST low-frequency variability with a first degree polynomial. Subsequent estimates of non-diurnal SST, diurnal SST anomalies, and total SST as their sum, are provided with their respective standard uncertainties. This Lagrangian SST dataset has been developed to match the existing hourly dataset of position and velocity from the Global Drifter Program.

Background & Summary

The Global Drifter Program (GDP) funded by the U.S. National Oceanic and Atmospheric Administration (NOAA) maintains an array of satellite-tracked water-following drifting buoys, hereafter referred to as *drifters*, designed to acquire in situ observations of near-surface ocean current, sea surface temperature (SST), and atmospheric sea level pressure¹. The requirement of the Global Ocean Observing System (GOOS) to achieve a nominal $5^\circ \times 5^\circ$ coverage of the world's ocean has been fulfilled since September 2005 with a pool of 1250 drifters². In near-real time, drifter locations and sensor data are relayed to the WMO's Global Telecommunication System (GTS), contributing to the collection of critical information needed for the World Weather Watch programme. Drifter data are also harvested by various national and international projects and organizations which aim at assembling in situ SST observations to produce quality-controlled and reformatted datasets for scientific analyzes, climate monitoring, and calibration and validation of satellite-based SST observations. In delayed-time, the GDP maintains the historical database of drifter data and metadata, and delivers regular updates of drifter data products of surface currents and SST following quality control and estimation procedures. The historical observations, with the earliest ones from 1979, have been processed in incremental steps to generate a 6-hour joint dataset of drifter position, velocity, and SST estimates, along with their uncertainty estimates^{3,4}. Because the frequency of drifter observations has increased since the onset of the array, an hourly product of drifter velocity estimates with uncertainties has been generated following a new estimation methodology, since 2016⁵. This paper describes the methods that have now been devised to generate a new dataset of SST estimates at hourly time steps along drifters' trajectories, aimed at accompanying the on-going hourly drifter velocity dataset⁵. A summary of the products generated by the GDP is contained in Table 1.

Hourly estimates of SST along drifters' trajectories are ultimately obtained from in situ sea water temperature observations. Estimates are obtained by least squares fitting a mathematical model of SST temporal evolution to temporally-uneven SST observations. The adopted fitting method is an adaptation of the locally weighted scatterplot smoothing method, known as LOWESS⁶. The method operates in an iterative manner in order to gradually reduce originally-uniform weights given to observations, eventually rejecting observations diagnosed as outliers. The method first generates SST estimates at the original times of the drifter SST sensor observations, and second generates SST estimates at regular top-of-hour times that typically do not coincide with the observation times. After fitting the mathematical model, the local error variance of the assumed observational process is estimated by summing the variance of the residuals from the fit and an ad hoc term aimed at taking into account the quantization error arising from temperature sensor resolution⁷. The ultimately chosen mathematical SST model is the sum of a polynomial function of order one, meant to capture non-diurnal variability, and the sum of three pairs of cosine and sine functions at harmonic frequencies of the diurnal frequency, meant to capture diurnal variability. The error variance

estimates are subsequently propagated through the least squares method to derive standard uncertainties of the model parameter and of the SST estimates. The parameters of the chosen model and of the fitting method have been chosen by analyzing two limited subsets of the global observational dataset. The choices made aim at minimizing both the mean square error calculated from the residuals of the fits and the estimates of the error variance of the observational process model. Ultimately, the variables added to the existing dataset of drifter positions and velocities consist of SST diurnal estimates, SST non-diurnal estimates, and total SST estimates. Each of these estimates is accompanied by its respective standard error estimates and specifically devised quality flag. Global statistics from our estimates indicate that the square root of the typical error variance is 0.02°C for drogued drifters' observations and 0.03°C for undrogued drifters' observations, and that the typical standard error for SST estimates is 0.016°C for drogued drifters' observations and 0.022°C for undrogued drifters' observations. There exist, however, marked geographical differences for these values across the world's ocean. The magnitudes of these uncertainties are an order of magnitude smaller than previously estimated measurement uncertainties for drifting buoys⁸ because of differences of methods.

Methods

The methods described in this paper define three levels of SST data denoted Level-0, Level-1, and Level-2, as explained in the following sections.

Data collation

In its basic configuration, a standard SVP drifter (from the Surface Velocity Program of the World Ocean Circulation Experiment) is composed of a surface float tethered to a holey-sock “drogue”, or sea anchor, centered at 15 m depth when deployed. As a result, the surface displacement of the float tracked by satellites is predominantly representative of oceanic velocity at 15 m⁹. The design of a GDP drifter has evolved, but by specification the drag area ratio of the drifters, defined as the product of the surface area of the drogue and its drag coefficient to the sum of the corresponding quantities computed for all the other components of a drifter, is 40:1. This specification ensures uniform and consistent water-following characteristics for the drifters composing the GDP array. With time, a drifter can lose its drogue and becomes “undrogued”² but still continues to transmit its position and its sensor data until it dies¹⁰. In addition to the standard SVP configuration, a number of drifters are equipped with additional sensors such as a barometer for sea-level atmospheric pressure¹¹ or a conductivity sensor to measure salinity¹² (see <https://www.aoml.noaa.gov/phod/dac/deployed.html> for the historical deployment log of the GDP). Yet, all drifters are equipped with a temperature sensor attached to the surface float and located at about 18 cm depth when at rest. Despite being an environmental variable of climate importance, sea surface temperature does not have a unique definition, and the depth at which a measurement is taken is crucial to interpret its value and variability. The definitions for near-surface seawater temperature from the Group for High Resolution Sea Surface Temperature (GHR SST, <https://www.ghrsst.org/ghrsst-data-services/products/>) would suggest to call the temperature data from surface drifters “observations of sea water temperature at a depth of 18 cm”. In the rest of this paper, for simplicity, we will refer to temperature observations from drifters, as well as temperature estimates derived from these, as SST data.

At the onset of the GDP in 1979, drifters were tracked by Argos which is both a positioning system and a data transmission system. At the beginning of the program, battery power and money was conserved by sampling location and transmitting data using different 1/3 and 2/3 schemes. As an example, data was transmitted for one day, followed by two days of no transmission, or data was transmitted for 8 hours, followed by 16 hours of no transmission¹³. Since 2000, this sampling scheme has been abandoned thanks to increased battery lives and other technological advancements. At the same time, the number of operational satellites of the Argos constellation increased with time, so that the typical time interval between two consecutive Argos fixes reduced to between 1 and 2 hours¹⁴. However, stemming from the original sampling pattern, the GDP has continued to routinely process and interpolate the location fixes and temperature observations to produce drifter locations and temperature estimates continuously along trajectories at 6-hour intervals. The general method of interpolation, called kriging³, provides an estimate of location, or of SST, at a given time as a weighted linear combinations of observations close in time (the five previous ones and the five subsequent ones in this case). Finding the optimal set of weights involves assuming a mathematical expression for a so-called structure function which is half of the variance of observation differences as a function of temporal lag. For the 6-hourly GDP product, structure functions for either location or temperature are fitted regionally and in discrete time periods to observations. As such, the kriging implementations for either location or temperature differ because of the structure function employed, and estimates of location and SST are independent from each others in the sense that no location information is used to estimate SST and vice versa. Drifter velocities are subsequently computed from the 6-hour locations by 12-hour central differencing³. The GDP started to phase out the Argos positioning system in 2014 in favor of the Global Positioning System (GPS). This system provides locations with estimated $O(10)$ -meter scale accuracy⁵ that is relayed almost instantly via the Iridium Satellite Communication system, along with sensor data, at regular temporal interval (typically hourly but not always), in contrast to Argos locations and transmissions. At the time of writing, the transition to GPS tracking and Iridium transmission is complete. A few drifters of the array were equipped with GPS receivers that transmitted their data via the Argos system⁵.

All transmitted locations and sensor data from Argos-tracked drifters were collected by Centre Localisation Satellite (CLS) which relayed them, first in near-real time to the WMO's GTS, and second to the GDP Data Assembly Center (DAC) located at the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) in Miami, Florida. The Argos data received by the DAC are organized in messages, each associated with an Argos localization from a single Argos satellite pass, with a time stamp contained within the 10 to 20 min duration of that pass^{15,16}. Each message may contain one or more sets of sensor observations, each set having its own sensor time which differs from the localization time, but is typically within plus or minus the pass' duration. Sometimes, an Argos message does not contain a location and a location time but nevertheless contains some sets of observations. In this case, the DAC assigns the location and time of the previous message to these observations. The sensor observations are subsequently processed by the DAC as follows. In the case of a message containing multiple and distinct sets of observations, the median value of all observation times and SST observations are retained for that message. For some drifters with specific sampling configurations, observations may explicitly include an age, which is a time interval that needs to be subtracted from the nominal observation time to obtain the true time of observations. Next, the DAC reorganizes Argos data as a row file, one per drifter, with each row containing in its columns Argos location (latitude and longitude), Argos location time, observation time, and observation data (SST and other sensors). Observation data are originally in sensor count but are decoded and converted throughout this process to physical units according to sensor equations found within each drifter specification sheet. Note that because several Argos satellites can be within the view of a single drifter at the same time, it is possible for the same set of observations to be transmitted by a drifter to different satellites, and to be eventually repeated in the dataset collated by the DAC, but with different locations and location times. As a result of the disconnection between Argos localization and acquisition of observations, there is no strict temporal coordination of location data and sensor data for Argos-tracked drifters.

For the modern GDP drifters that relay their data through the Iridium Satellite System, the geographical location from a GPS receiver is treated as another sensor variable, like the sensor SST, and as such is not subject to the semi-aleatory transmission schedule like with the Argos system. The data are transmitted in Short Burst Data (SBD) format which contains a number of parameters that depends on the type of drifter and the manufacturer, but typically includes date and time and sensor data including GPS when available. GPS location times and sensor data times are therefore concurrent. If a GPS position is not available at an observation time, the previous position is reported with a recorded time delay. Depending on a drifter's firmware, the GPS location sampling interval may differ from the sensor data sampling intervals. Drifter locations and sensor data are relayed in near-real time to the GDP Data Processing Center (DPC) located at the Lagrangian Drifter Laboratory (LDL) of the Scripps Institution of Oceanography and, from there, are sent the WMO's GTS after decoding the GPS and sensor data according to manufacturers' specification sheets. The drifter messages decoded by the LDL are also made available as text files to the DAC at AOML for inclusion in the GDP database.

Pre-processing and initial quality controls

Out of the data collated by the DAC, we consider for this product the SST observations from 20-Dec-1978 02:00:00 to 06-Jul-2020 22:59:31, which totals 285,886,818 data tuples of SST values and observation times. We apply a number of pre-processing and quality control procedures to these data to form what we call here Level-0 data. These initial procedures, as well as all subsequent estimation methods, are applied to all drifters irrespective of their tracking and data transmission systems. As we saw for Argos drifters, the nominal sampling patterns for drifter SST and location acquisition are essentially independent. Note that at this stage, as described in the previous section, an approximate or "raw" geographical location with varied, or unknown, uncertainty is associated with a SST data point.

The GDP DPC and DAC harvest drifter deployment sheets filled out at sea by operators, as well as conducts a number of diagnostics based on location and sensor data, in order to maintain a directory file at the DAC. This directory file lists the dates, times, and locations of trajectory starts, the dates, times, and locations of trajectory ends (i.e. drifter deaths), and the estimated dates and times of drogue losses^{10,17}. The start and death dates and times are used to truncate if needed the SST time series for data points before oceanic deployment and after oceanic death ("post-death" sensor data may exist in the transmitted data for example if a drifter had been picked up by a vessel or run aground but continued to transmit its sensor data). Next, we use some information from a quality-control step that has been in place as part of the production of the 6-hourly SST dataset. The NOAA Optimum Interpolation (OI) SST V2¹⁸ at monthly time steps is used to calculate a monthly climatology which is subsequently interpolated to the raw locations associated with the drifter SST observations. These interpolated values are then visually compared to drifter SST observations to determine a first and a last "good" SST observation per drifter trajectory, based on an expert human assessment. The corresponding dates and times of these two points are recorded in a dedicated master file for all drifters. Another master file is maintained by the DAC that lists periods of time for which it appears that an SST sensor failed temporarily, also based on the comparison to climatological values. These potential periods of sensor failure, and the periods before and after the first and last good points, are subsequently discarded from the SST observation time series. Note that this comparison to a climatology is not used to remove seemingly outlying points, but rather to determine

blocks of time for which SST observations are deemed invalid. After this stage, we find that some data records contain filling values for missing data points, which are consequently removed. Next, we find exactly repeated data tuples of SST values and observation times and the redundant data points are removed. Next, we find two or more SST observations existing at the same time for a single drifter. In that case, all observations are kept and will be processed for obtaining Level-1 estimates at the same times. Next, we identify a number of drifter SST time series with only a single point, which are removed, and constant-value time series which are also removed. In the end, the final Level-0 dataset consists of 197,916,695 tuples of SST and time data originating from 24,597 drifter trajectories.

As a result of the differing technologies of the data transmission systems (Argos and Iridium), of the number of different drifter manufacturers for the GDP, and of changing firmwares with time, the Level-0 dataset of time series of SST observations is heterogeneous in its sampling intervals and apparent levels of noisiness. Two-dimensional histograms of occurrences of time differences and absolute temperature differences between two subsequent data points for all trajectories (Figure 1) generally indicate that larger absolute SST differences are found for smaller time differences, for both Iridium and Argos drifters. For Iridium drifters, time differences are concentrated around multiples of one hour or 30 min, but with deviations from these because of possible delays of GPS signal acquisitions compared to a specified schedule. The distribution of time differences for Argos drifters is more continuous but exhibits local peaks near one hour and 101 min, the latter corresponding to the orbital period for an Argos satellite⁵. Even when considering the distribution of the median of time differences (or sampling interval) per trajectory, the Argos drifters exhibit a much varied set of values compared to the Iridium drifters (Figure 2).

Model of SST temporal evolution

We seek to obtain SST estimates by fitting a local temporal model to the temperature data acquired along drifter trajectories. In-situ observations of SST from drifting or moored platforms, as well as remote sensing observations, suggest that two types of temporal variability typically co-exist: a relatively fast evolution on a diurnal time scale, sometimes referred to as a diurnal warming, and a relatively slower background evolution. For this background evolution (also referred to as non-diurnal in the rest of this manuscript), there is a priori no expectation of a dominant physical process acting at all times and all places. As such, it is reasonable to model this evolution with a local polynomial model as an approximation of a Taylor series expansion of an unknown underlying function¹⁹. For the diurnal evolution, we follow a number of previous studies^{20–23} and model this evolution as the sum of cosine functions with fundamental frequency $\omega = 2\pi$ radians per day. In contrast to some previous studies however, our diurnal model is exactly periodic in the sense that it is locally zero-mean. A mean SST value and a possible difference of SST between the beginning and the end of a diurnal period will be both captured by the background non-diurnal polynomial model (as it will be at least of order 1). In addition, the amplitudes and phases of each of the cosine functions contributing to the diurnal model are not constant within a day, but rather vary locally since they are fitted at every time step using data within a sliding window centered on that time step. This diurnal model allows us to accommodate various environmental conditions (e.g. momentum and heat fluxes) affecting the shape of the diurnal signal in time and space as a drifter is advected by ocean currents. Note that since the diurnal SST estimate is locally zero-mean and does not represent solely a diurnal warming, the contemporaneous non-diurnal SST estimate differs from what is called a foundation temperature, that is a temperature free of diurnal temperature variability. In other words, the non-diurnal SST estimate typically contains the local mean of the SST diurnal variability.

In summary, the complete SST model is the sum of a polynomial s_P of order P , and a sum s_D of N cosine functions at harmonic frequencies of the diurnal frequency $\omega = 2\pi$ radians per day. Next, we consider that a number of drifter SST observations s_i , found in the temporal vicinity of a single observation s_k at time t_k , are generated by the process

$$s_i = s_m(t_i; t_k) + \sigma_i \varepsilon_i. \quad (1)$$

The noise component, ε_i , is expected to be zero-mean, to have unit variance [$E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = 1$], and to be independent of other noise components at other times. The noise is locally scaled by the square root of σ_i^2 which is the error variance of the observations, is conditional to time t_i , and will be estimated a posteriori.

The temporal evolution model is

$$s_m(t_i; t_k) = s_P(t_i; t_k) + s_D(t_i; t_k) \quad (2)$$

$$= \sum_{p=0}^P s_{p,k}(t_i - t_k)^p + \sum_{n=1}^N A_{n,k} \cos[n\omega(t_i - t_k) + \phi_{n,k}] \quad (3)$$

$$= \sum_{p=0}^P s_{p,k}(t_i - t_k)^p + \sum_{n=1}^N [\alpha_{n,k} \cos n\omega(t_i - t_k) + \beta_{n,k} \sin n\omega(t_i - t_k)], \quad (4)$$

with

$$\alpha_{n,k} = A_{n,k} \cos \phi_{n,k}, \quad (5)$$

$$\beta_{n,k} = -A_{n,k} \sin \phi_{n,k}. \quad (6)$$

The last form (4) of the model shows that the $P + 1 + 2N$ parameters of this model can be estimated by forming a linear system of equations. Ultimately, once the model parameters are estimated, the SST estimate itself at time t_k is evaluated by setting $t = t_k$ in (4) to obtain:

$$\hat{s}_{m,k} \equiv s_m(t_k; t_k) = s_{0,k} + \sum_{n=1}^N \alpha_{n,k}, \quad (7)$$

which involves only $N + 1$ parameters of the $P + 1 + 2N$ estimated parameters. The other $N + P$ parameters nevertheless provide further physical information such as the SST tendency for the non-diurnal evolution (e.g. $s_{1,k} = \partial s_P(t_k; t_k) / \partial t$ if $P \geq 1$) or the phase and amplitude of the diurnal harmonics:

$$\phi_{n,k} = \arctan \left(\frac{-\beta_{n,k}}{\alpha_{n,k}} \right), \quad (8)$$

$$A_{n,k} = \frac{\alpha_{n,k}}{\cos \phi_{n,k}}. \quad (9)$$

Ultimately, we will select $P = 1$ and $N = 3$ on the basis of the analyses of two subsets of surface drifters, as explained in the section *Model selection*. As explained further in the next section, the model is first fitted at all original observation times of a drifter trajectory in an iterative manner in order to gradually adjust the weight of the data in the estimation, as well as identify outlier data points. After a given number of iterations, the model is ultimately fitted once at regular, top-of-the-hour, times that do not typically coincide with the original times. This method therefore defines three levels of data denoted thereafter Level-0, Level-1, and Level-2:

- Level-0 corresponds to the original, temporally unevenly distributed data as reported by the SST sensor and transmitted to the GDP DAC via Service Argos system or by the LDL at SIO.
- Level-1 corresponds to our SST estimates calculated from (7) after fitting model (2) at the same unevenly distributed times as Level-0;
- Level-2 corresponds to SST estimates at a regular hourly interval, at the top of each hour, also calculated from (7) after fitting model (2); Level-2 estimates of this current data release corresponds to the same times as the position and velocity estimates for drifters of the of the GDP hourly dataset, release 2⁵.

This paper describes the derivations of Level-1 and Level-2 datasets and announces the release of the Level-2 data as a research product.

Estimation of model parameters and SST

The devised method to estimate SST continuously along a drifter trajectory is adapted from the method known as the locally weighted scatterplot smoothing or LOWESS⁶. This method is iterative, and thus robust to outlying data points which are commonly observed in SST time series from surface drifters (see an example in Figure 3). Our method goes as follows. For a given SST time series from a single drifter, for each SST observation s_k at time t_k , we compute by weighted least squares the $P + 1 + 2N$ parameters of the model s_m that minimize

$$\sum_{i=1}^K [s_i - s_m(t_i; t_k)]^2 K_{h_k, i}, \quad (10)$$

where $K_{h_k, i}$, is a set of weights given by

$$K_{h_k, i} = K \left(\frac{t_i - t_k}{h_k} \right), \quad (11)$$

with K the tricube kernel function⁶:

$$K(\tau) = (1 - |\tau|^3)^3 I_{[-1, 1]}(\tau), \quad \text{with} \quad I_{[-1, 1]}(\tau) = \begin{cases} 1, & |\tau| \leq 1 \\ 0, & |\tau| > 1. \end{cases} \quad (12)$$

In (11), h_k is called the bandwidth of the kernel K , that is the half-width of the temporal window around the observation time t_k within which the weights $K_{h_k,i}$ are different from zero. The least squares calculation therefore involves practically only those data points with non-zero weights. Because the complete model s_m includes a diurnal oscillation, we initially set $h_k = 1$ day for all points, but this value is automatically and gradually increased as needed in 1-hour steps in order to include more data points to ensure that the least squares system of equations is not undetermined, up to a maximum value of 2 days. If not enough data points are available within the temporal window of maximum length, then no SST estimate is obtained. For the data selected to match the GDP hourly dataset version 1.04c (released in February 2021, with data through June 2020), fewer than 0.4% of the data points require a half-bandwidth longer than 1 day. With the addition of the SST estimates, we call this update version 2.00.

Using matrix notation for convenience, the minimization problem (10) can be written

$$\min_{\beta} (\mathbf{s} - \mathbf{X}\beta)^T \mathbf{W}(\mathbf{s} - \mathbf{X}\beta), \quad (13)$$

with solution

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{s}. \quad (14)$$

In (13), \mathbf{X} is the design matrix for linear model (4):

$$\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \mathbf{X}_3], \quad (15)$$

with

$$\mathbf{X}_1 = \begin{bmatrix} 1 & (t_1 - t_k) & \cdots & (t_1 - t_k)^P \\ \vdots & \vdots & & \vdots \\ 1 & (t_i - t_k) & \cdots & (t_i - t_k)^P \\ \vdots & \vdots & & \vdots \\ 1 & (t_K - t_k) & \cdots & (t_K - t_k)^P \end{bmatrix}, \quad (16)$$

$$\mathbf{X}_2 = \begin{bmatrix} \cos\{\omega(t_1 - t_k)\} & \cos\{2\omega(t_1 - t_k)\} & \cdots & \cos\{N\omega(t_1 - t_k)\} \\ \vdots & \vdots & & \vdots \\ \cos\{\omega(t_i - t_k)\} & \cos\{2\omega(t_i - t_k)\} & \cdots & \cos\{N\omega(t_i - t_k)\} \\ \vdots & \vdots & & \vdots \\ \cos\{\omega(t_K - t_k)\} & \cos\{2\omega(t_K - t_k)\} & \cdots & \cos\{N\omega(t_K - t_k)\} \end{bmatrix}, \quad (17)$$

$$\mathbf{X}_3 = \begin{bmatrix} \sin\{\omega(t_1 - t_k)\} & \sin\{2\omega(t_1 - t_k)\} & \cdots & \sin\{N\omega(t_1 - t_k)\} \\ \vdots & \vdots & & \vdots \\ \sin\{\omega(t_i - t_k)\} & \sin\{2\omega(t_i - t_k)\} & \cdots & \sin\{N\omega(t_i - t_k)\} \\ \vdots & \vdots & & \vdots \\ \sin\{\omega(t_K - t_k)\} & \sin\{2\omega(t_K - t_k)\} & \cdots & \sin\{N\omega(t_K - t_k)\} \end{bmatrix}. \quad (18)$$

The weighing matrix \mathbf{W} is defined by

$$\mathbf{W} = \text{diag} \{K_{h_k,i}\}, \quad (19)$$

and \mathbf{s} and β are the vector of data points and the vector of dimension $P + 1 + 2N$ of parameters to be estimated, respectively:

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}, \quad \beta = \begin{bmatrix} s_{0,k} \\ \vdots \\ s_{P,k} \\ \alpha_{1,k} \\ \vdots \\ \alpha_{N,k} \\ \beta_{1,k} \\ \vdots \\ \beta_{N,k} \end{bmatrix}. \quad (20)$$

Next, following the initial iteration of estimating the model parameters and calculating the corresponding SST estimates at all times t_k , we consider the residuals at all observation times:

$$r_k = s_k - \hat{s}_{m,k}, \quad (21)$$

and compute M , the median value of the distribution of their absolute values. A set of robust weights are next calculated as

$$\delta_k = B\left(\frac{r_k}{DM}\right), \quad (22)$$

where

$$B(t) = (1 - |t|^2)^2 I_{[-1,1]}(t), \quad (23)$$

is the biweight kernel function⁶ and D is a real factor to be determined.

The next step of the method consists in iterating the weighted least squares estimation of all parameters of the model at all times t_k , but this time using modified weights $\delta_i K_{h_k,i}$ instead of $K_{h_k,i}$ in (10). How many data points are down-weighted is dependent on the coefficient D in the denominator of (22) which is typically set to 6⁶ but here is set to 14, as discussed in the section *Model selection*. The number of iterations is chosen here to be three after the initial least squares estimation without modified weights. The modified weights can effectively become zero when $\delta_k = 0$, that is when the absolute value of a residual is larger than D times M for a given SST time series associated with one drifter trajectory. This implies that such data points are ultimately not used for any estimation but SST values at the corresponding time are nevertheless obtained using all available non-zero-weighted data points within the temporal window centered on any of these points. This method effectively flags as outliers some of the Level-0 SST data point like a “de-spiking” procedure would do, for example by applying a median filter³. An example of flagged outliers in a Level-0 drifter SST time series is shown in Figure 3. One implicit assumption of using (22) to modify the weights of the data is that all residuals from a given drifter SST time series originate from a common distribution, or equivalently that the statistics of the observations are constant within a given trajectory. This assumption may be violated if a drifter trajectory is long enough to experience environmental condition changes, or the characteristics of the SST sensor changes in an undetected fashion. An illustration of a Level-1 estimation step is provided in Figure 4.

Finally, as the last step of the method, the SST model (4) with the same number of parameters is fitted to the data but at times t_k corresponding to the top of the hour UTC (00:00, 01:00, etc.), in one iteration with weights $\delta_i K_{h_k,i}$ where the δ_i were calculated prior to the last iteration for the original data times (not posterior). As before, the bandwidth is set to 1 day but is allowed to increase in increments of one hour, up to two days, to make sure the linear estimation problem is not under-determined. This last step generates the final Level-2 data product. An illustration of Level-2 estimated data is provided in Figure 5.

Error variance estimates and uncertainty estimates

As part of the method, we quantify the uncertainties of the parameter estimates and thus the uncertainties of the diurnal SST estimates, of the non-diurnal SST estimates, and of the total SST estimates. Formally, the covariance matrix of the weighted least squares solution at time t_k is

$$\mathbf{C}_\beta \equiv \text{Var}(\hat{\beta}) = (\mathbf{X}\mathbf{W}^*\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}^*\Sigma\mathbf{W}^*\mathbf{X})(\mathbf{X}\mathbf{W}^*\mathbf{X})^{-1}, \quad (24)$$

where \mathbf{W}^* is the weight matrix containing in its diagonal the modified weights $\delta_i K_{h_k}(t_i - t_k)$ from the penultimate iteration of the least squares estimation, and Σ is the unknown covariance matrix of the observation errors from the process model (1). In order to proceed, we assume local homoscedasticity and that the errors are independent which results in $\Sigma = \sigma^2(t_k)\mathbf{I}$, where the local error variance $\sigma^2(t_k)$ is unknown and needs to be estimated. In the case of a local polynomial regression of order P , it is recommended¹⁹ to re-conduct a polynomial fit of order $P + 2$ and estimating the error variance from the residuals of that fit. In our case, which is not a sole polynomial regression since model (2) also includes trigonometric functions, the optimal course of action is unclear. Yet, to proceed, we classically calculate a first estimate of the error variance from the normalized weighted residual sum of squares:

$$\hat{\sigma}_1^2(t_k) = \frac{(\mathbf{s} - \mathbf{X}\hat{\beta})^T \mathbf{W}^* (\mathbf{s} - \mathbf{X}\hat{\beta})}{\text{tr}\{\mathbf{W}^* - \mathbf{W}^* \mathbf{X} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*\}} \quad (25)$$

$$= \frac{\sum_i [s_i - \hat{s}_m(t_i; t_k)]^2 \delta_i K_{h_k}(t_i - t_k)}{\nu}. \quad (26)$$

The denominator of (25), referred to as ν in (26), is the effective number of degrees of freedom for the residuals for weighted least squares¹⁹. For ordinary least squares, ν would simply be the number of data points used to calculate $\hat{\beta}$ minus the number of parameters of the model ($P + 1 + 2N$), but for weighted least squares cases, ν is smaller.

The majority of drifters from the GDP database are equipped with temperature sensors returning a bit count n used to calculate SST following the sensor equation:

$$SST = an + b, \quad (27)$$

where a is the resolution of the temperature sensor. As such, the Level-0 data should be rounded due to the resolution of the instrument recording. In the signal processing literature this is known as quantization, and has the effect of removing high resolution information in the data. As a result, the estimated error variance [$\hat{\sigma}_1^2(t_k)$, (25)] should be increased to reflect the additional uncertainty created through quantization, as this information cannot be recovered. In the extreme case that the input data is the same value within a full window length then the increase to the error variance is $a^2/12$ ⁷, following from the properties of the uniform distribution. As a result, adjusting for resolution, our total error variance is

$$\hat{\sigma}^2(t_k) = \hat{\sigma}_1^2(t_k) + \frac{a^2}{12}. \quad (28)$$

This adjustment is conservative, in that the effect of resolution to the error variance will decrease as the input values have more variance⁷. For simplicity, we use the conservative adjustment proposed above as this ensures the reported standard errors always include the resolution effect which should not be ignored.

For about 85% of the drifters, representing 83% of the Level-1 estimates, the resolution a can be obtained from the individual specification sheets provided by the manufacturers. We identify in this way 179 different resolutions, ranging from 0.00260877°C to 0.17°C. The three most common resolutions are 0.01°C, 0.05°C, and 0.08°C. For the remaining 15% of the drifters, some have an SST equation which is not a linear function of a sensor single bit count and the impact of the quantization error cannot be simply modeled as in (28). Some other drifters have an unknown resolution because of the lack of available metadata. For these drifters, we estimate the resolution from the data as follows: we consider the time series of absolute SST difference, bin these differences in 0.001°C bins, and assign the resolution to the most common value that is not zero. In this way, the three most commonly estimated resolutions are 0.05°C, 0.08°C., and 0.043°C. This method is successful in 92% of cases when tested on the data of the drifters for which the resolution is known from the metadata. The overall distribution of all resolution values, as well as their temporal distribution, are illustrated in Figure 6.

We found it necessary to consider the resolution error for two reasons. First, since we have allowed our estimation algorithm to obtain a solution with as little data points as the number of model parameters to be estimated, and because of numerical precision errors, we find a small number of instances (0.33% of the Level-1 data) for which ν , and therefore the first estimated error variance $\hat{\sigma}_1^2(t_k)$, is small and negative. These instances are resolved by adding the resolution error. Second, in some other instances (0.22% of the Level-1 results, see Figure 7), we find that the residuals, and hence the estimated error variance and the parameter uncertainties, are locally zero within numerical precision despite an ample number of data points available for the estimation. This occurs when the Level-0 SST data does not change in value within the estimation window for reasons which are not clear. Once again, these instances are resolved by adding the resolution error, resulting in more realistic error estimates. Nevertheless, these two instances define two populations of the results that are clearly separated within a two-dimensional histogram of $\hat{\sigma}_1^2(t_k)$ and ν (Figure 7). As such, we can flag these results using an empirical and ad-hoc condition:

$$\log_{10}[\hat{\sigma}_1^2]^{1/2} < -\frac{1}{2} \log_{10} |\nu| - 10. \quad (29)$$

The final error variance estimate $\hat{\sigma}^2(t_k)$ (28) is a function of time t_k and specific to a drifter because of the sensor resolution. This estimate is subsequently used to calculate an estimate of the local covariance matrix of the observations $\hat{\Sigma} = \hat{\sigma}^2(t_k)\mathbf{I}$ and to calculate the covariance matrix \mathbf{C}_β [expression (24)].

From the expression for the SST estimate (7), its variance is

$$\sigma_m^2 \equiv \text{Var}[s_m(t_k; t_k)] = \text{Var}[s_{0,k} + \sum_{n=1}^N \alpha_{n,k}] \quad (30)$$

$$= \text{Var}[s_{0,k}] + 2\text{Cov}[(s_{0,k}), (\sum_{n=1}^N \alpha_{n,k})] + \text{Var}[\sum_{n=1}^N \alpha_{n,k}] \quad (31)$$

$$= \sigma_p^2 + 2\text{Cov}[(s_{0,k}), (\sum_{n=1}^N \alpha_{n,k})] + \sigma_D^2. \quad (32)$$

This last expression describes how the variance of the total SST estimate (σ_m^2) is the sum of the variance of the non-diurnal SST estimate (σ_p^2 , containing one term), of the variance of the diurnal estimate (σ_D^2 , containing N^2 terms), and of $2N$ additional cross-covariance terms. The $(N + 1)^2$ needed terms to estimate σ_m^2 , σ_p^2 , and σ_D^2 are extracted and summed appropriately from

the calculated covariance matrix C_β [expression (24)] at each time step. The square root of each of these three estimated variances, referred subsequently as $\hat{\sigma}_m$, $\hat{\sigma}_p$, and $\hat{\sigma}_D$, define the standard errors, or standard uncertainties, of the three SST estimates. Illustrations of estimated square roots of error variances and SST uncertainties is provided in Figures 4, 5, and 8. Additional discussions of error variance and uncertainty estimates are provided in section *Interpretation of uncertainty estimates*.

Model selection

In order to fit the total SST model to the data, choices need to be made for the order P of the polynomial of the non-diurnal model and the number N of harmonics of the diurnal model. We consider a total of 14 models with P varying between 0 and 3, and N between 2 and 6 (Table 2). We test and assess the performances of the models by fitting them to two limited subsets of GDP drifters as it would be computationally prohibitive to conduct tests on the entire Level-0 data.

The first subset is from the Salinity Processes in the Upper Ocean Regional Study (SPURS) in the subtropical North Atlantic^{12,24}. The drifters released as part of SPURS were manufactured by Pacific Gyre Inc. but differed from standard SVP-type drifters of the GDP⁹. Instead of a temperature sensor on their buoys, these drifters were equipped with an unpumped Sea-Bird Electronics SBE37-SI MicroCAT CTD placed underneath the surface buoy with its sensors located at a depth of 50 cm. The Microcat instruments were set to acquire conductivity and temperature at 30-min intervals by sampling once a minute for 5 min and averaging the values. According to the manufacturer, the initial accuracy and resolution of the temperature sensor are 0.002°C and 0.0001°C respectively, but the data transmitted and relayed to the DAC exhibit a resolution of 0.01°C. For this study, we select 80 drifters which generated temperature data (considered to be SST observations) for time periods spanning between 29 and 660 days. These drifters transmitted their locations and sensor data via the Argos satellite system, including their position data from GPS receivers. These GPS data were previously used as a test set to devise the methodology being used to generate the global dataset of hourly position and velocity for the GDP⁵. However, here, the original Argos message data files for these drifters are re-processed to eliminate redundant and corrupted data by taking into consideration a previously ignored checksum flag indicating the integrity of Argos data transmissions. Next, the SST time series are further truncated to match the beginnings and ends of the regular hourly time series of position and velocity for these drifters, as well as truncated for their first and last good data points as diagnosed by the QC procedures of the DAC. The resulting dataset consists of 80 time series of SST at uneven temporal intervals multiple of 30 minutes, totalling nearly 1.26M data points over 29,018 drifter days.

The second subset of drifter SST data, hereafter referred to as the “test” subset, is built from the global database by selecting at random 14 drifters within each 10° latitude band between -70°S and 70°N with an average SST temporal sampling interval of between 50 and 70 minutes, resulting in a total of 98 individual SST time series which are further truncated in time for deployment times etc. The resulting dataset consists of 98 time series of SST at uneven temporal intervals, totalling 697,045 data points over 29,408 drifter days. These test drifters constitute a limited subset but represent a variety of drifter types deployed between years 2000 and 2019. Fifty of them are drifters with barometer (SVPB type), and 48 are standard SVP drifters. Fifty-seven of them were Iridium drifters and 41 Argos drifters. The test drifters were built by a variety of manufacturers: 18 by DBi, 19 by Metocean, 9 by Clearwater, 30 by Pacific Gyre, 18 by Scripps Institution of Oceanography, 2 by Technocean, 1 by Marlin-Yug, and 1 by NKE. Finally, the stated resolution of their SST sensors as specified by their respective specification sheets were varied: 0.01°C for 68 of them, 0.05°C for 20, 0.04°C for 2, 0.043°C for 2, 0.04329°C for 1, 0.04343°C for 1, 0.08°C for one, and unknown for three of them.

We proceed to fit the 14 models listed in Table 2 to the SPURS and test subsets of drifter SST time series, and subsequently consider two statistics calculated for each time series. The first statistic is the weighted root mean square error (WRMSE) calculated from the residuals of a given fit. For this calculation, the weights are the robust weights calculated by the algorithm described previously after the penultimate iteration (that is the weights calculated before the last estimation at the original times), but with a further normalization to ensure that weights sum to one:

$$\text{WRMSE} = \left[\sum_{k=1}^M w_k (s_k - \hat{s}_{m,k})^2 \right]^{1/2}, \quad (33)$$

$$w_k = \frac{\delta_k K_{k,k}}{\sum_{i=1}^M \delta_i K_{i,i}}. \quad (34)$$

The second statistic considered is the square root of the weighted median of the error variance estimates: after the final iteration, we consider the error variance estimates $\hat{\sigma}^2(t_k)$ given by (28) and using the weights defined by (34), we calculate the weighted median defined as the value $\hat{\sigma}^2(t_n)$ such that

$$\sum_{k=1}^{n-1} w_k \leq 1/2 \quad \text{and} \quad \sum_{k=n+1}^M w_k \leq 1/2. \quad (35)$$

For these two statistics, using weighted calculations effectively filters out the outlier data points diagnosed from the method (that for which $w_k = 0$). We also tried non-weighted calculations that include all data points: besides shifting numerical values of the results in the sense of worsening performances, this did not change the relative performances of the models nor our overall conclusions and model selection choice.

We find that varying the number of parameters of either the diurnal model or the non-diurnal model affects the two statistics differently. We present the results in Figure 9, displaying the WRMSE and the square root of the weighted median error variance, both in unit of degrees Celsius. The figures display scatter plots of the two statistics averaged over each of the subsets, along with ellipses representing the 95% confidence intervals for the means in order to illustrate the scatter of the results. Not surprisingly, the scatter of the results is relatively smaller for the SPURS subset for which the SST records have the same nominal characteristics compared to the test subset composed of heterogeneous records.

For both data sets, we find that for a fixed number of harmonics of the diurnal model, increasing the polynomial order of the non-diurnal model (going right through the columns of Table 2) reduces the error variance with little change to the WRMSE. The most dramatic reduction occurs when going from models for which $P = 0$ (models 1, 2, and 3) to models for which $P = 1$ (models 4 and higher) for which the square root of the weighted median error variance is at least approximately halved. Conversely, for a fixed order of the polynomial non-diurnal model, we find that increasing the number of harmonics (going down the rows of Table 2) decreases the WRMSE with little change to the error variance. Considering these two general tendencies together, as well as the scatter of the results as depicted by the ellipses, we find that model 5 ($P = 1$ and $N = 3$) provides a good balance between the two statistics. Further, we find that from model 5, no significant improvement is obtained for the WRMSE error variance by increasing the polynomial order from 1 to 2 (going to model 8), and no significant improvement is obtained for the error variance by increasing the number of harmonics from 3 to 4 (going to model 6). Significant improvements are obtained for both statistics by both increasing the polynomial order from 1 to 2 and the number of harmonics from 3 to 4 (going from model 5 to model 9) for the SPURS subset but not for the test subset, which is expected to be representative of a much greater fraction of the total data. We also tested models with $P = 1$ and $N = 5, 6$ (models 13 and 14) but these, while reducing significantly the WRMSE from model 5, did not reduce significantly the error variance, and started to show larger error variances for the test subset. As a result, model 5 is our final choice of model to be fitted to the entire SST drifter dataset to generate the Level-1 and Level-2 datasets.

We now discuss briefly the choice of the bandwidth length $[h_k, (11)]$ and the choice of the factor D for the robust weights [see (22)]. The sensitivity of the results to these choices is summarized in Figure 10 for model 5 only. The choice of h_k technically implies that data points within a $2h_k$ window centered on the estimation time are considered [eqs. (10) and (11)]. Yet, because the weighing window is not uniform but a tricube kernel, the effective number of degrees of freedom used for each estimation is closer to the number of data points one would find in a uniform window of length h_k . Here, our choice $h_k = 1$ day is based on observations that the characteristics of diurnal SST oscillations change on a daily time scale²⁵. Yet, we examine the summary statistics for model 5 for h_k varying between 0.25 and 1.25 days at 0.25 day interval for the test subset (Figure 10). We find that decreasing h_k to less than 1 day significantly decreases the error variance yet does not decrease the WRMSE, and thus does not overall improve the performances of model 5. In contrast, we find that increasing h_k to 1.25 days significantly increases the WRMSE and increases the error variance. The results for the SPURS subset are similar (not shown). These overall results therefore suggest that $h_k = 1$ day is an appropriate choice for the bandwidth.

The choice of the factor D in the denominator of the biweight kernel for calculating the robust weights [eq. (22)] effectively sets the threshold for labeling data points as outliers. Our final choice of $D = 14$ is compared to alternatively choosing D between 4 and 20 at intervals of 2. The sensitivity of the summary statistics to the value of D is displayed in Figure 10 for the test subset. We find that varying D has a modest impact on the performances of model 5 and only for D less than 6 does model 5 exhibit significantly better WRMSE, but no better standard deviation error. We also consider the ensemble average of the fraction of data points not labeled as outliers as a function of the choice of D (Figure 10), and find that this fraction starts to decrease strongly as D decreases from 8. In the original LOWESS method⁶, D is set to 6 without justification, and such a choice in our case would result in around 10% of the data points labeled as outliers. In the end, we settled on $D = 14$ which results in only between 1% and 4% of the data points being labeled as outliers, but maintains approximately the performance of model 5 compared to $D = 6$. The results are similar for the SPURS subset (not shown).

Quality indication

The Level-2 data product is intended to provide SST estimates contemporaneous to the estimated positions and velocities of drifters at hourly top-of-the-hour times from the hourly GDP dataset version 1.04c⁵, which with SST now included we shall call version 2. Since the sampling of SST sensors onboard drifters can be independent from the positioning, it sometimes occurs that our methodology is able to provide an SST estimate at times when no location estimate is available. Since there is little use for an SST estimate with no associated location estimate, these are not included in the Level-2 data product (see Table 1).

We devise three different quality indication flag schemes, one for each component of SST (non-diurnal, diurnal, total),

with flag values ranging from 0 (worst) to 5 (best). For all three schemes, when no SST estimate could be obtained from the methodology (for example for lack of enough Level-0 data within the sliding temporal window), or when SST data was simply not transmitted by a drifter (as an example because of a faulty sensor), the estimate is assigned quality flag 0 (and the NetCDF file contains a standard filling value). When an SST estimate could be obtained but not an SST uncertainty estimate, the estimate is assigned quality flag 1 (and the NetCDF file contains a filling value for the uncertainty estimate).

For higher flag values, the schemes for the non-diurnal SST estimates and for the total SST estimates are the same, as illustrated in Figure 11. When an SST estimate and an uncertainty estimate both exist, the quality flag is based on the relative position of the interval formed by the SST estimate plus or minus its standard error estimate with respect to the $[-2,50]^{\circ}\text{C}$ range of physically-acceptable temperature values²⁶. If the estimated interval is completely contained within this range, the assigned quality flag is the highest, at 5. If one or two end points of the interval are located outside of the range but the SST estimate is inside the range, the assigned quality flag is 4. If the SST estimate is outside the range but one of the end point of the interval is within the range, the assigned quality flag is 3. Finally, if the interval is located completely outside the physical range, then the quality flag is 2.

The quality flag scheme for the diurnal SST estimates differs from the scheme described above because a diurnal SST estimate is an anomaly around zero for which a range of physically plausible values is not straightforward to define. A climatology of SST diurnal variability²³ constructed by fitting a model to temperature observations from drifters within zonal bands, by seasons, and by environmental categories (clear or cloudy sky, wind speed) provides amplitude of SST diurnal anomalies no larger than 2.4°C (from observations) or 0.689°C (from modeled values). Locally in coastal regions, diurnal warming as large as 6.6°C has been detected²⁷. As such, rather than defining here an acceptable amplitude threshold for diurnal SST anomalies, we consider three criteria for the quality flag of a diurnal SST estimate:

1. Is the diurnal anomaly estimate smaller than its standard error estimate?
2. Is the standard error estimate for the diurnal estimate smaller than 1°C ?
3. Were more than 24 Level-0 data points used to obtain an estimate?

As illustrated in Figure 12, criteria (1) and (2) define specific sub-regions in the parameter space defined by the absolute value of the diurnal estimates and the value of the standard error estimate of the diurnal estimate. In contrast, criterion (3) does not strictly defines a sub-region in that parameter space, but rather an average region which can be visualized by mapping in that space the average number of data points used for the estimations. On average, estimates obtained with 24 data points or more are found in the parameter space for which the diurnal anomaly estimates are smaller than 10° and the standard error of the estimates are most often smaller than 1° . In conclusion, we use the three criteria listed above to define self-exclusive quality flags as follows: a quality flag 5 indicates that all criteria (1), (2), and (3) are fulfilled; a quality flag 4 indicates that (1) and (2) are fulfilled but not (3); a quality flag 3 indicates that (1) is fulfilled but not (2) nor (3); and quality flag 2 indicates that none are fulfilled.

The inventory of Level-2 estimates for each type (total, non-diurnal, and diurnal) and each quality flag class (0 to 5) is provided in Table 3. The number of position and velocity estimates for the GDP hourly dataset version 2.0 is 165,754,333 from 17,324 individual drifter trajectories. Of this target number, 95.59% are with a quality flag 5 for the total SST estimates, 95.60% are with a quality flag 5 for the non-diurnal SST estimates, but only 75.58% are with quality flag 5 for the diurnal SST estimates. Note that estimates of total SST and non-diurnal SST with quality flag 3 or 2 are outside the physically-acceptable range of values and should be used and interpreted with extreme caution. We assessed that diurnal SST estimates with quality flag 5 are plausible but we could not conclude the same for lesser quality flags.

Interpretation of uncertainty estimates

In order to interpret our uncertainty estimates, we examine the distribution of the residuals of all model fits, normalized by their associated estimates of error standard deviations. This constitutes an assessment of the distribution of the error term ε_i of the process model (1):

$$\hat{\varepsilon}_i = \frac{s_k - \hat{s}_{m,k}}{\hat{\sigma}(t_k)}. \quad (36)$$

The results are shown in Figure 13 for both the SPURS and test drifter subsets. For both sets, the distributions are never Gaussian for any of the models. The distributions are nearly centered but exhibit central peaks more narrow than Gaussian distributions with the same means and standard deviations (only comparisons to model 5 are shown). We observe that increasing the number of harmonics of the diurnal oscillation model consistently renders the peak of the residual distribution to be narrower and higher, and the tails to be slightly lighter. The opposite is true when increasing the order of the non-diurnal polynomial model, while still being non Gaussian in the sense of exhibiting a higher kurtosis. We find that a t location-scale distribution

(also known as non-standardized Student's t distribution), previously used to model Argos location errors⁵, is a better fit to the observed distributions than Gaussian distributions, yet still does not completely capture their shapes (not shown). An implication of the non-gaussianity of the normalized residuals is that the error term ε_i of the process model (1) is also not Gaussian-distributed. As a result, a classic least squares estimation of the parameters of the models would tend to give too much weight to outliers in the data. Fortunately, we are applying an iterative least squares estimation method based on the LOWESS⁶ which is expected to temper such outliers, but the exact impact on the estimation is difficult to quantify here.

Nevertheless, a further implication of the non-gaussianity is that caution should be taken when interpreting the standard errors for the SST estimates described above: whereas for Gaussian-distributed errors one standard error can be used to calculate a 68% confidence interval for an estimate, in our case, a standard error represents an interval encompassing more probable values of the true unknown values of a quantity, thus a more conservative confidence interval. As shown in Figure 13 for model 5, the 16-th and 84-th percentiles, encompassing 68% of the residual distribution, define an interval narrower than the interval defined by plus or minus one sample standard deviation around the sample mean. Plus or minus one standard deviation actually encompasses approximately 78% of the distribution of the residuals for model 5 (and approximately the same percentage for the other models, not shown). In other words, the standard error for our estimates can be interpreted as being representative of a 78% confidence interval rather than a 68% confidence interval. In contrast, the 2.5-th and 97.5-th percentiles, encompassing 95% of the distribution, define an interval slightly wider but close to the one defined by plus or minus 1.96 sample standard deviation around the sample mean, which encompasses approximately 94% of the distribution of the residuals for model 5 (and approximately the same percentage for the other models, not shown). In conclusion, considering 1.96 standard errors to quantify uncertainty in this case happens to represent an approximate 95% confidence interval, as would be the case if the errors were Gaussian-distributed. Note that for the Level-2 hourly product (Table 1), the uncertainty estimates provided for location and velocity is 95% confidence intervals, whereas for SST estimates the uncertainty estimates are standard error estimates.

Global characteristics of error variance estimates and uncertainty estimates

In Figure 14, we examine the distribution of error variance estimates from residuals [$\hat{\sigma}_r^2$, eq. (25)] not including data points for which the error estimation failed, and the distribution of total error variance estimates incorporating the resolution error variance [$\hat{\sigma}^2$, eq. (28)] for Level-2 estimates. We show the distributions for Level-2 estimates only because the ones for Level-1 estimates are extremely similar. We also report some statistics rounded to the nearest 0.001 in Table 4, which differ by no more than 0.001°C between Level-1 and Level-2 estimates. Based on the distributions in Figure 14, we assess that the mode value, or most probable value, of the square root of the error variance estimates from residuals is 0.020°C for drogued drifters, but is 50% larger at 0.030°C for undrogued drifters. Over all data, the mode value is 0.026°C. Further, we assess that the mode value of the square root of the error variance estimates incorporating the resolution error variance is 0.031°C for drogued drifters and 0.036°C for undrogued drifters. Over all data, the mode value of the total error variance estimates is 0.033°C. Median values of each of these variables are typically higher by a few 1/1000-th of a degree (See Table 4): the overall median value of the square root of the total error variance estimates is 0.036°C. The distribution of the total error variance is however not unimodal (Figure 14, right) because of the resolution error variance is dominated by a few discrete values (Figure 6).

The error variance estimates are however very heterogeneous in space, which is revealed when these estimates are averaged in half-degree geographical bins (Figure 15, top). The spatial distribution of the error variance estimates is clearly related to ocean surface dynamics: it is found to be the highest in regions of high surface kinetic energy such as western boundary currents and equatorial regions²⁸, but is also relatively high at mid-latitudes within regions of high wind stress variability. Largest mean error variance estimates are found on average within the Agulhas Retroreflection region in the Indian Ocean and north of the Gulf Stream in the North Atlantic Ocean.

Whereas an error variance estimate provides a local quantification of the magnitude of the background noise, an uncertainty estimate for SST [eq. (24)] provides a statistical characterization of the distance between a SST estimate and the true, but unknown, SST value. In Figure 16, we examine the distributions of standard error estimates for the non-diurnal SST estimates, the diurnal SST estimates, and the total SST estimate for Level-2 data, and we report overall statistics rounded to the nearest 0.001 in Table 5. The results for Level-1 data are extremely similar and their distributions are not shown. Overall, the uncertainty estimates for diurnal SST estimates are a factor of 2 to 3 times larger than the uncertainty estimates for non-diurnal estimates. In turn, the uncertainty estimates for total SST estimates are larger than the uncertainty estimates for diurnal SST estimates but by no more than a few 1/1000-th of a degree. The most probable value of the uncertainty estimate for non-diurnal SST estimate is 0.006°C for all data. For drogued drifters only it is 0.005°C, and for undrogued drifters only it is 0.006°C. The most probable value of the uncertainty estimate for diurnal SST estimate is 0.016°C for all data. For drogued drifters only it is 0.015°C, and for undrogued drifters only it is 0.017°C. The most probable value of the uncertainty estimate for total SST estimate is 0.018°C for all data. For drogued drifters only it is slightly smaller at 0.016°C, and slightly higher for undrogued drifters at 0.019°C. The spatial distribution of the uncertainty estimates follow closely the spatial distribution of the error variance estimates (Figure 15, middle and bottom). The spatial distribution of the uncertainty estimates for total SST estimates is not shown as it is extremely

similar to the spatial distribution of the uncertainty estimates for diurnal SST estimates.

The overall statistics of uncertainty estimates for SST estimates (Table 5) are an order of magnitude smaller than previously estimated measurement uncertainties for drifting buoys⁸. Such uncertainty estimates range between 0.1 °C and 0.7 °C and are typically based on analyses of collocated SST observations from drifting buoys, ships, and satellites^{29–32}. These uncertainty estimates encompass not only the instrumental error of the drifter SST sensors but also the spatial and temporal differences between the different measurands that are targeted by the different observational platforms, such as a SST satellite’s ground footprint versus a pointwise drifter measurement. With this current SST product, the quantified uncertainties are on the order of 1/100-th of a Kelvin, rather than on the order of a 1/10-th of a Kelvin, because these estimated uncertainties are benefiting from time series of observations, as opposed to a single observation, that typically provide approximately 22 effective degrees of freedom over a 2-day observational estimation window (see mode value of v in Figure 7). As a result, sources of error arising from instrumental and communication noise, as well as sub-hourly unresolved geophysical variability, are averaged downward for our estimates. Our estimates are consistent with the back-of-the-envelope calculation of dividing by $\sqrt{22}$ an a priori error due to communication and transmission, and pre-deployment calibrated accuracy, overall on the order of 0.05 °C to 0.1 °C³³.

Data Records

The Level-2 estimates of total SST, non-diurnal SST, diurnal SST, and each of their respective standard error estimates, along with quality flag variables for each of the three SST estimates, are distributed as part of the hourly drifter dataset of the GDP⁵, now in its version 2.0 with the addition of these SST estimates. The dataset, assembled as a contiguous ragged array in a single file, is officially available from the NOAA National Center for Environmental Information (NCEI) at [doi and url pending]. The data are also available via the ERDDAP server of the NOAA Observing System Monitoring Center at http://osmc.noaa.gov/erddap/tabledap/gdp_hourly_velocities.html where subsets of the data can be selected according to a number of temporal and spatial criteria [update to version 2.00 pending].

Table 6 lists the names of the variables included in the NetCDF files, including the new SST-related variables. Usage of this SST data product in combination with any of the position and/or velocity data⁵ for release 2.00 or subsequent releases must cite this present paper as well as the original 2016 paper describing the hourly position and velocity dataset (Elipot et al. 2016⁵).

Technical Validation

The spatial and temporal distributions of the Level-2 hourly SST estimates are displayed in Figure 17. The map of spatial data density is the result of historical deployments and the efforts of the GDP to fulfill the requirement of the array, and of the patterns of the convergence and divergence of the near-surface oceanic circulation^{2,10}. The temporal histogram of SST estimates closely follows the distribution of hourly position and velocity estimates, showing the maturity of the array at the beginning of 2006 as well as the drop in the amount of data between 2011 and 2014 because of unfortunately numerous short-lived instruments. To support the technical validation of the new SST dataset, we compute the mean and standard deviation of SST estimates globally within $0.5^\circ \times 0.5^\circ$ geographical bins (Figure 18). The mean total SST map exhibits the expected meridional gradients as well as the west-east asymmetries within each ocean basins. As also expected, the standard deviation map of total SST estimates exhibits larger values within regions of higher surface kinetic energy such as in western boundary current regions³⁴ but also within the mid-latitude regions where high variability of air-sea fluxes is expected to enhance SST variance. The map of diurnal SST standard deviation exhibits different patterns resulting from the competing effects of the spatial pattern of solar heating increasing diurnal variability, and the spatial pattern of wind speed decreasing diurnal variability. At the scales displayed here, the maps of mean and standard deviation of non-diurnal SST estimates (not shown) are indistinguishable from the maps for the total SST estimates. The map of mean diurnal SST estimates (not shown) is approximately zero everywhere as expected from the model of temporal SST evolution used to derive this product.

Usage Notes

In the NetCDF files, all SST estimates are provided to three decimal places, with the last digit rounded towards the nearest 0.001. Total SST estimates are the sum of non-diurnal SST estimates and diurnal SST estimates but because of rounding, discrepancies exist within the NetCDF files for about 44% of values between the numerical value the user will read for the total SST value and the value of the sum of the non-diurnal and diurnal SST values.

The uncertainty estimates are also provided to three decimal places but with the last digit rounded “up” (i.e. towards infinity) to the nearest 0.001. The reason for the rounding up of the uncertainty estimates is to prevent reporting null uncertainties for 13,502 non-diurnal SST estimates for which the calculated uncertainty is smaller than 0.001 but larger than 0.0001. Rounding up uncertainties is acceptable as this provides more conservative uncertainties but only increasing their values by typically 6% for the non-diurnal SST uncertainties, and by typically 2% for the diurnal and total SST uncertainties.

Code availability

A software associated with this manuscript is licensed under MIT and published on GitHub at <https://github.com/selipot/sst-drift.git> (Elipot et al. 2021, [zenodo artifact]). This software allows the user to fit model (2) to temperature observations and derive the resulting SST estimates and their uncertainties. Input arguments to the model fitting function include an arbitrary order for the background non-diurnal SST model and arbitrary frequencies for the diurnal oscillatory model. A sample of Level-0 data from drifter AOML ID 55366 is provided in order to test the routines and produce figures similar to Figures 4 and 5. Alternatively, the main code can also generate stochastic data for testing purposes.

References

1. Centurioni, L. *et al.* Global in situ Observations of Essential Climate and Ocean Variables at the Air–Sea Interface. *Front. Mar. Sci.* **6**, 1–23, [10.3389/fmars.2019.00419](https://doi.org/10.3389/fmars.2019.00419) (2019).
2. Lumpkin, R., Centurioni, L. & Perez, R. C. Fulfilling Observing System Implementation Requirements with the Global Drifter Array. *J. Atmos. Ocean. Technol.* **33**, 685–695, [10.1175/JTECH-D-15-0255.1](https://doi.org/10.1175/JTECH-D-15-0255.1) (2016).
3. Hansen, D. V. & Poulain, P.-M. Quality Control and Interpolations of WOCE-TOGA Drifter Data. *J. Atmos. Ocean. Technol.* **13**, 900–909, [10.1175/1520-0426\(1996\)013<0900:QCAIOW>2.0.CO;2](https://doi.org/10.1175/1520-0426(1996)013<0900:QCAIOW>2.0.CO;2) (1996).
4. Lumpkin, R. & Centurioni, L. NOAA Global Drifter Program quality-controlled 6-hour interpolated data from ocean surface drifting buoys. NOAA National Centers for Environmental Information. Dataset., <https://doi.org/10.25921/7ntx-z961> (2019). Accessed November 29, 2021.
5. Elipot, S. *et al.* A global surface drifter data set at hourly resolution. *J. Geophys. Res. Ocean.* **121**, 2937–2966, [10.1002/2016JC011716](https://doi.org/10.1002/2016JC011716) (2016).
6. Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **74**, 829, [10.2307/2286407](https://doi.org/10.2307/2286407) (1979).
7. Chiorboli, G. Uncertainty of mean value and variance obtained from quantized data. *IEEE Trans. Instrum. Meas.* **52**, 1273–1278, [10.1109/TIM.2003.816820](https://doi.org/10.1109/TIM.2003.816820) (2003).
8. Kennedy, J. J. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **52**, 1–32, [10.1002/2013RG000434](https://doi.org/10.1002/2013RG000434) (2014).
9. Lumpkin, R., Özgökmen, T. & Centurioni, L. Advances in the Application of Surface Drifters. *Ann. Rev. Mar. Sci.* **9**, 59–81, [10.1146/annurev-marine-010816-060641](https://doi.org/10.1146/annurev-marine-010816-060641) (2017).
10. Lumpkin, R., Maximenko, N. & Pazos, M. Evaluating Where and Why Drifters Die*. *J. Atmos. Ocean. Technol.* **29**, 300–308, [10.1175/JTECH-D-11-00100.1](https://doi.org/10.1175/JTECH-D-11-00100.1) (2012).
11. Centurioni, L., Horányi, A., Cardinali, C., Charpentier, E. & Lumpkin, R. A global ocean observing system for measuring sea level atmospheric pressure : Effects and impacts on numerical weather prediction. *Bull. Am. Meteorol. Soc.* **98**, 231–238, [10.1175/BAMS-D-15-00080.1](https://doi.org/10.1175/BAMS-D-15-00080.1) (2017).
12. Centurioni, L. R. *et al.* Sea surface salinity observations with lagrangian drifters in the tropical North Atlantic during SPURS: Circulation, fluxes, and comparisons with remotely sensed salinity from aquarius. *Oceanography* **28**, 96–105, [10.5670/oceanog.2015.08](https://doi.org/10.5670/oceanog.2015.08) (2015).
13. Hansen, D. V. & Herman, A. Temporal Sampling Requirements for Surface Drifting Buoys in the Tropical Pacific. *J. Atmos. Ocean. Technol.* **6**, 599–607, [10.1175/1520-0426\(1989\)006<0599:TSRFS>2.0.CO;2](https://doi.org/10.1175/1520-0426(1989)006<0599:TSRFS>2.0.CO;2) (1989).
14. Elipot, S. & Lumpkin, R. Spectral description of oceanic near-surface variability. *Geophys. Res. Lett.* **35**, L05606, [10.1029/2007GL032874](https://doi.org/10.1029/2007GL032874) (2008).
15. *Argos User's Manual* (2015). [Http://www.argos-system.org/manual/](http://www.argos-system.org/manual/), Copyright 2007-2015 CLS.
16. Lopez, R., Malarde, J.-P., Royer, F. & Gaspar, P. Improving Argos Doppler Location Using Multiple-Model Kalman Filtering. *IEEE Trans. Geosci. Remote. Sens.* **52**, 4744–4755, [10.1109/TGRS.2013.2284293](https://doi.org/10.1109/TGRS.2013.2284293) (2014).
17. Lumpkin, R. *et al.* Removing Spurious Low-Frequency Variability in Drifter Velocities. *J. Atmos. Ocean. Technol.* **30**, 353–360, [10.1175/JTECH-D-12-00139.1](https://doi.org/10.1175/JTECH-D-12-00139.1) (2013).
18. Huang, B. *et al.* Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *J. Clim.* **34**, 2923–2939, [10.1175/JCLI-D-20-0166.1](https://doi.org/10.1175/JCLI-D-20-0166.1) (2021).
19. Fan, J. & Gijbels, I. *Local Polynomial Modelling and Its Applications* (Routledge, 2018), 1st edn.

20. Gentemann, C. L. Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.* **30**, 1140, [10.1029/2002GL016291](https://doi.org/10.1029/2002GL016291) (2003).
21. 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.* **34**, 1–5, [10.1029/2006GL028920](https://doi.org/10.1029/2006GL028920) (2007).
22. Lindfors, A. V., Mackenzie, I. A., Tett, S. F. B. & Shi, L. Climatological Diurnal Cycles in Clear-Sky Brightness Temperatures from the High-Resolution Infrared Radiation Sounder (HIRS). *J. Atmos. Ocean. Technol.* **28**, 1199–1205, [10.1175/JTECH-D-11-00093.1](https://doi.org/10.1175/JTECH-D-11-00093.1) (2011).
23. Morak-Bozzo, S., Merchant, C. J., Kent, E. C., Berry, D. I. & Carella, G. Climatological diurnal variability in sea surface temperature characterized from drifting buoy data. *Geosci. Data J.* **3**, 20–28, [10.1002/gdj3.35](https://doi.org/10.1002/gdj3.35) (2016).
24. Hormann, V., Centurioni, L. R. & Reverdin, G. Evaluation of Drifter Salinities in the Subtropical North Atlantic. *J. Atmos. Ocean. Technol.* **32**, 185–192, [10.1175/JTECH-D-14-00179.1](https://doi.org/10.1175/JTECH-D-14-00179.1) (2015).
25. Kawai, Y. & Wada, A. Diurnal sea surface temperature variation and its impact on the atmosphere and ocean: A review. *J. Ocean.* **63**, 721–744, [10.1007/s10872-007-0063-0](https://doi.org/10.1007/s10872-007-0063-0) (2007).
26. Merchant, C. J. *et al.* Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data* **6**, 223, [10.1038/s41597-019-0236-x](https://doi.org/10.1038/s41597-019-0236-x) (2019).
27. Flament, P., Firing, J., Sawyer, M. & Trefois, C. Amplitude and Horizontal Structure of a Large Diurnal Sea Surface Warming Event during the Coastal Ocean Dynamics Experiment. *J. Phys. Ocean.* **24**, 124–139, [10.1175/1520-0485\(1994\)024<0124:AAHSAO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1994)024<0124:AAHSAO>2.0.CO;2) (1994).
28. Laurindo, L. C., Mariano, A. J. & Lumpkin, R. An improved near-surface velocity climatology for the global ocean from drifter observations. *Deep. Sea Res. Part I Ocean. Res. Pap.* **124**, 73–92, [10.1016/j.dsr.2017.04.009](https://doi.org/10.1016/j.dsr.2017.04.009) (2017).
29. Emery, W. J., Baldwin, D. J., Schlüssel, P. & Reynolds, R. W. Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *J. Geophys. Res. Ocean.* **106**, 2387–2405, [10.1029/2000JC000246](https://doi.org/10.1029/2000JC000246) (2001).
30. O’Carroll, A. G., Eyre, J. R. & Saunders, R. W. Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations. *J. Atmos. Ocean. Technol.* **25**, 1197–1207, [10.1175/2007JTECHO542.1](https://doi.org/10.1175/2007JTECHO542.1) (2008).
31. Xu, F. & Ignatov, A. Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals. *J. Geophys. Res. Ocean.* **115**, 1–18, [10.1029/2010JC006129](https://doi.org/10.1029/2010JC006129) (2010).
32. Merchant, C. J. *et al.* A 20year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. *J. Geophys. Res. Ocean.* **117**, 1–18, [10.1029/2012JC008400](https://doi.org/10.1029/2012JC008400) (2012).
33. Poli, P. *et al.* The Copernicus Surface Velocity Platform drifter with Barometer and Reference Sensor for Temperature (SVP-BRST): genesis, design, and initial results. *Ocean. Sci.* **15**, 199–214, [10.5194/os-15-199-2019](https://doi.org/10.5194/os-15-199-2019) (2019).
34. Lumpkin, R. & Johnson, G. C. Global ocean surface velocities from drifters: Mean, variance, El Niño–Southern Oscillation response, and seasonal cycle. *J. Geophys. Res. Ocean.* **118**, 2992–3006, [10.1002/jgrc.20210](https://doi.org/10.1002/jgrc.20210) (2013).

Acknowledgements

This research was partially carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a Cooperative Institute of the University of Miami and the National Oceanic and Atmospheric Administration, cooperative agreement #NA20OAR4320472. R. Lumpkin was supported by NOAA’s Global Ocean and Monitoring Program and the Atlantic Oceanographic and Meteorological Laboratory. L. Centurioni was supported by NOAA’s grant NA20OAR4320278 “The Global Drifter program”. The authors thank Sofia Olhede for her advice on some of the statistical aspects of this work, and Bertrand Dano and Philippe Miron for generating the final NetCDF files for distribution.

Author contributions statement

S. Elipot, A. Sykulski, and R. Lumpkin conceived the dataset and designed the associated methods. M. Pazos assembled and conducted the quality control of the drifter data forming the basis for the Level-0 data. All authors reviewed the manuscript.

Competing interests

(mandatory statement)

All authors declare no conflict of interest.

Figures & Tables

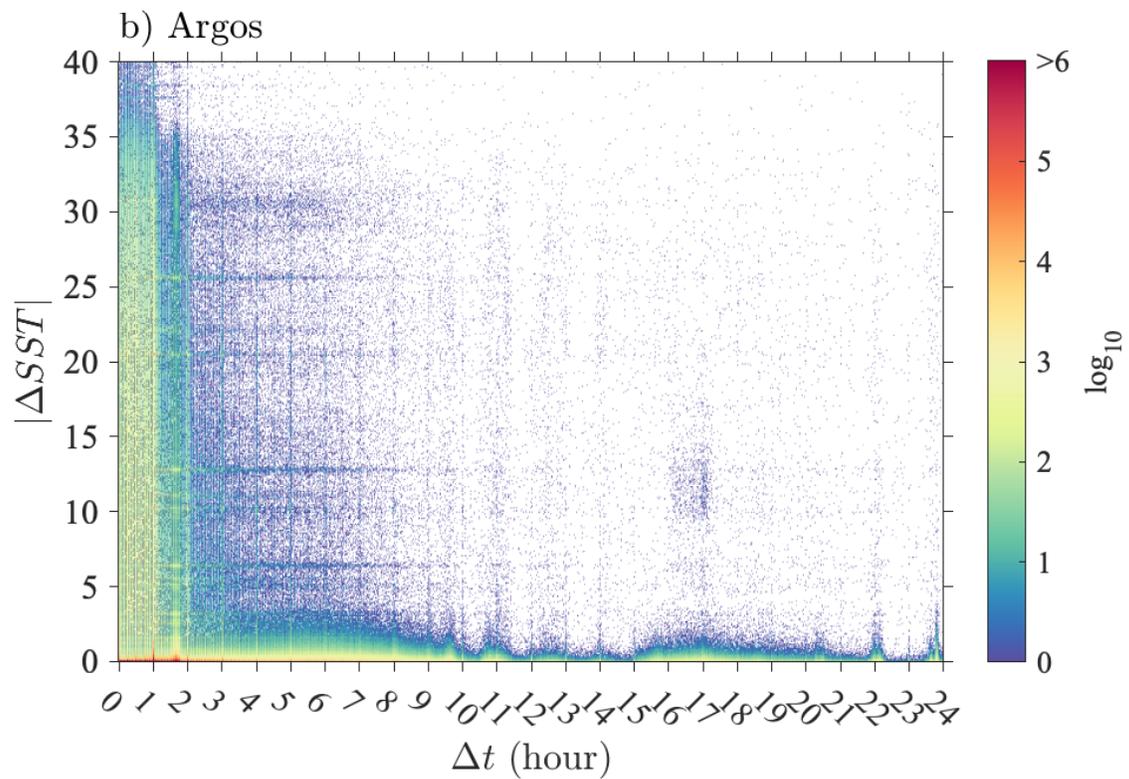
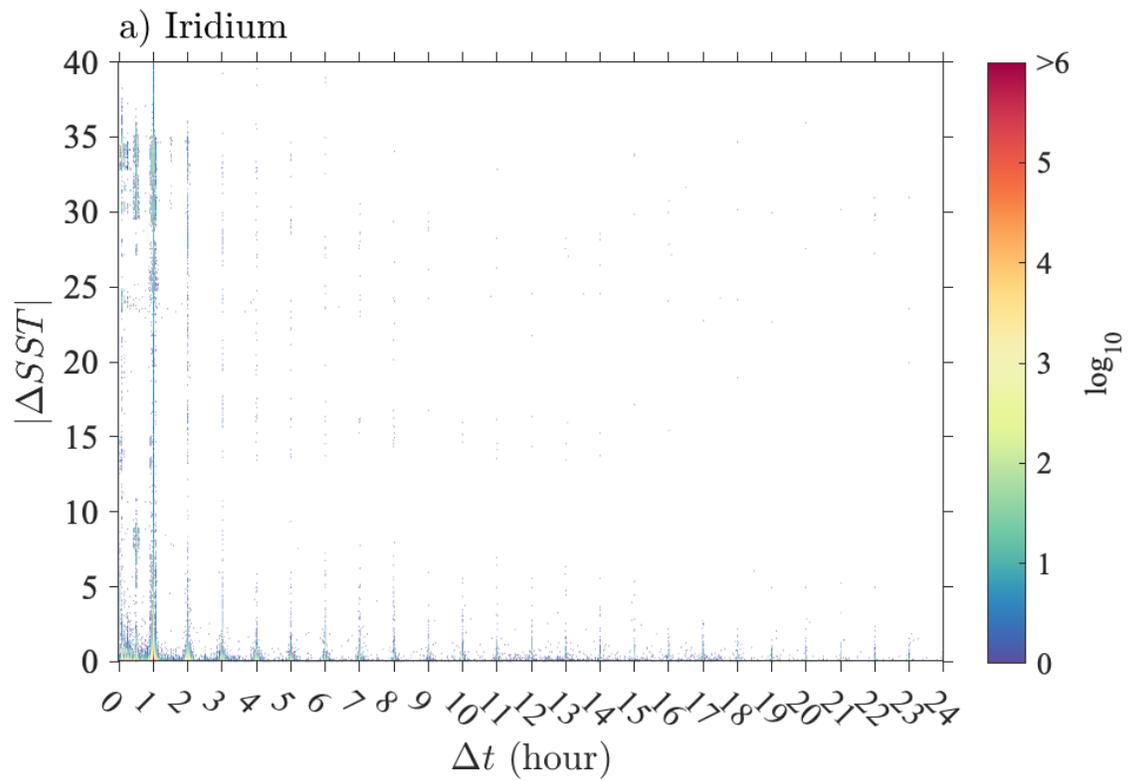


Figure 1. Two-dimensional histograms of time differences (Δt) and absolute temperature differences ($|\Delta SST|$) between two consecutive SST observations of the Level-0 data, for Iridium drifters (a, top panel) and Argos drifters (b, bottom panel). ~~10.37~~ values of Δt less than 24 hours and 40°C are shown.

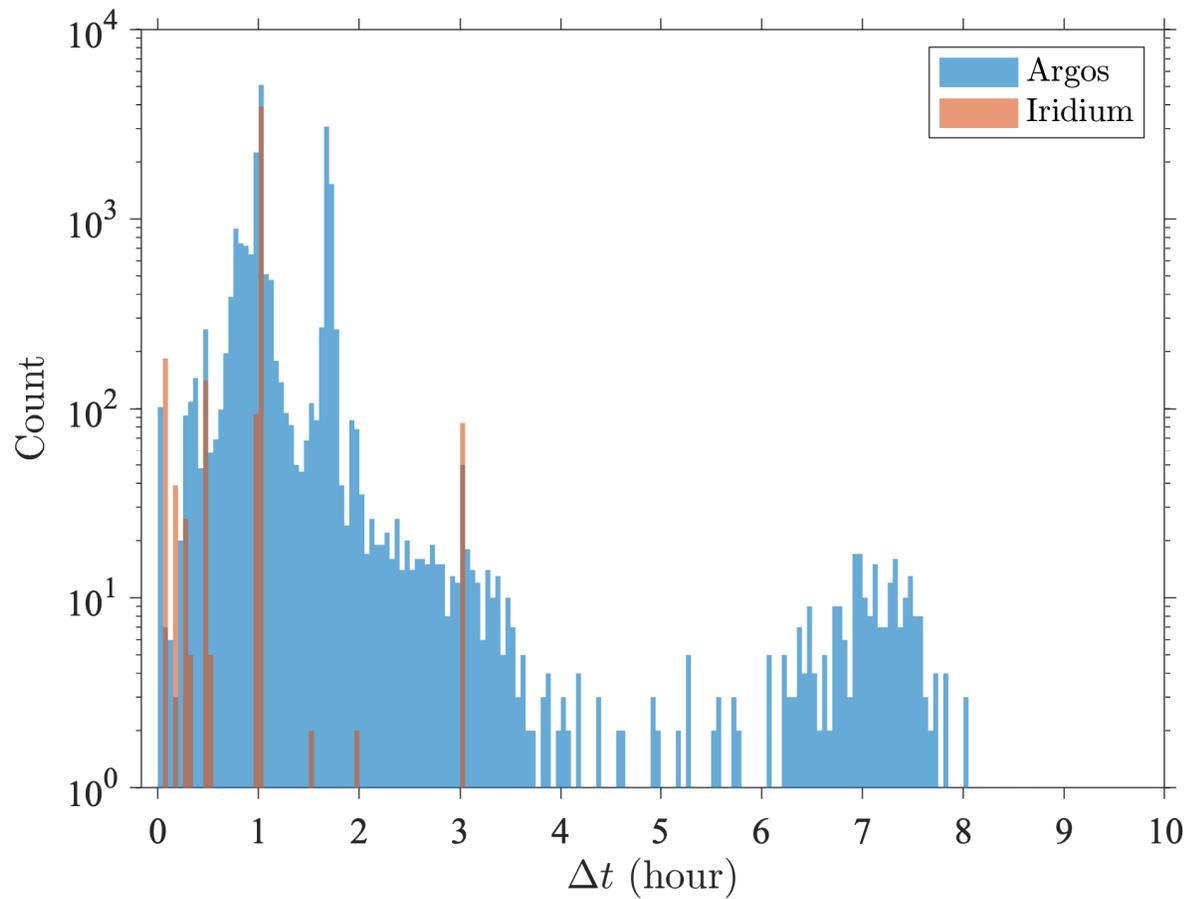


Figure 2. Histograms in 3-minute bins of median SST temporal sampling intervals per drifter trajectory for Level-0 data (24,597 SST time series from 4,495 Iridium drifters and 20,102 Argos drifters). Only values smaller than 10 hours are displayed. 105 Argos time series have a median sampling larger than 10 hours.

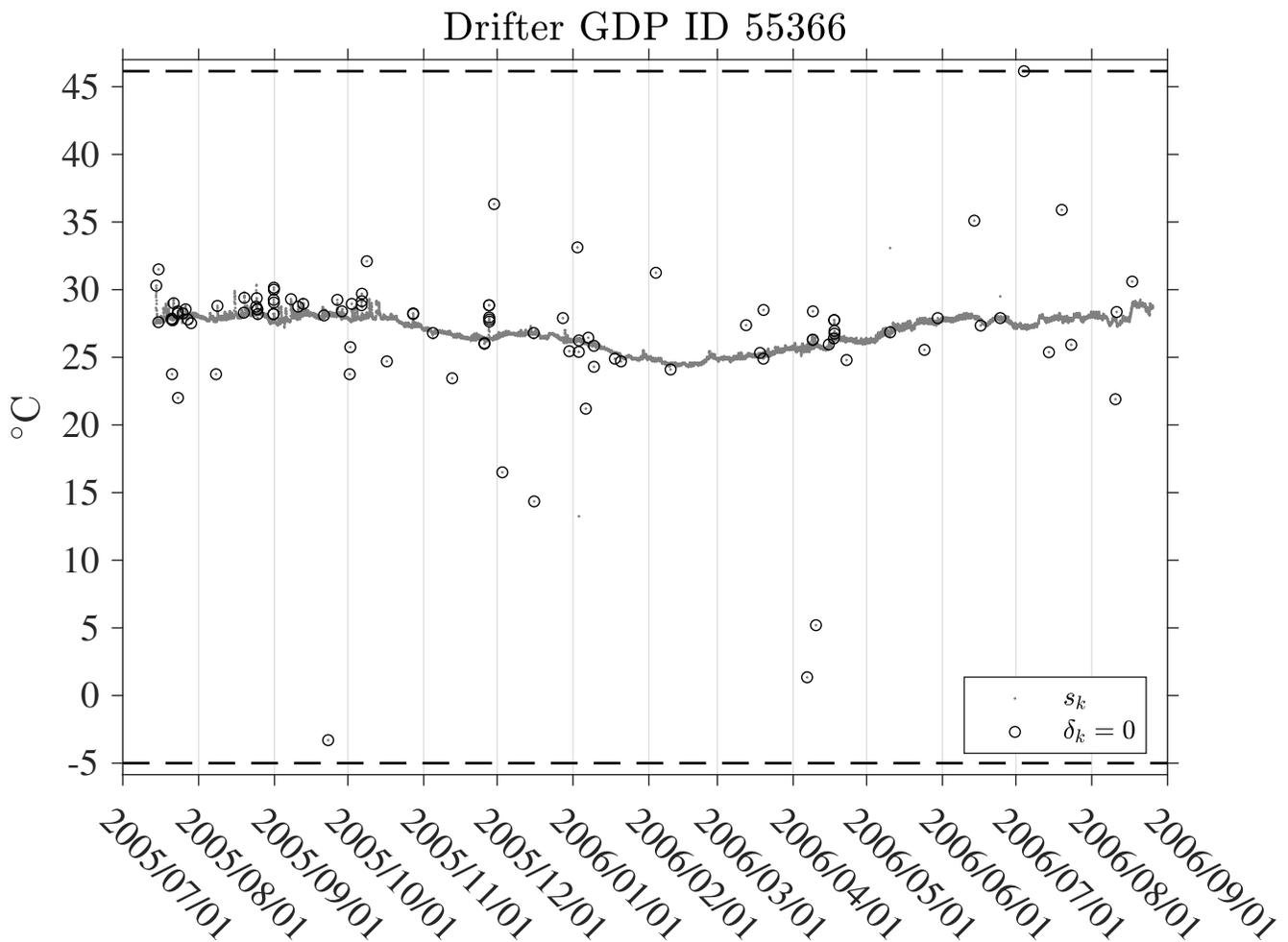


Figure 3. Time series of SST data for GDP drifter ID 55366 (WMO number 3100541). This drifter was built by Pacific Gyre and is of the Surface velocity Program (SVP) type, tracked by the Argos positioning system. The median time interval between SST observations for this drifter is about 52 min. The SST equation for this drifter is $SST(^{\circ}C) = 0.05 \times n - 5.00$ where n is a 10-bit sensor count. This equation defines the data resolution ($0.05^{\circ}C$) as well as the minimum value ($0.05 \times 0 - 5.00 = -5.00^{\circ}C$) and maximum value ($0.05 \times (2^{10} - 1) - 5.00 = 46.15^{\circ}C$) that should be returned by the temperature sensor, indicated by the horizontal dashed lines on this figure. The Level-0 data (s_k) are indicated by gray dots. The circled dots are the data points that are ultimately down-weighted to zero by the iterative estimation method, and thus flagged as outliers ($\delta_k = 0$).

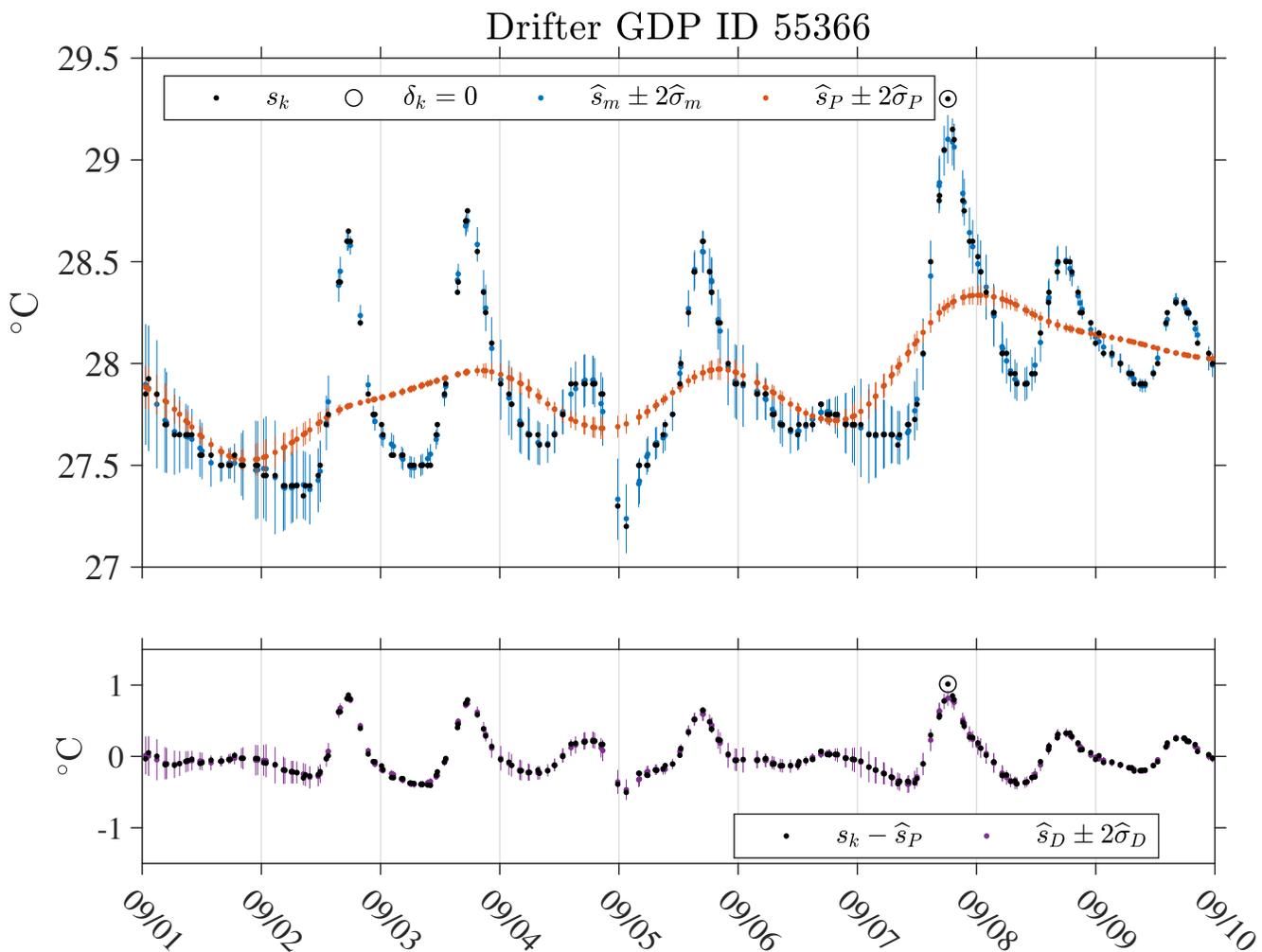


Figure 4. Time series of SST estimates for GDP drifter ID 55366 (WMO number 3100541) between 2005/9/1 and 2005/9/10. Top panel: Black dots are the original SST data (level-0, s_k) and circles are the data points down-weighted to zero ($\delta_k = 0$). The blue dots with vertical lines show the total SST estimates and their plus or minus two standard errors ($\hat{s}_m \pm 2\hat{\sigma}_m$). The red dots and vertical lines show the non-diurnal SST estimates and their plus or minus two standard errors ($\hat{s}_P \pm 2\hat{\sigma}_P$). The blue estimates are the sum of the red estimates and purple estimates shown in the lower panel of the figure. Lower panel: Black dots show SST data minus the non-diurnal SST estimate ($s_k - \hat{s}_P$). The purple dots and vertical lines show the corresponding diurnal SST estimates and their plus or minus two standard errors ($\hat{s}_D \pm 2\hat{\sigma}_D$).

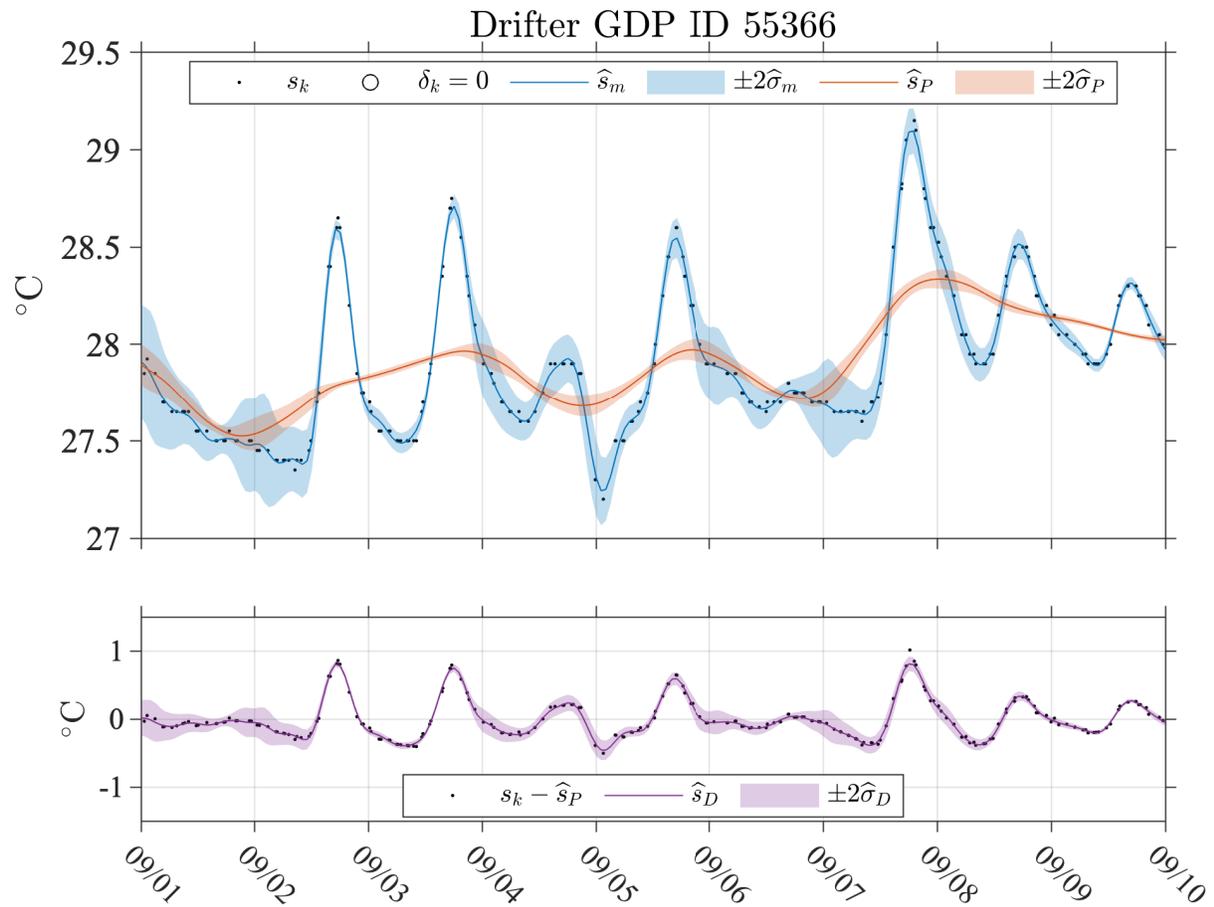


Figure 5. Time series of continuous hourly SST estimates for GDP drifter ID 55366 (WMO number 3100541) between 2005/9/1 and 2005/9/10. Top panel: Black dots show the original SST data (level-0, s_k). The blue line and shaded region show continuously the hourly total SST estimates and twice their standard errors ($\hat{s}_m \pm 2\hat{\sigma}_m$). The red line and shaded region show continuously the hourly non-diurnal SST estimates and twice their standard errors ($\hat{s}_m \pm 2\hat{\sigma}_P$). The blue line is the sum of the red line and the purple line shown in the lower panel. Lower panel: Black dots show SST data minus the non-diurnal SST estimate ($s_k - \hat{s}_P$). The purple line and shaded region show continuously the hourly diurnal SST estimates and twice their standard errors ($\hat{s}_D \pm 2\hat{\sigma}_D$).

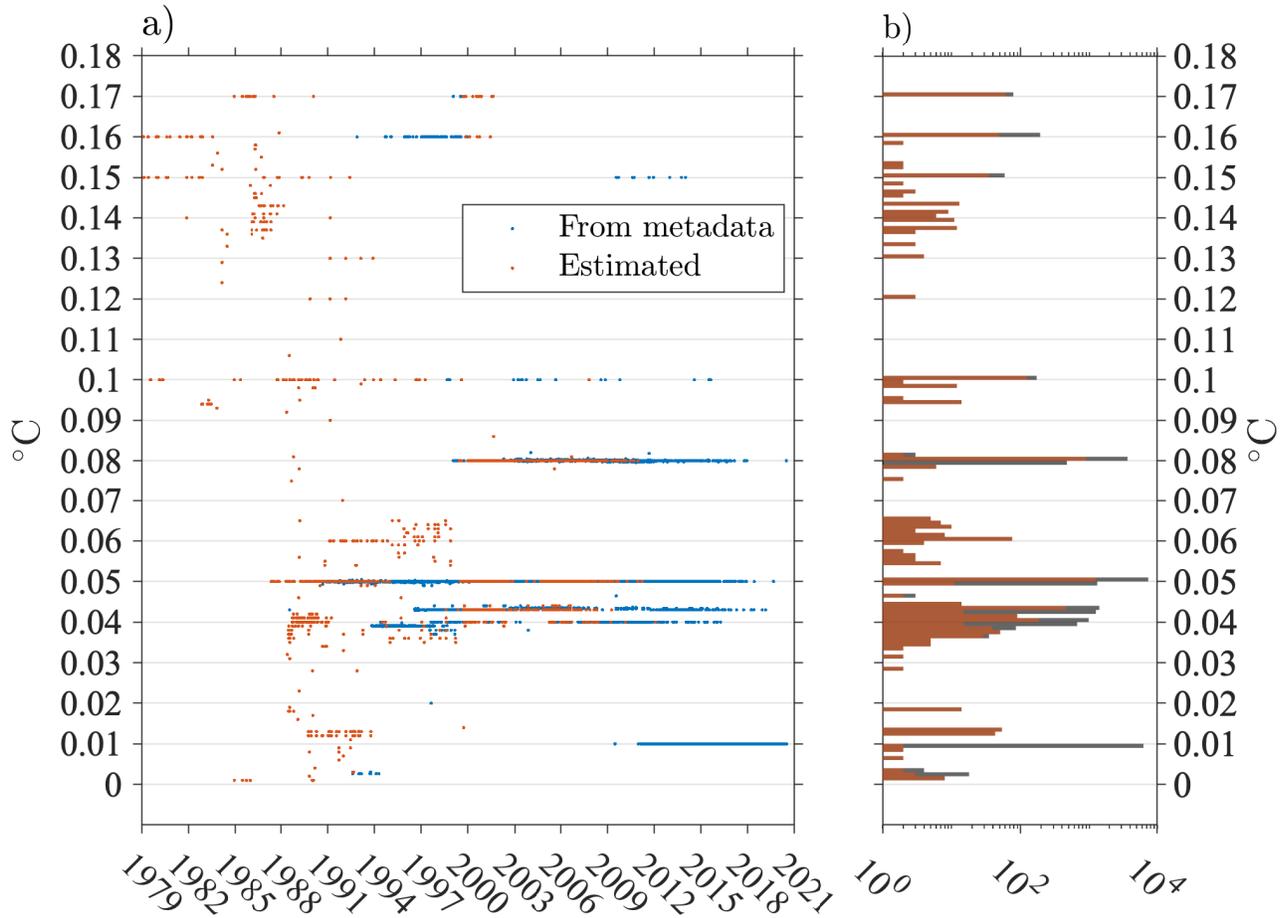


Figure 6. Distributions of drifter SST sensors resolution. a) Temporal distribution of drifter SST resolution in the GDP database from February 1979 to July 2020. The blue points corresponds to resolution a obtained from the drifter metadata from the SST equation: $SST(^{\circ}C) = a \times n + b$ where n is a bit sensor count. The red points corresponds to drifters for which the resolution is not available from the metadata and was estimated directly from the data (see text). b) Histogram of drifter SST resolution values in $0.001^{\circ}C$ bins. The red bars correspond to the estimated resolution values and the gray bars correspond to all values. Note that the horizontal axis is on a log scale. The three most common resolution values in the dataset are in order: 0.05, 0.01, and $0.08^{\circ}C$.

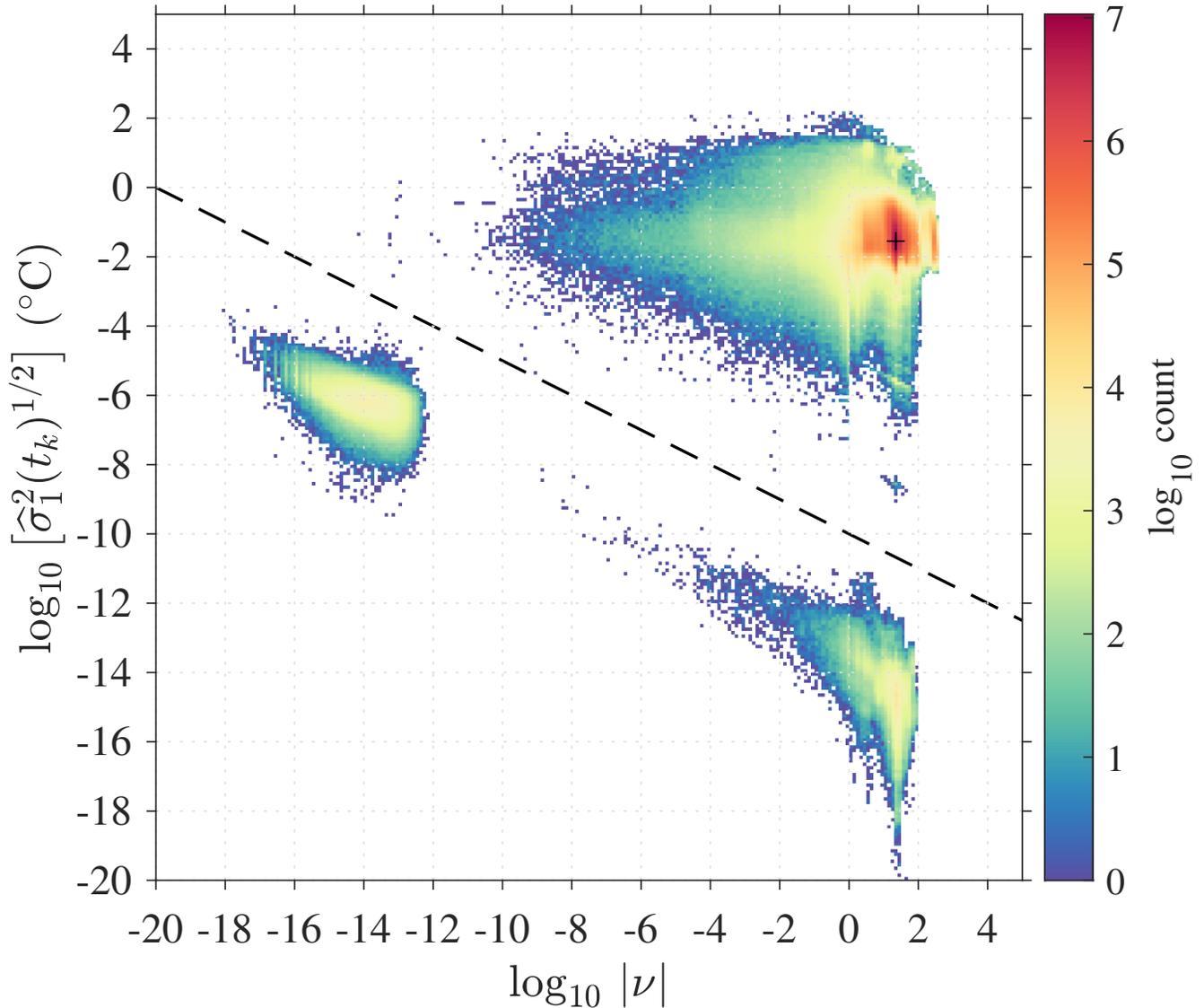


Figure 7. Two-dimensional histogram of the effective degrees of freedom for the residuals (ν) and estimates of the error variance neglecting the resolution error variance $[\hat{\sigma}_1^2(t_k)]$ for Level-1 data results. The two populations found below the black dashed line ($\log_{10}[\hat{\sigma}_1^2]^{1/2} < -\frac{1}{2}\log_{10}|\nu| - 10$) correspond to failed estimations of the error variance and are flagged with quality flag 1. The left population below the dashed line (0.33% of the data) corresponds to negative estimated variance (see text). The right population below the dashed line (0.22% of the data) corresponds to uncharacteristically flat SST records leading to unrealistic near-zero estimated error variances. The peak of the distribution is found within the upper-right population for $\nu \approx 22.4$ and $[\hat{\sigma}_1^2(t_k)]^{1/2} \approx 0.028^\circ\text{C}$.

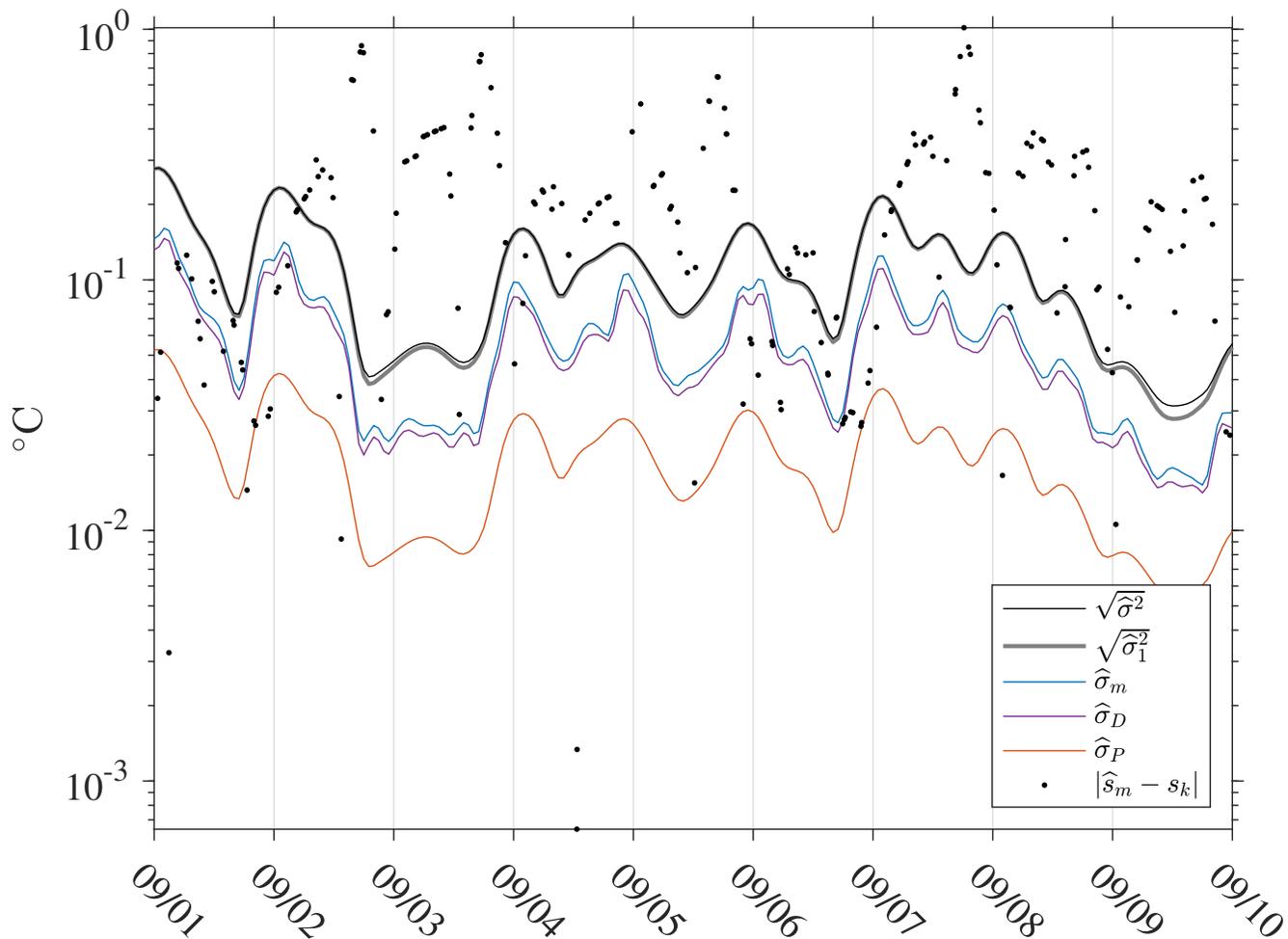


Figure 8. Time series for GDP drifter ID 55366 (WMO ID 3100541) between 2005/9/1 and 2005/9/10 of SST standard error estimates ($\hat{\sigma}_m$), non-diurnal SST standard error estimates ($\hat{\sigma}_P$), diurnal SST standard error estimates ($\hat{\sigma}_D$), square root of error variance estimates from residuals [$\sqrt{\hat{\sigma}_1^2}$, eq. (25)], square root of total error variance estimates [$\sqrt{\hat{\sigma}^2}$, eq. (28)], and absolute residuals ($|\hat{s}_m - s_k|$). The curves for $\sqrt{\hat{\sigma}_1^2}$ and $\sqrt{\hat{\sigma}^2}$ are most often indistinguishable except around 09/03 and 09/09-09/10.

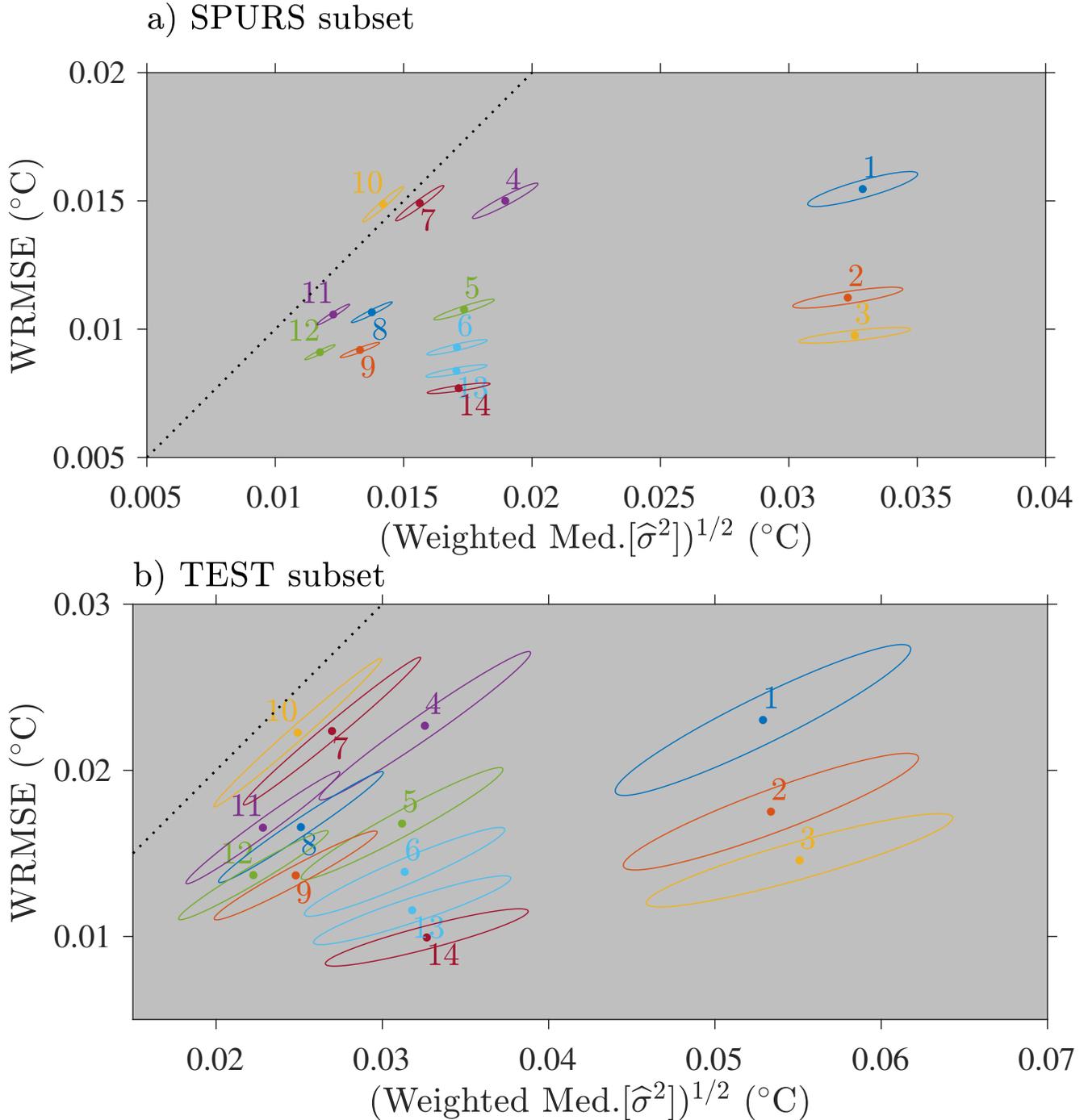


Figure 9. Summary statistics for the 14 models listed in Table 2 for (a) the subset of 80 drifters from the SPURS experiment and (b) the subset of 98 “test” drifters selected from the global database. Colored dots with numbers indicate the average values of the square root of the weighted median of the error variance (horizontal axis) versus the average values of the weighted root mean square error (WRMSE, vertical axis). Ellipses correspond to 95% confidence intervals across ensemble statistics. Note the different axis ranges between panels a) and b). The black dotted line indicates the slope-1 intercept-0 curve.

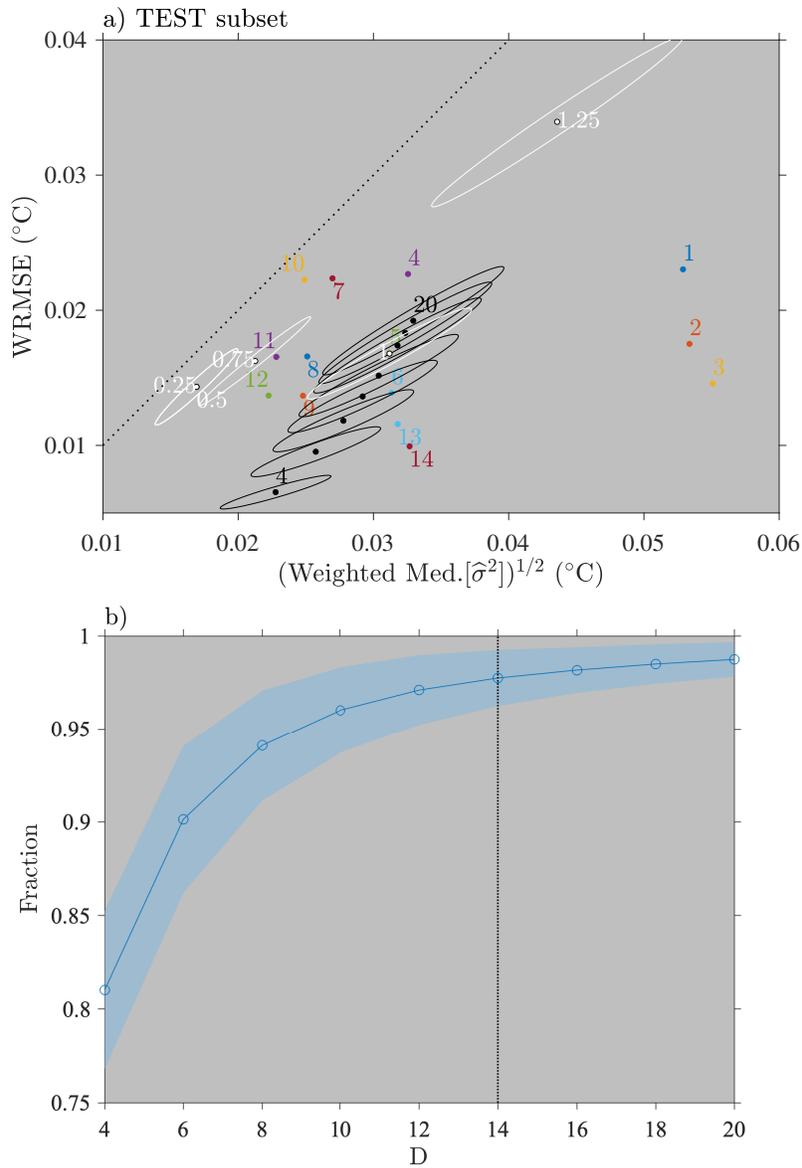


Figure 10. Panel a): Summary statistics for the 14 models for the test set of drifters as in Figure 9 (b). Here are also shown the results of varying the D factor in eq. (22) from 4 to 20 in increments of 2 for model 5 (black dots and ellipses) and the results of varying the bandwidth parameter h_k from 0.25 days to 1.25 days in increments of 0.25 (white dots and ellipses). The white ellipse around the black dot for model 5 corresponds to $h_k = 1$ and $D = 14$ as in Figure 9. Panel b): Ensemble averages of the fraction of data points not labeled as outliers as a function of factor D . The shading indicates plus or minus one standard deviation around the ensemble averages. The vertical dotted line indicates $D = 14$ ultimately chosen here.

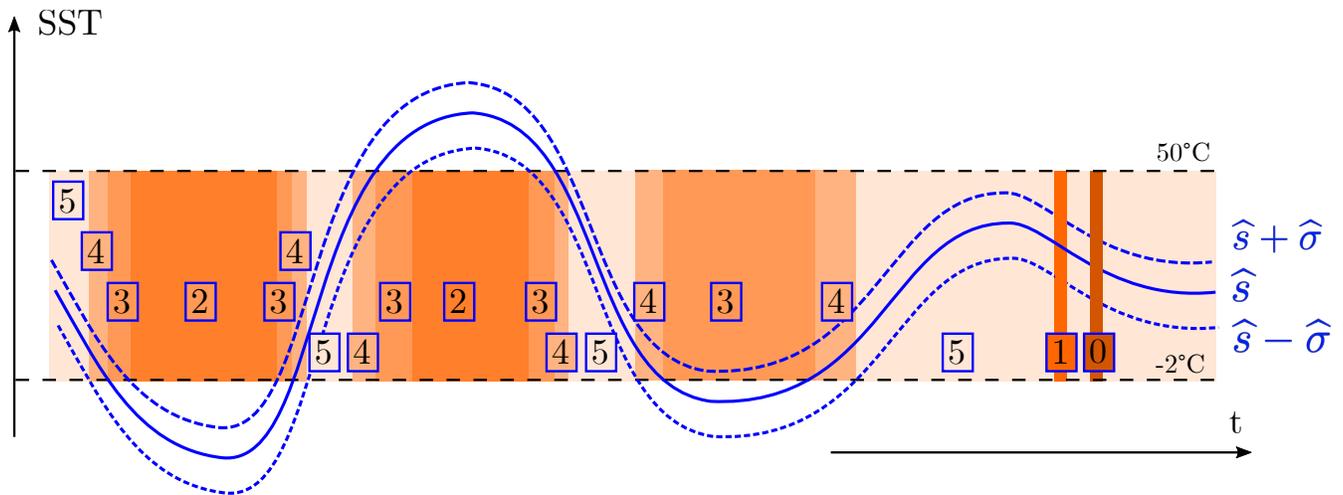


Figure 11. Illustration of the quality flag determination scheme for time series of non-diurnal and total SST estimates. Flag 0 indicates both missing SST estimates (\hat{s}) and uncertainty estimates ($\hat{\sigma}$). Flag 1 indicates a missing or failed uncertainty estimate only. Flags 2, 3, 4, and 5 are based on the values of \hat{s} and $\hat{s} \pm \hat{\sigma}$ with respect to the range of acceptable temperature values ($[-2, 50]^{\circ}\text{C}$).

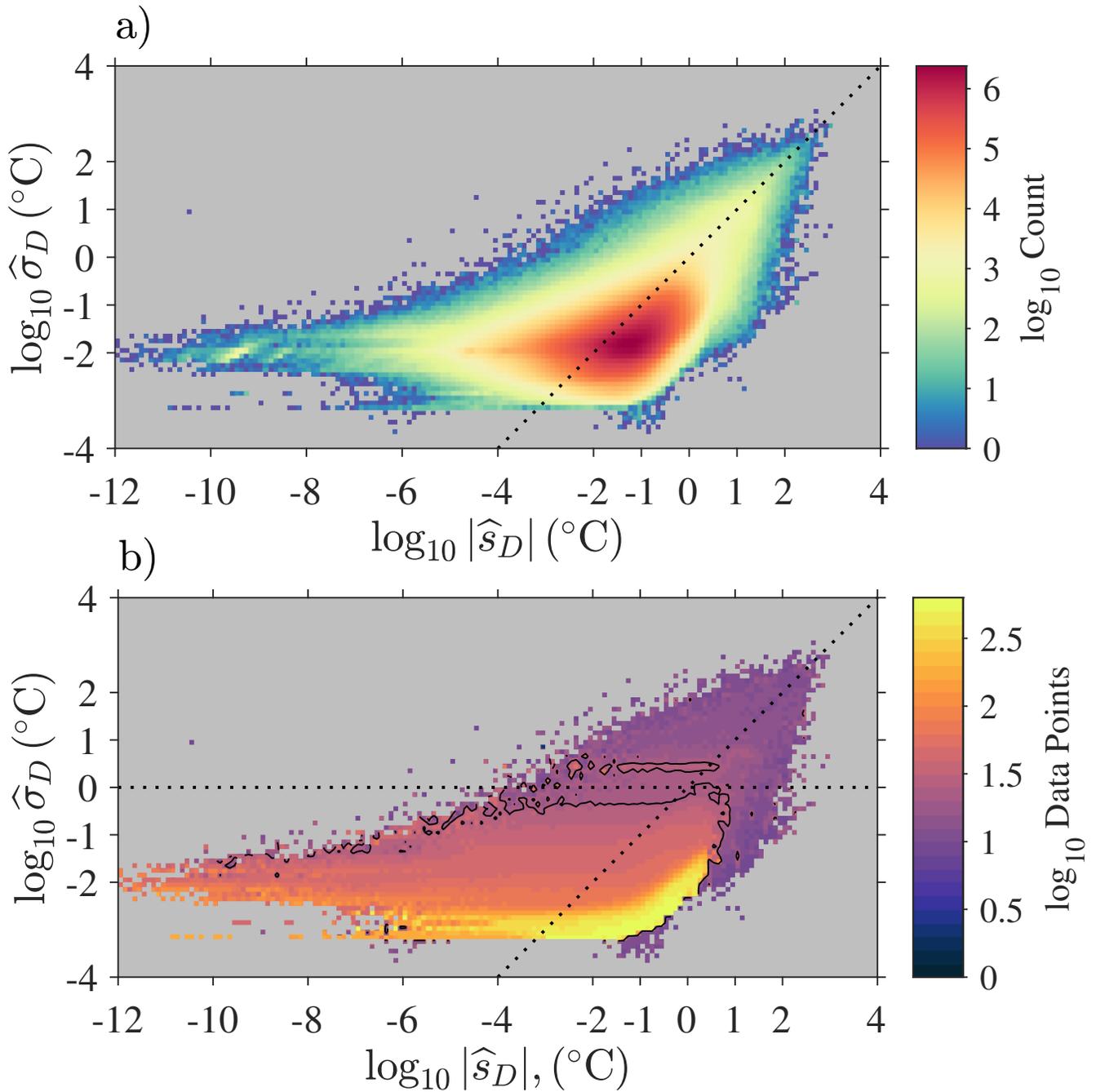


Figure 12. a) Two-dimensional histogram of absolute diurnal SST estimates (\hat{s}_D) and standard error estimates for diurnal estimates ($\hat{\sigma}_D$) for Level-2 data. The dotted line corresponds to the slope 1 line ($\hat{\sigma}_D = \hat{s}_D$). b) Average number of Level-0 data points used for estimating SST mapped onto the two-dimensional distribution shown in the top panel. The black contour corresponds to 24 data points on average.

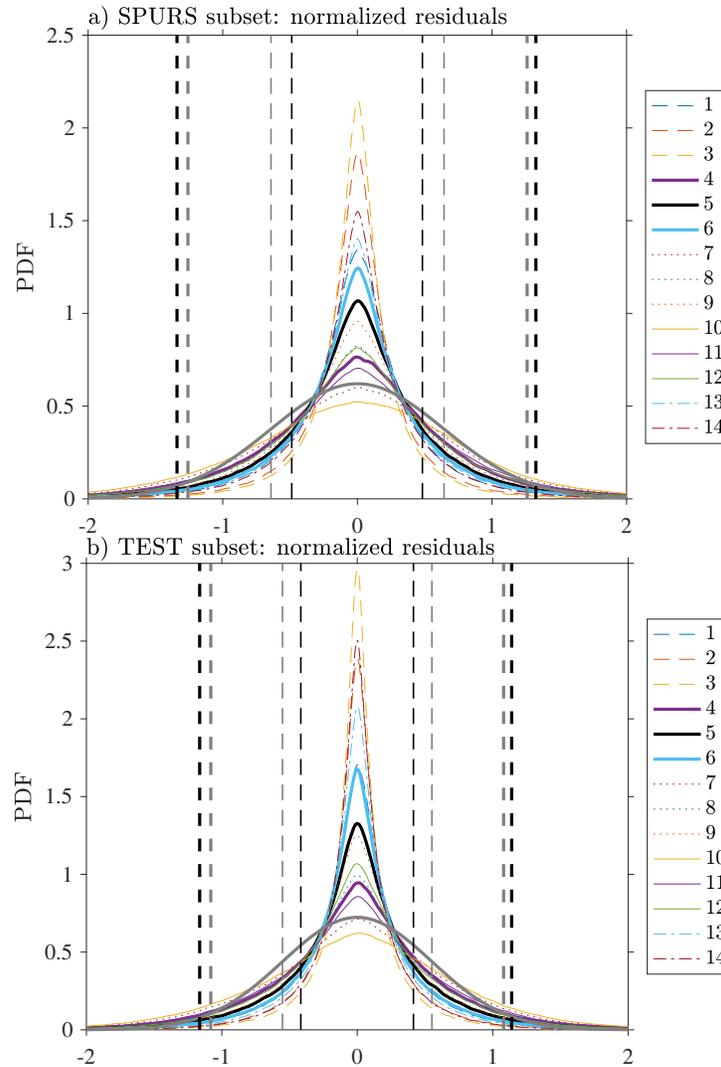


Figure 13. Probability density function (PDF) estimates of normalized residuals following the fitting of the 14 models for (a) the subset of 80 drifters from the SPURS experiment and (b) the subset of 98 “test” drifters selected from the global database. The PDFs are estimated using an Epanechnikov kernel¹⁹ at 0.01 resolution using only residuals with non-zero final robust weights. In each panel, the thin dashed black lines indicate the 16-th and 84-th percentiles of the distribution of residuals for model 5 whereas the thick dashed black lines indicate the 2.5-th and 97.5-th percentiles. The gray curve in each panel corresponds to the fit to a normal distribution for the residuals for model 5 and the gray dashed vertical lines indicate the mean plus or minus 1 (thin line) and 1.96 (thick line) standard deviation and therefore correspond to the 2.5-, 16-, 84-, and 97.5-th percentiles of that fitted normal distribution.

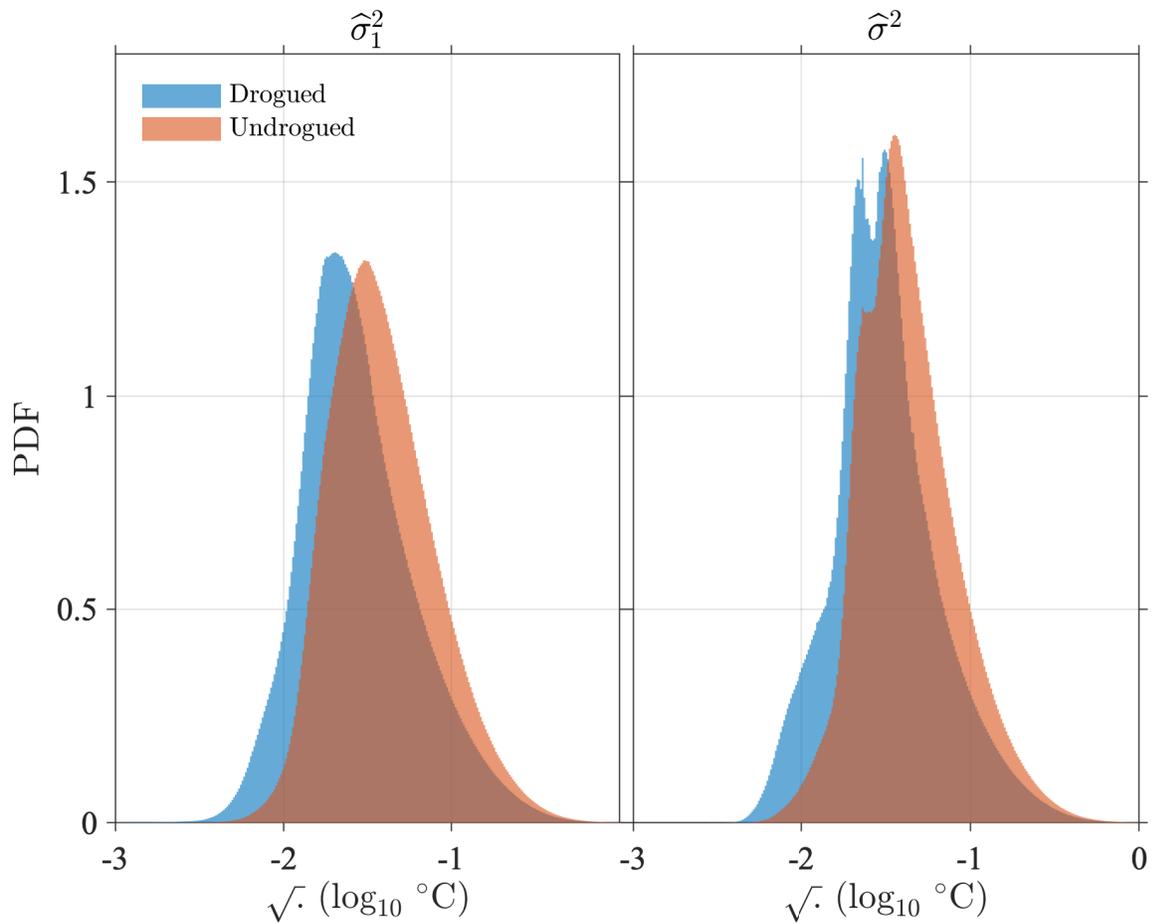


Figure 14. Distribution of error variance estimates from residuals [left, $\hat{\sigma}_1^2$, eq. (25)] and total error variance estimates incorporating the resolution error [right, $\hat{\sigma}^2$, eq. (28)] for Level-2 data. The histograms of the decimal logarithm of the square root of the estimates are displayed. Mode values at the peak of the distributions and 50-th percentile values are listed in Table 4.

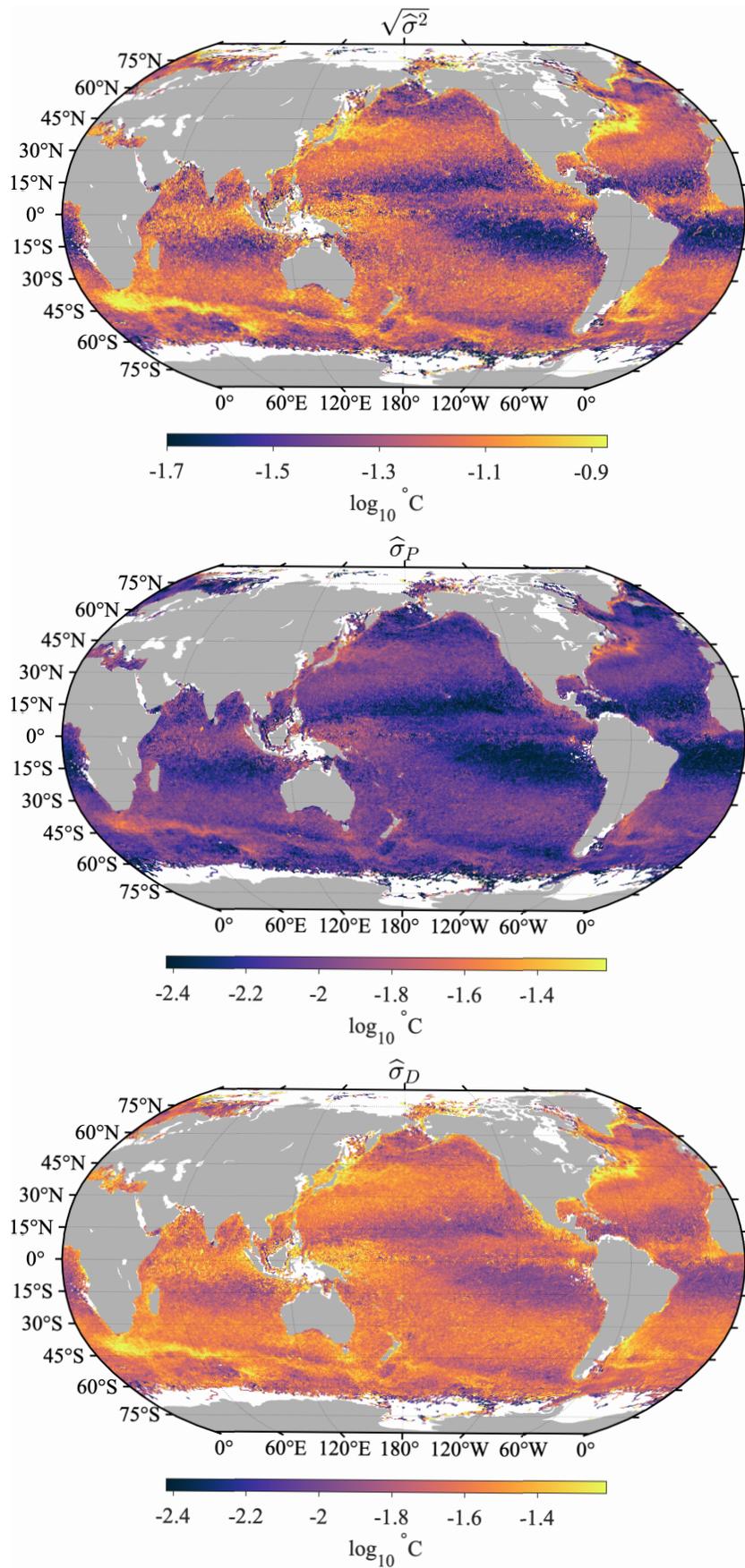


Figure 15. Top: Square root of total error variance estimates $[(\hat{\sigma}^2)^{1/2}]$ averaged in half-degree spatial bins. Middle: Non-diurnal SST uncertainty estimates $(\hat{\sigma}_P)$ averaged in half-degree spatial bins. Bottom: total SST uncertainty estimates $(\hat{\sigma}_D)$ averaged in half-degree spatial bins. The maps are obtained with Level-2 data with quality flags 5 for all estimates. In all three panels the units are decimal logarithm of degrees Celsius.

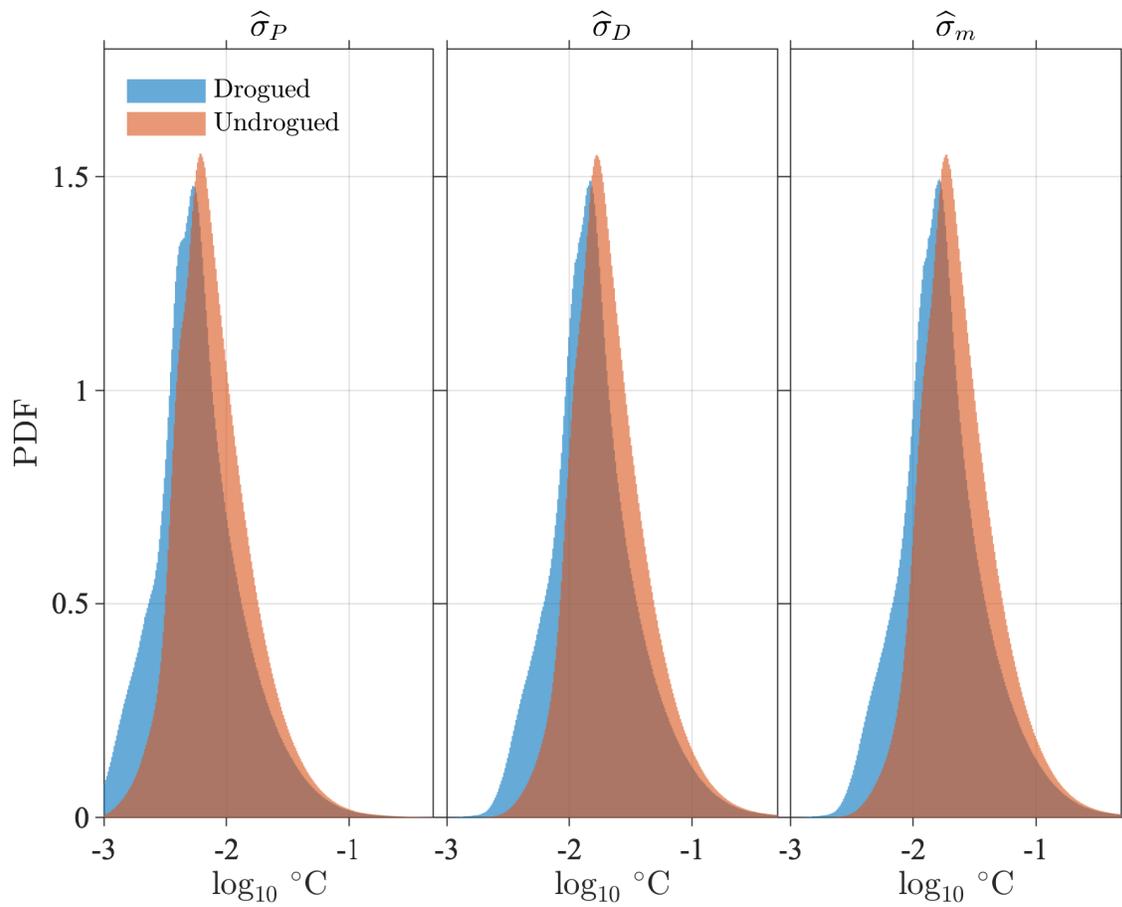


Figure 16. Probability density function (PDF) estimates of standard error estimates for the non-diurnal ($\hat{\sigma}_P$), diurnal ($\hat{\sigma}_D$), and total ($\hat{\sigma}_m$), SST estimates, separated between data from drogued and undrogued drifters. The normalized histograms of the decimal logarithm of the estimates are displayed. Mode values at the peak of the distributions and 50-th percentile values are listed in Table 5.

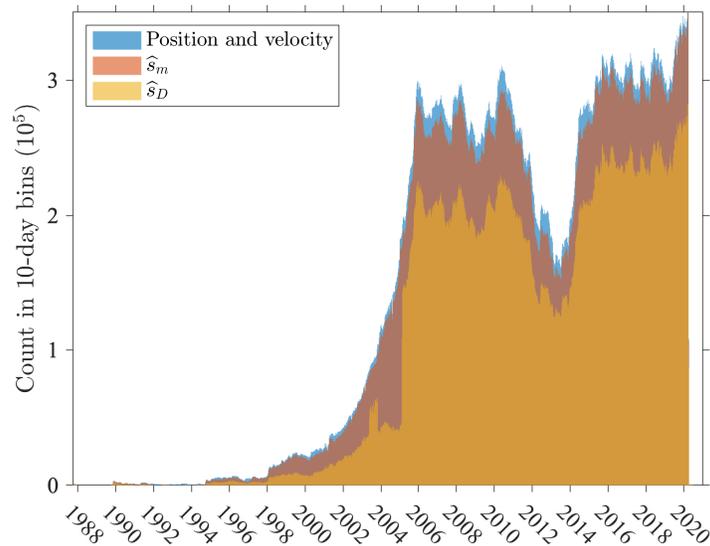
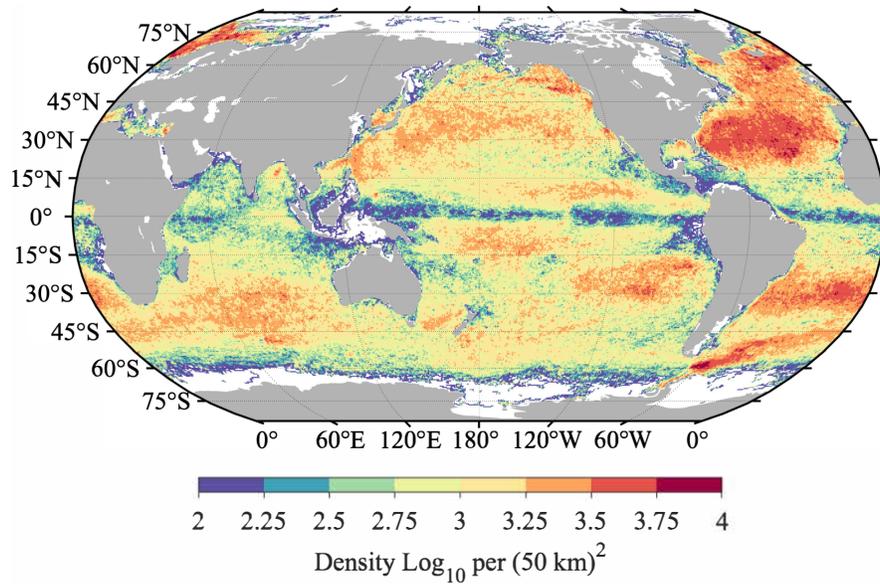


Figure 17. Top panel: Spatial distribution of Level-2 total SST estimates expressed as a density per $(50 \text{ km})^2$ in half-degree spatial bins. Only quality flag 5 data are counted. Bottom panel: Level-2 total (\hat{s}_m) and diurnal (\hat{s}_D) SST estimates temporal distribution in 10-day bins from 03-Oct-1987 13:00:00 to 30-Jun-2020 23:00:00. The temporal distribution of the matching position and velocity hourly dataset⁵ release 2 is also displayed. The temporal distribution of non-diurnal SST estimates is not displayed as it would be indistinguishable from the distribution of the total SST estimates.

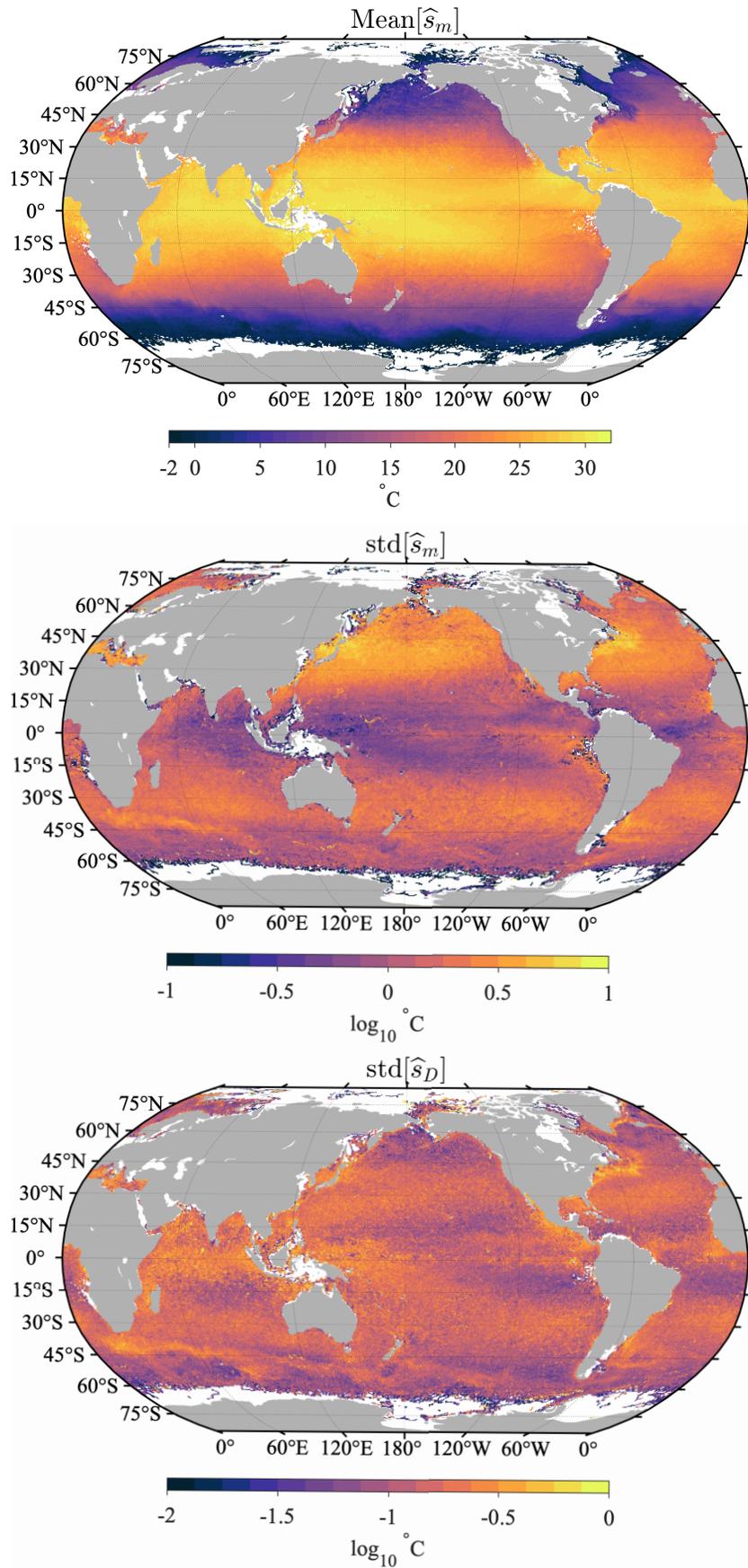


Figure 18. Top: Level-2 total SST estimates averaged in half-degree spatial bins. Middle: Level-2 total SST estimates standard deviation in half-degree spatial bins. Bottom: Level-2 diurnal SST estimates standard deviation in half-degree spatial bins. Only quality flag 5 estimates for each respective variable is used to produce these maps.

| | Argos drifters | | Iridium drifters | |
|---------|--|---|--|---|
| | Location (X, Y) & velocity (U, V) | Temperature (S) | Location (X, Y) & velocity (U, V) | Temperature (S) |
| Level-0 | X, Y | S | X, Y | S |
| Level-1 | - | S^* <i>This paper</i> | X^*, Y^*, U^*, V^* <i>Elipot et al. (2016)</i> | S^* <i>This paper</i> |
| Level-2 | 6-hourly X^*, Y^*, U^*, V^* <i>Hansen and Poulain (1996)</i> | S^* <i>Hansen and Poulain (1996)</i> | X^*, Y^*, U^*, V^* <i>Hansen and Poulain (1996)</i> | S^* <i>Hansen and Poulain (1996)</i> |
| | hourly X^*, Y^*, U^*, V^* <i>Elipot et al. (2016)</i> | S^* <i>This paper</i> | X^*, Y^*, U^*, V^* <i>Elipot et al. (2016)</i> | S^* <i>This paper</i> |

Table 1. Table of location, velocity, and temperature data products and availability from the Global Drifter Program (GDP) as described in Hansen and Poulain (1996), Elipot et al. (2016), and this paper which defines three levels of data processing. Variables without asterisks are observations and variables with asterisks are estimates. Only Level-2 data products are readily available from the GDP.

| Polynomial order (P) | 0 | 1 | 2 | 3 |
|-----------------------|--------------|----------------|---------------|----------------|
| Diurnal harmonics (N) | | | | |
| 2 | 1 (5) | 4 (6) | 7 (7) | 10 (8) |
| 3 | 2 (7) | 5 (8) | 8 (9) | 11 (10) |
| 4 | 3 (9) | 6 (10) | 9 (11) | 12 (12) |
| 5 | | 13 (12) | | |
| 6 | | 14 (14) | | |

Table 2. Table of temporal models considered for SST as a function of polynomial order and number of diurnal harmonics. Boldface numbers indicate the model identification numbers discussed in the text. Number in parentheses indicate the number of parameters of each model ($2N + P + 1$). The model ultimately chosen is number 5.

| Data Flag | 0 | 1 | 2 | 3 | 4 | 5 |
|-----------------------------|--------------------|------------------|----------------------|-----------------|--------------------|-----------------------|
| Total (\hat{s}_m) | 6,909,407 4.17% | 302,743 0.18% | 80,542 0.05% | 10,458 0.01% | 14,413 0.01% | 158,436,770 95.59% |
| Non-diurnal (\hat{s}_p) | 6,909,407 4.17% | 302,743 0.18% | 77,844 0.05% | 4,425 <0.01% | 6,823 <0.01% | 158,453,091 95.60% |
| Diurnal (\hat{s}_D) | 6,909,407 4.17% | 302,743 0.18% | 34,391,616 20.75% | 51,181 0.03% | 3,787,063 2.28% | 120,312,323 75.58% |

Table 3. Inventory of quality flags for Level-2 estimates. The target number of data points is 165,754,333 from 17,324 time series for the Global Drifter program hourly dataset version 2.0.

| $\sqrt{\hat{\sigma}_1^2}$ | | All | Drogued | Undrogued |
|---------------------------|----------------|--------------|--------------|--------------|
| Mode | Level-1 | 0.026 | 0.020 | 0.030 |
| | Level-2 | 0.026 | 0.020 | 0.030 |
| 50-th percentile | Level-1 | 0.031 | 0.025 | 0.036 |
| | Level-2 | 0.031 | 0.025 | 0.036 |
| $\sqrt{\hat{\sigma}^2}$ | | All | Drogued | Undrogued |
| Mode | Level-1 | 0.033 | 0.031 | 0.036 |
| | Level-2 | 0.033 | 0.031 | 0.036 |
| 50-th percentile | Level-1 | 0.035 | 0.030 | 0.040 |
| | Level-2 | 0.036 | 0.030 | 0.040 |

Table 4. Statistics of error variance estimates from residuals [$\hat{\sigma}_1^2$, eq. (25)] and final error variance estimates incorporating the resolution error [$\hat{\sigma}^2$, eq. (28)]. The square root values are displayed, rounded to the nearest 0.001. Units are degrees Celsius. Bold values highlights the values discussed in the text: the 50-th percentile (median) and mode values of error standard deviation for level-2 data.

| $\hat{\sigma}_P$ | | All | Drogued | Undrogued |
|------------------|----------------|--------------|--------------|--------------|
| Mode | Level-1 | 0.006 | 0.005 | 0.006 |
| | Level-2 | 0.006 | 0.005 | 0.006 |
| 50-th percentile | Level-1 | 0.006 | 0.006 | 0.007 |
| | Level-2 | 0.007 | 0.006 | 0.007 |
| $\hat{\sigma}_D$ | | All | Drogued | Undrogued |
| Mode | Level-1 | 0.016 | 0.015 | 0.016 |
| | Level-2 | 0.016 | 0.015 | 0.017 |
| 50-th percentile | Level-1 | 0.017 | 0.015 | 0.019 |
| | Level-2 | 0.018 | 0.015 | 0.020 |
| $\hat{\sigma}_m$ | | All | Drogued | Undrogued |
| Mode | Level-1 | 0.017 | 0.016 | 0.018 |
| | Level-2 | 0.018 | 0.016 | 0.019 |
| 50-th percentile | Level-1 | 0.018 | 0.016 | 0.020 |
| | Level-2 | 0.019 | 0.016 | 0.022 |

Table 5. Statistics of SST standard uncertainty estimates. Units are degrees Celsius. Values are rounded to the nearest 0.001.

| Description of the content of variables in the NetCDF files | Names of variables |
|---|--|
| Time of estimate | time |
| Estimated latitude from drifting locations | latitude |
| Estimated longitude from drifting locations | longitude |
| Estimated northward velocity from drifting locations | surface northward sea water velocity |
| Estimated eastward velocity from drifting locations | surface eastward sea water velocity |
| Estimated 95% confidence interval in latitude from drifting locations | 95% confidence interval in latitude |
| Estimated 95% confidence interval in longitude from drifting locations | 95% confidence interval in longitude |
| Estimated 95% confidence interval in northward velocity from drifting locations | 95% confidence interval in northward velocity |
| Estimated 95% confidence interval in eastward velocity from drifting locations | 95% confidence interval in eastward velocity |
| Estimated near-surface sea water temperature from drifting buoy measurements | fitted sea water temperature |
| Estimated near-surface non-diurnal sea water temperature from drifting buoy measurements | fitted non-diurnal sea water temperature |
| Estimated near-surface diurnal sea water temperature from drifting buoy measurements | fitted diurnal sea water temperature anomaly |
| Estimated one standard error of near-surface sea water temperature estimate from drifting buoy measurements | standard uncertainty of fitted sea water temperature |
| Estimated one standard error of near-surface non-diurnal sea water temperature estimate from drifting buoy measurements | standard uncertainty of fitted non-diurnal sea water temperature |
| Estimated one standard error of near-surface diurnal sea water temperature estimate from drifting buoy measurements | standard uncertainty of fitted diurnal sea water temperature anomaly |
| Quality flag for fitted sea water temperature | fitted sea water temperature quality flag |
| Quality flag for fitted non-diurnal sea water temperature | fitted non-diurnal sea water temperature quality flag |
| Quality flag for fitted diurnal sea water temperature anomaly | fitted diurnal sea water temperature anomaly quality flag |

Table 6. Data record information for the Level-2 data product and details of variables in drifter NetCDF files.