AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2016JC011716

A global surface drifter data set at hourly resolution

Shane Elipot[1], Rick Lumpkin[2], Renellys C. Perez[2,3], Jonathan M. Lilly[4], Jeffrey J. Early[4], and Adam M. Sykulski[5]

[1]Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida, USA, [2]NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida, USA, [3]Cooperative Institute of Marine and Atmospheric Studies, University of Miami, Miami, Florida, USA, [4]Northwest Research Associates, Redmond, Washington, USA, [5]Department of Statistical Science, University College London, London, UK

Key Points:
- Surface drifters tracked by Argos exhibit location errors which are not normally distributed
- An hourly interpolation method is devised, taking into account the location error distributions
- A new global data set of hourly drifter locations and velocities evidences high-frequency oceanic motions at all latitudes

Correspondence to:
S. Elipot,
selipot@rsmas.miami.edu

Citation:
Elipot, S., R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski (2016), A global surface drifter data set at hourly resolution, *J. Geophys. Res. Oceans*, *121*, 2937–2966, doi:10.1002/2016JC011716.

Received 11 FEB 2016
Accepted 30 MAR 2016
Accepted article online 4 APR 2016
Published online 4 MAY 2016

Abstract The surface drifting buoys, or drifters, of the Global Drifter Program (GDP) are predominantly tracked by the Argos positioning system, providing drifter locations with *O*(100 m) errors at nonuniform temporal intervals, with an average interval of 1.2 h since January 2005. This data set is thus a rich and global source of information on high-frequency and small-scale oceanic processes, yet is still relatively understudied because of the challenges associated with its large size and sampling characteristics. A methodology is described to produce a new high-resolution global data set since 2005, consisting of drifter locations and velocities estimated at hourly intervals, along with their respective errors. Locations and velocities are obtained by modeling locally in time trajectories as a first-order polynomial with coefficients obtained by maximizing a likelihood function. This function is derived by modeling the Argos location errors with *t* location-scale probability distribution functions. The methodology is motivated by analyzing 82 drifters tracked contemporaneously by Argos and by the Global Positioning System, where the latter is assumed to provide true locations. A global spectral analysis of the velocity variance from the new data set reveals a sharply defined ridge of energy closely following the inertial frequency as a function of latitude, distinct energy peaks near diurnal and semidiurnal frequencies, as well as higher-frequency peaks located near tidal harmonics as well as near replicates of the inertial frequency. Compared to the spectra that can be obtained using the standard 6-hourly GDP product, the new data set contains up to 100% more spectral energy at some latitudes.

1. Introduction

The Global Drifter Program (GDP) maintains a global array of more than 1000 satellite-tracked surface drifting buoys, hereafter referred to as drifters (http://www.aoml.noaa.gov/phod/dac/index.php). The GDP is part of NOAA's Global Ocean Observing System and is a scientific project of the international Data Buoy Cooperation Panel (DBCP, http://www.jcommops.org/dbcp/). Drifters not only provide information on oceanic drift at 15 m, the nominal depth of their drogue, but also measure sea surface temperature and, for a subset of the drifters, sea level pressure, sea surface salinity, and surface winds [*Lumpkin and Pazos*, 2007]. The GDP has been instrumental in describing and advancing the dynamical understanding of large-scale and regional oceanic variability on monthly to climate time scales [*Maximenko et al.*, 2013; *Lumpkin and Johnson*, 2013]. Drifter data are routinely utilized to improve global and regional weather forecasts (L. *Centurioni et al.*, Global observing system for measuring sea level atmospheric pressure: Effects and impacts on numerical weather prediction, submitted to *Bulletin of the American Meteorological Society*, 2016).

Drifters have historically been tracked by the Argos positioning system [*CLS*, 2011], but in recent years, the GDP has been gradually implementing a tracking of drifters by the Global Positioning System (GPS), which provides drifter locations with meter-scale accuracy and precision at best, at hourly time scales or shorter, relayed almost instantly via Iridium satellites. The growing data set from GPS-tracked drifters should become a very valuable tool for global and regional studies of small-scale processes in the ocean [e.g., *Centurioni et al.*, 2015]. However, drifters tracked by Argos, which provides drifter locations with much less accuracy and precision than GPS [e.g., *Lopez et al.*, 2014], have already allowed many investigators to study oceanic processes characterized by high-frequencies and small spatial scales such as inertial oscillations, tides, and submesoscale vortices [e.g., *Elipot and Lumpkin*, 2008; *Elipot et al.*, 2010; *Chaigneau et al.*, 2008;

Lumpkin and Elipot, 2010; *Poulain and Centurioni*, 2015]. Yet the Argos sampling is temporally nonuniform, and to facilitate analyses, the aforementioned studies required ad hoc processing of the drifter trajectories in order to obtain at regular temporal intervals not only drifter locations but also drifter velocities. In general, the global data set can constitute a challenge for global and regional analyses, especially when the focus of investigations is directed toward short temporal and spatial scales.

From the onset of the GDP, drifter battery power has often been conserved by sampling location for 1 day followed by 2 days of nonsampling [*Hansen and Herman*, 1989]. Thus, the Data Assembly Center (DAC) of the GDP is routinely processing and interpolating the Argos fixes to produce drifter locations continuously along trajectories at 6 h intervals, using an objective interpolation method commonly called Kriging [*Hansen and Poulain*, 1996], tuned to the original sampling scheme [*Hansen and Herman*, 1989]. Since 2000, the 1 day on, 2 day off sampling scheme has been abandoned thanks to increased battery lives and other technological advancements [*Lumpkin and Pazos*, 2007]. In parallel, the number of operational satellites of the Argos constellation has continuously increased (up to six to date), so that the typical time interval between two consecutive Argos fixes has dramatically reduced to between 1 and 2 h [*Elipot and Lumpkin*, 2008]. Despite this increased sampling frequency, the readily available drifter trajectory and velocity product generated by the DAC are still provided with a 6 h sampling period, which is generally inadequate for the study of high-frequency processes in the ocean, as will be demonstrated in this paper.

Thus, the first goal of this study is to justify and present the methods we have selected to produce a new global data set of hourly geographical locations and horizontal velocities along drifter trajectories, derived from the temporally nonuniform data set of locations from Argos, and from GPS. The second goal is to provide a few examples of how this new global data set can be utilized to study oceanic velocity variance at high frequencies. For Argos drifters, our global interpolation method was determined after testing four interpolation methods (including an update on the Kriging method currently used by the DAC) using the trajectories of 82 drifters tracked contemporaneously by both Argos and by GPS in the North Atlantic Ocean as part of the Salinity Processes in the Upper Ocean Regional Study (SPURS) [*Centurioni et al*., 2015]. For GPS drifters, we have chosen to apply a smoothing and linear interpolating method.

This paper is organized as follows. Section 2 presents an assessment of the accuracy of Argos locations by comparing them to GPS locations. This first step is found to be a prerequisite for assessing and selecting a method of interpolation. Section 3 presents the four methods of interpolations. The performances of the methods are compared in section 4. Section 5 presents the implementation of the chosen interpolation methods on the global Argos data set, and highlights new global observations of high-frequency drifter motions. Concluding statements are given in section 6.

2. Assessment of Argos Locations Using the SPURS Drifter Data

In this section, we assess the errors of Argos locations by comparing them to GPS locations. We first describe the quality controls applied to Argos locations, then describe the quality controls and processing applied to GPS locations, and finally compare the two to quantify the Argos errors. We note here that GPS location accuracy is expected to be on the order of a few meters, and thus GPS errors are considered negligible compared to Argos errors [*NSTB/WAAS T&E Team*, 2014].

2.1. Quality Control of Argos Locations

To devise an interpolation method, we use 82 drifters deployed for SPURS, tracked by both GPS and Argos between 21 August 2012 and 1 October 2013 (Figure 1). These data are available through the SPURS data repository of the NASA PODAAC website (https://podaac.jpl.nasa.gov/SPURS). Argos locations were transmitted to the DAC at the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) in Miami, FL. The standard data quality control procedures described in *Hansen and Poulain* [1996] and *Lumpkin and Pazos* [2007] are applied to the Argos data set to remove spurious data, and the status of the drogue (which when attached keeps the drifter following the 15 m currents) is assessed as described in *Lumpkin et al*. [2013]. The final quality-controlled (QC) Argos locations total 202,404 data points, 79% with drogue on, amounting to 17,667 drifter days with trajectory lengths ranging between 29 and 405 days. Location data with or without drogue attached will be all considered here as the two types of data do not lead to significant differences in our analyses, unless specifically stated.

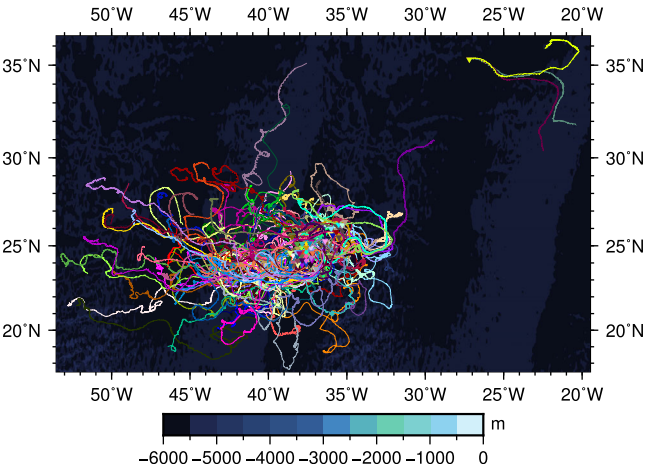

Figure 1. Trajectories of 82 drifters from Argos tracking deployed as part of the SPURS experiment from 21 August 2012 to 1 October 2013. Inverted triangles indicate the deployment locations.

2.2. Quality Control and Processing of the GPS Locations

For SPURS, the GPS receivers on the drifters were set up to acquire location fixes at 30 min intervals, and to transmit data via the Argos system. Because of the intermittency of Argos satellite coverage, potential overlapping of Argos satellite passes, and the temporal precision of the data archived at the DAC (1/1000th of day or 1.44 min), the time intervals $\Delta t = t_{k+1} - t_k$ between pairs of consecutive GPS fixes are not uniform. The distribution of Δt exhibits peaks near zero and near multiples of 30 min (not shown). Outside of a few exceptions, the Δt values near zero correspond to consecutive locations which are not distinguishable within the longitudinal and latitudinal precision of the data archived at the DAC (1/1000 of a degree or about 111 m in the meridional direction). Thus, we eliminate such redundant locations by discarding the second point of each pair when $\Delta t < 20$ min or when the separation distance is less than 50 m, or both. Note that angular separation is calculated using the haversine formula, and converted to separation distance using an Earth's radius of 6371 km.

The GPS locations still include occasional extreme outliers, with unphysically large and apparently random changes of longitude and latitude. All GPS locations are accompanied by a quality index which we do not find useful for our purpose, as it does not appear to accurately flag such outliers. To detect these, we devised a multiple-step procedure now applied routinely by the DAC to all GPS data entering the GDP database. As a first step, for each GPS location, we look for the temporally closest Argos location of quality index 2 or 3 (we will explain Argos classes in the next section) and calculate the absolute speed required to go between the GPS location and this Argos location. If the resulting speed is larger than 3 m s^{-1}, these GPS locations are discarded. The second step consists of comparing the time series of longitude and latitude to filtered versions of these time series, obtained by applying a one-dimensional five-point median filter, with the time series mirrored at either end. Original values are flagged and removed if they are more than five standard deviations from the five-point median. This operation is repeated five times. The remaining fixes constitute what we call the QC GPS fixes (Figure 2).

Location data from GPS are taken as truth for evaluating Argos locations (next subsection) and for assessing the four interpolation methods applied to Argos locations (section 4). We assess the interpolation methods based on estimated locations and also on estimated horizontal drifter velocities, since both quantities are estimated at the same time by the methods (except Kriging). Thus, it is necessary to derive velocities from GPS locations as well for comparisons. A straightforward method to calculate GPS velocities is to first linearly interpolate longitude and latitude coordinates to hourly time steps, and second to compute velocity by a central difference scheme, which is the average of arriving and departing velocities at each step. Instead, we choose to implement another method which consists of estimating both locations and horizontal velocities at the nonuniform times by modeling locally in time the GPS trajectory using a first-order polynomial. For this, we apply a robust regression method called the Locally Weighted Scatterplot Smoother, or LOWESS [*Cleveland*, 1979], described in Appendix A. This choice of method is motivated by the impact of the central difference scheme on the velocity variability at high frequencies. Figure 3 is a comparison between rotary spectra calculated for velocity time series from the LOWESS method (after subsequently linearly interpolating these velocities onto uniform hourly times), and rotary spectra calculated for velocity time series obtained by directly linearly interpolating the nonuniform QC GPS locations onto uniform hourly times and then using a central differencing scheme. The central difference scheme overattenuates power at high frequencies, while the LOWESS velocities renders an approximately constant spectral slope in the same

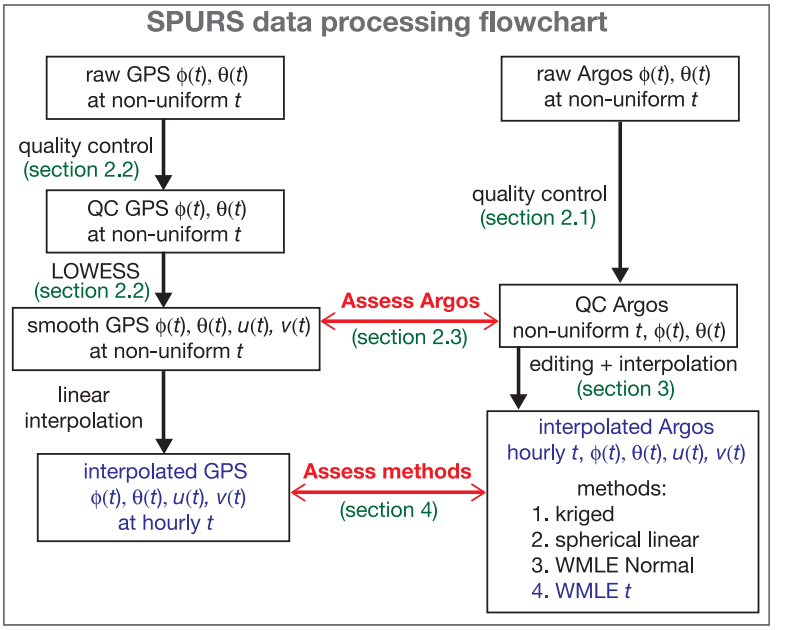

Figure 2. Processing chart of the SPURS data set for this study. The green labels reference sections of this article. The blue labels indicate the final products of this study. Raw GPS and raw Argos longitude (ϕ) and latitude (θ) data are recorded at different nonuniform time steps t. Interpolated GPS and Argos ϕ, θ, and horizontal velocity components (u, v) are estimated on common hourly GMT time steps.

frequency range. The resulting smoothed GPS data (Figure 2) consist of 676,201 smooth estimates of GPS locations and velocities, amounting to 17,826 drifters days, with trajectory lengths ranging between 29 and 405 days.

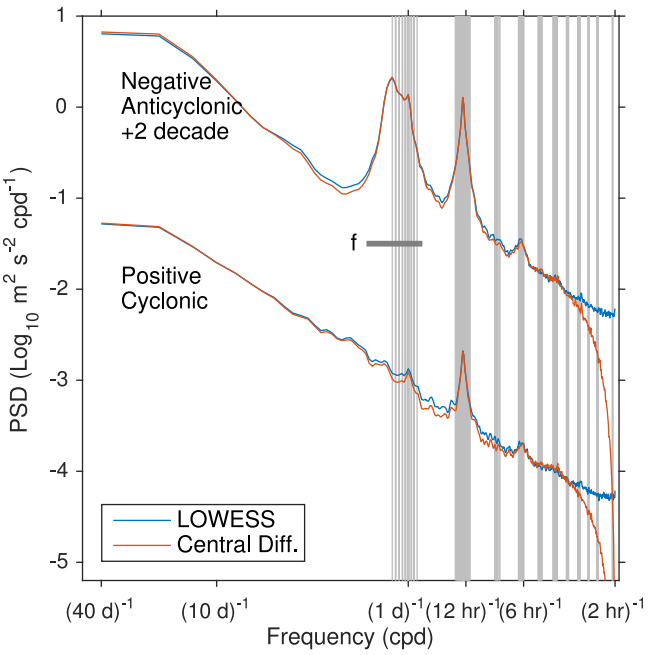

Figure 3. Rotary velocity spectra estimates for GPS velocities obtained from a LOWESS filter or by central differences of linearly interpolated GPS locations. The negative, anticyclonic, spectra are offset vertically by two decades for legibility. The thin vertical gray lines correspond to astronomical tidal frequencies. The horizontal bar marked by f indicates the range of inertial frequencies for these data.

2.3. Assessment of Argos Location Errors

In this section, we assess the quality of Argos locations, which are determined using the Doppler shift on the transmission frequency between Argos platforms and an Argos satellite flying above [*CLS*, 2011]. The localization algorithm selected from Argos by the DAC has historically been a least squares analysis, producing location as well as an associated location error. Since March 2011, a multimodel Kalman filter solution has been made available by Argos, and has been selected by the DAC. Argos location errors are not horizontally uniform, and are represented by an ellipse with semimajor axis a and semiminor axis b [*CLS*, 2011]. However, by specification, the GDP is provided with a location error characterized by a class, associated with an equivalent radius error $\sqrt{ab}$, assumed to be representative of one standard deviation of the errors. The error classes are labeled 0 to 3, each associated with a range of values for the radius error. The nominal

Table 1. Previous Studies Assessing Argos Location Errors for Drifters

	Zonal Error (Std, m)	Meridional Error (Std, m)
Hansen and Herman [1989]: 45 grounded drifters	277	173
Lumpkin and Pazos [2007]: one grounded drifter	630	270
Lopez et al. [2014]: 44 moving drifters		
Least squares algorithm	622	481
Kalman filter	516	433

error values are less than 250 m for class 3, 250–500 m for class 2, 500–1500 m for class 1, and greater than 1500 m for class 0. Two worse location classes are possible, classes A and B, but these are not transmitted to the DAC. *Lopez et al.* [2014] concluded that the Kalman filter positioning was significantly better than the least squares only for locations classes A and B.

The statistical description of Argos location errors is crucial for analyses of Argos location data [e.g., *Jonsen et al.*, 2005; *Boyd and Brightsmith*, 2013], yet previous assessments of Argos location errors for drifters are few. A common result of previous studies is that the error in the zonal direction is generally larger than in the meridional direction (Table 1). We estimate and describe the location errors for the SPURS drifters by comparing GPS and Argos locations. This result will later be used for the interpolation methods tested in section 3. To compute errors, we linearly interpolate the smooth GPS locations at the QC Argos nonuniform times (Figure 2). We define Argos location errors as the signed longitude differences ($\Delta\phi$), signed latitude differences ($\Delta\theta$), and positive angular separations ($\Delta\lambda$) between interpolated GPS locations and Argos locations. In order to minimize the impact on errors of the linear interpolation itself, we consider the errors calculated only when a GPS location is available both within an hour before and an hour after Argos sampling times. This results in considering 92% of the nonuniform QC Argos positions (186,892 Argos data points). The SPURS data set does not contain any location class 0, and therefore this class cannot be assessed here.

We estimate the error distributions using kernels [e.g., *Fan and Gijbels*, 1996, p. 47] at 10^{-3} degrees resolution, for all Argos location classes and for each of the classes. We also fit the $\Delta\phi$ and $\Delta\theta$ values (not the kernel estimate curves) to two types of probability distribution functions (PDFs) (Figure 4): the normal, or Gaussian, PDF, and the *t* location-scale PDF, also called nonstandardized Student's *t* PDF [e.g., *Jackman*, 2009, p. 507], hereafter simply called *t* PDF. Both PDFs are symmetric around a location parameter μ (corresponding to the mean for the normal PDF) which indicates the value with maximum probability. The PDF curves both decrease from this maximum at a rate characterized by a scale parameter σ (corresponding to the standard deviation for the normal distribution). The *t* PDF is characterized by an additional parameter, ν, which is called the shape parameter. This last parameter permits heavier tails and a narrower central peak compared to a normal PDF with the same parameter σ. The *t* PDF tends to a normal PDF as ν tends to infinity. The analytical expressions for the normal and *t* location-scale PDFs, respectively, are

$$p(z|\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}, \tag{1}$$

$$p(z|\mu,\sigma,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{1}{\nu}\left(\frac{z-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}, \tag{2}$$

where $\Gamma(a)=\int_0^\infty t^{a-1}\exp(-t)\,dt$ is the Gamma function. We also fit the angular separation error values to the exponential PDF, $p(z|\sigma)=e^{-z/\sigma}/\sigma$. This last PDF may not be the best model but its simplicity allows us to obtain qualitative assessments of the distributions. We compute maximum likelihood estimates (MLE) of the parameters of the PDF models using routines from the statistical toolbox of MATLAB. See Tables 2, 3, and 5 for the parameter values and Figure 4 for the corresponding curves. Note that the sample mean and biased sample standard deviation statistics are the MLE of μ and σ, respectively, for the normal PDF, so that these values are both statistics of the data and parameter estimates. In the same way, for the exponential PDF, the sample mean is the MLE of σ. Finally, we also compute a 2-D histogram of the longitude and latitude errors, in bins of 10^{-3} degrees, to investigate a possible dependence between the two types of errors (Figure 6).

The longitude errors are nearly centered, with mean errors of 10 m or less in magnitude, but the latitude errors exhibit positive biases larger than 60 m for classes 2 and 3 (Table 2). The standard deviations of errors are commensurable in magnitude with previous findings (Table 1), are larger in the zonal direction than in

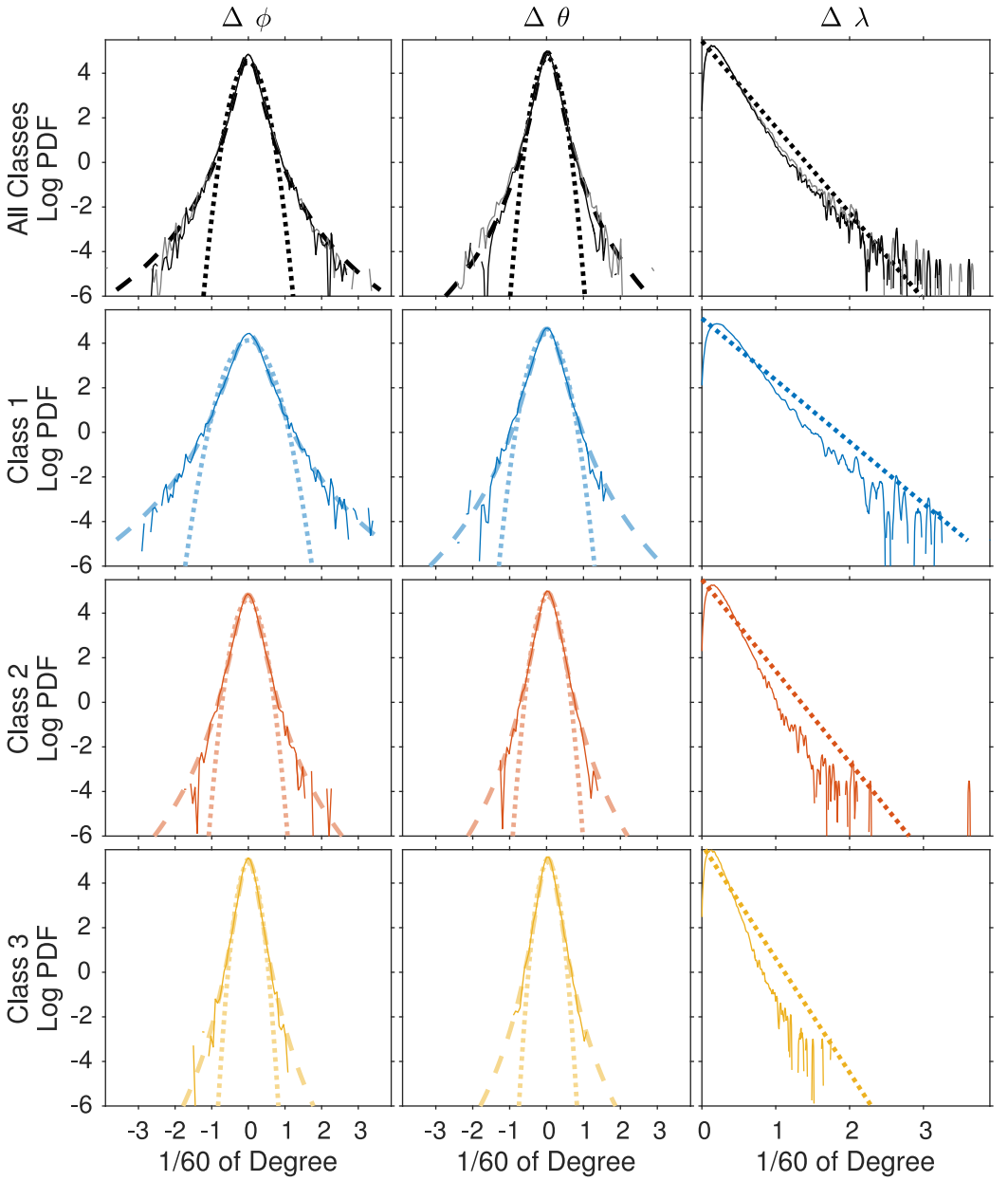

Figure 4. (left column) Longitude error ($\Delta\phi$), (middle column) latitude error ($\Delta\theta$), and (right column) angular separation error ($\Delta\lambda$) distributions for all Argos location classes, and location classes 1, 2, and 3 (from top to bottom). In each plot, the thin solid curve is a kernel estimate. For longitude and latitude errors, the dotted heavy curve is a fit to a normal PDF, and the dashed heavy curve is a fit to a *t* location-scale PDF. In the top plots, the kernel estimate of the distribution of all errors (including "nonverifiable" ones) is shown as a thin solid gray line.

the meridional direction, and decrease with increasing class, as expected. We compute the observed probabilities that $\Delta\phi$ and $\Delta\theta$ fall within the intervals defined by ± 1 and ± 1.96 standard deviations from their means. These intervals would, respectively, comprise 68% and 95% of the distributions if the errors were normally distributed (Table 2). For all classes and for each class, the probability of falling within one standard deviation of the mean is higher ($>$73%) than the expectation for a normal PDF. The observed probabilities of falling within 1.96 standard deviations of their means are sometimes higher and sometimes lower, depending on the coordinate and the class, yet close (94.1–97.1%) to the normal expectation. Thus, the normal PDF model does not appears to be a good representation of the error distributions for geographical coordinates. Indeed, the kernel curves clearly exhibit heavier tails and narrower central peaks compared to the normal distributions fitted to the data (Figure 4). To investigate further such nonnormal character, we

Table 2. Results of Fits of Argos Longitude and Latitude Errors to Normal PDFs for Each of the Location Classes, and for All Classes[a]

	All Classes	Class 1	Class 2	Class 3
Lon. Err. $\Delta\phi$				
$10^2\times\mu_{\Phi_k}$ (sample mean)	0.003° (−3 m)	0.010° (10 m)	−0.009° (−9 m)	−0.005° (−6 m)
$10^2\times\sigma_{\Phi_k}$ (sample std)	0.445° (450 m)	0.641° (648 m)	0.390° (394 m)	0.294° (297 m)
Prob. within $\mu_{\Phi_k}\pm\sigma_{\Phi_k}$ (68%)	77.3%	77.0%	74.2%	75.2%
Prob. within $\mu_{\Phi_k}\pm1.96\sigma_{\Phi_k}$ (95%)	94.5%	94.8%	94.1%	97.1%
Lat. Err. $\Delta\theta$				
$10^2\times\mu_{\Theta_k}$ (sample mean)	0.045° (50 m)	0.006° (6 m)	0.055° (62 m)	0.062° (69 m)
$10^2\times\sigma_{\Theta_k}$ (sample std)	0.364° (405 m)	0.473° (526 m)	0.340° (379 m)	0.280° (312 m)
Freq. within $\mu_{\Theta_k}\pm\sigma_{\Theta_k}$ (68%)	73.6%	74.2%	73.3%	74.1%
Freq. within $\mu_{\Theta_k}\pm1.96\sigma_{\Theta_k}$ (95%)	92.8%	93.8%	94.3%	96.2%

[a]Longitude parameters are converted to distance using the 24.7°N median latitude of the data. The observed probability of the errors to fall within ±1 and ±1.96, the standard deviation from the mean are also listed, which should be 68% and 95%, respectively, for a normal PDF.

conduct the one-sample Kolmogorov-Smirnov (KS) test [*Massey*, 1951] for the null hypotheses that coordinate errors are normally distributed with the location and scale parameters fitted to the data (Table 4). This test fails for all cases.

The previous results motivate us to consider instead *t* PDFs. We find that the fits to these PDFs are closer to the kernel estimates (Figure 4), and capture the narrowing of the distributions with a scale parameter σ that decreases with increasing class, and also capture the decreasing tendency for heavy distribution tails with a shape parameter ν that increases with increasing class (Table 3). We conduct the KS test for the *t* PDFs and we find that the test does not reject the null hypothesis at the 5% level for longitude errors for Argos classes 1–3. However, the test does reject the null hypothesis of *t* PDFs for latitude errors, yet the *p* value of the test statistics is more favorable than for the null hypothesis of normal PDF (Table 4). The *p* value is the probability of observing a KS test statistic as large as, or larger than, the observed value under the null hypothesis. A preference for the *t* PDF model over the normal PDF model is also demonstrated by scatterplots between observed quantile values and theoretical quantile values for either normal or *t* PDFs with the fitted parameters (Figure 5). In these plots, the scatter points for the *t* PDFs are closer to the regression-one curve, compared to the points for normal PDFs. Only toward extremes values does the *t* PDF model appear not to represent the data correctly. Finally, we also calculate the Aikake Information Criterion statistics [*Akaike*, 1974]. This statistics assess whether there is statistical justification for using more parameters when choosing between models that are nested. We can use this to assess whether there is significant evidence to reject normal PDFs in favor of *t* PDFs. We can do this as the normal PDF is functionally (although not formally) nested with a *t* PDF when $\nu=\infty$. The calculated statistics (not listed) indicate that the *t* PDF is better suited than the normal PDF to represent the $\Delta\phi$ and $\Delta\theta$ distributions.

Table 3. Results of Fits of Argos Longitude and Latitude Errors to *t* Location-Scale PDFs for Each of the Location Classes, and for All Classes[a]

	All Classes	Class 1	Class 2	Class 3
Lon. Err. $\Delta\phi$				
$10^2\times\mu_{\Phi_k}$	−0.004° (−4 m)	0.009° (9 m)	−0.009° (−9 m)	−0.005° (−5 m)
$10^2\times\sigma_{\Phi_k}$	0.294° (297 m)	0.448° (453 m)	0.297° (300 m)	0.236° (238 m)
ν_{Φ_k}	3.38	3.75	4.7	5.55
Prob$(\phi<\Delta\Phi_k-\mu_{\Phi_k}<\phi)=0.68$				
$10^2\times\phi$	0.342° (346 m)	0.513° (519 m)	0.330° (334 m)	0.257° (260 m)
Prob$(\phi<\Delta\Phi_k-\mu_{\Phi_k}<\phi)=0.95$				
$10^2\times\phi$	0.879° (888 m)	1.278° (1291 m)	0.779° (787 m)	0.589° (595 m)
Lat. Err. $\Delta\theta$				
$10^2\times\mu_{\Theta_k}$	0.048° (54 m)	0.014° (16 m)	0.054° (60 m)	0.060° (67 m)
$10^2\times\sigma_{\Theta_k}$	0.266° (296 m)	0.357° (397 m)	0.268° (299 m)	0.221° (246 m)
ν_{Θ_k}	4.08	4.44	5.12	5.11
Prob$(\theta<\Delta\Theta_k-\mu_{\theta_k}<\theta)=0.68$				
$10^2\times\theta$	0.301° (335 m)	0.400° (445 m)	0.296° (329 m)	0.243° (270 m)
Prob$(\theta<\Delta\Theta_k-\mu_{\theta_k}<\theta)=0.95$				
$10^2\times\theta$	0.733° (815 m)	0.954° (1061 m)	0.685° (762 m)	0.564° (627 m)

[a]Longitude parameters are converted to distance using the 24.7°N median latitude of the data. Also listed are the calculated error values around the mean that define 68% and 95% of the fitted *t* location-scale PDFs.

Table 4. Results of One-Sample Kolmogorov-Smirnov Tests for the Null Hypotheses That the Longitude and Latitude Errors Are Distributed Like Normal or *t* Location-Scale PDFs[a]

	All Classes	Class 1	Class 2	Class 3
Lon. Err. $\Delta\phi$				
Normal	0	0	0	0
t location scale	0.0444	**0.4274**	**0.2405**	**0.7279**
Lat. Err. $\Delta\theta$				
Normal	0	0	0	0
t location scale	0.0002	0.0044	0.0020	0.0180

[a]For each type of errors and for each class, the table lists the *p* value of the hypothesis test. $p < 0.05$ indicates that the test rejects the null hypothesis at the 5% level, thus bold values indicate acceptance of the test at the 5% level.

Since *t* PDFs are a better match, we use the analytical expression of the *t* PDF (2) with the fitted parameters for each class in order to compute the values of $\Delta\phi$ and $\Delta\theta$ that define 68% and 95% of the distributions around their location parameters (Table 3). In particular, the 68% values can be compared to the Argos equivalent radius for each class. Class 3 is supposed to be representative of a 250 m radius: this is an underestimation of the 68% value for $\Delta\phi$ (260 m) and for $\Delta\theta$ (270 m). Class 2 is supposed to be representative of a 250–500 m radius, which is appropriate for both longitude and latitude errors 68% values (334 and 329 m, respectively). Finally, class 1 is supposed to be representative of a 500–1500 m radius range, which is appropriate for the 68% value for $\Delta\phi$ (519 m) but is an overestimation for the 68% value for $\Delta\theta$ (445 m). Alternatively, one can ask if the class equivalent radii are representative of the 68th percentile of the $\Delta\lambda$ distribution (Table 5): 250 m is an underestimation for class 3, 250–500 m is an underestimation for class 2, and 500–1500 m is appropriate for class 1.

An implication for the interpolation problem is that if the assumption of normality of the data is made, as is the case in standard least squares estimation methods, then outliers in the data may be given unrealistic weights, as we have shown that the normal model is not a suitable description of the observed distributions. In addition, formal confidence intervals based on normal distributions may inaccurately represent the true errors due to the original distributions of the data. We will demonstrate in the rest of the paper that the method of interpolation of the Argos data that we favor does not assume that the errors are normally distributed, but rather distributed as *t* PDFs.

3. Methods of Interpolation

In this section, we introduce four interpolation methods tested on the SPURS data set to motivate the choice of the one ultimately chosen for the global data set. The first method, Kriging, is presented because it is the method currently used by the DAC to produce the 6-hourly GDP product. Here we present an adaptation of this method for hourly time steps, and the results should constitute a benchmark for other methods. The second method is a refinement of the simple linear interpolation in order to take into account the sphericity of the Earth, hence it is called "spherical linear interpolation." Linear interpolation is considered and discussed because it straightforward and widely used, and any other more sophisticated method would be expected to perform better. The third and fourth methods are called "weighted maximum likelihood estimates," and consist of modeling locally drifter coordinates as linear functions of time, which is conceptually one step forward compared to linear interpolation. The model parameters are obtained simultaneously for both coordinates by maximizing a likelihood function, either assuming a normal PDF of the data for the third method (making this method equivalent to a least squares fit, and this is why it is considered here) or assuming a *t* PDF for the fourth method. This fourth method is eventually selected to be applied to the global Argos drifter data set.

Table 5. Statistics of Angular Separation $\Delta\lambda$ Argos Errors[a]

Angular Separation $\Delta\lambda$	All Classes	Class 1	Class 2	Class 3
10^2 mode	0.210° (234 m)	0.340° (378 m)	0.220° (245 m)	0.210° (234 m)
10^2 50th percentile	0.351° (390 m)	0.496° (552 m)	0.345° (384 m)	0.280° (312 m)
10^2 68th percentile	0.485° (539 m)	0.684° (761 m)	0.469° (521 m)	0.376° (418 m)
10^2 95th percentile	1.044° (1161 m)	1.423° (1582 m)	0.933° (1037 m)	0.738° (820 m)
Exp. Fit: $10^2 \times \sigma$ (sample mean)	0.432° (480 m)	0.602° (669 m)	0.406° (452 m)	0.326° (362 m)

[a]The sample mean is also the result of the fit of the scale parameter σ of an exponential PDF model.

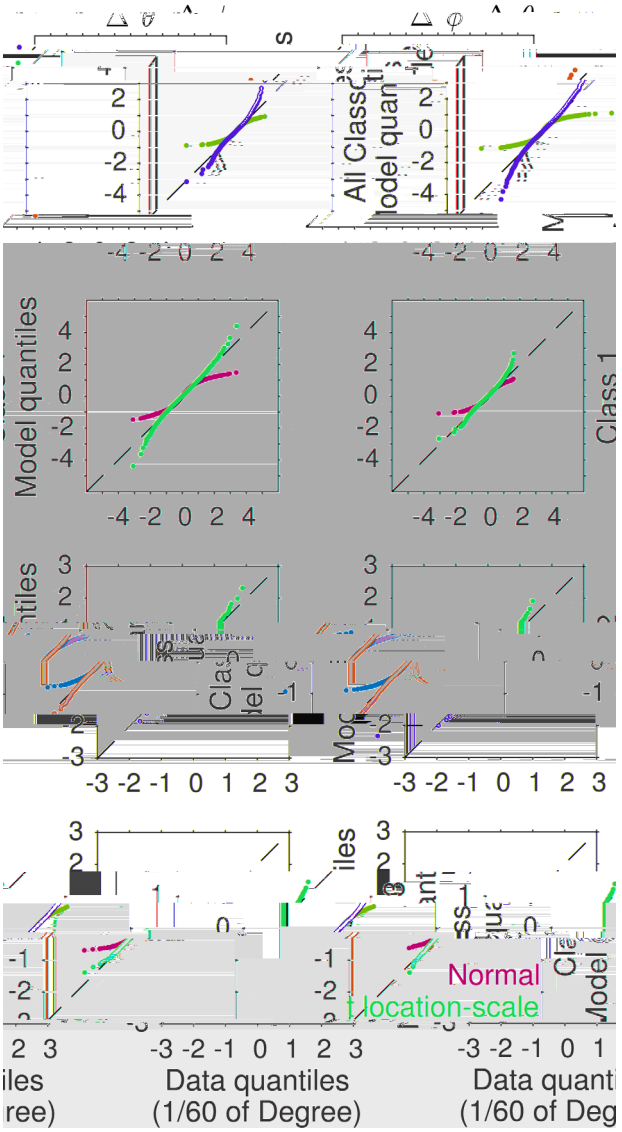

Figure 5. Quantile-quantile plots for (left column) longitude errors ($\Delta\phi$) and (right column) latitude errors ($\Delta\theta$) for all Argos location classes and location classes 1, 2, and 3 (from top to bottom). Each plot shows the quantile values of the observations on the *x* axis and the theoretical quantile values on the *y* axis for normal (blue) and t location-scale (red) PDFs with the parameters listed in Tables 2 and 3.

3.1. Editing of Argos Locations

After experimentation, we found it beneficial to further edit the QC Argos locations before interpolating (Figure 2). An Argos drifter fix is determined during the 10–20 min duration of a satellite's pass [*CLS*, 2011; *Lopez et al.*, 2014]. Because there are several Argos operational satellites, it is often possible that a drifter is located within more than one satellite's view at a time, and thus several location estimates may exist within overlapping pass durations. Averaging such locations within fixed windows should reduce the location error, yet it is not obvious what location class or error value should be associated with such averages. One option would be to keep all available observations, but when fitting a trajectory model with a fixed number of observations, as we do when we estimate hourly positions, instabilities in finding numerical solutions to our interpolations may arise if divergent fixes are too closely spaced together. Thus, before interpolation, we choose to process the Argos locations as follows.

Starting from the first location of each trajectory, we look for observations within a 20 min temporal window in the future. These observations are ranked by their Argos classes and only the ones with the best location class are kept. The selected locations are then reduced to a single location with this best class, and with averaged dates, longitudes, and latitudes. Most often, this selection algorithm amounts to selecting the best location between only two locations with different classes within a 20 min window. A sometimes second and necessary step is to detect the rare consecutive two observations that lead to a null displacement. In this case, the lesser quality location, or the first of two locations with the same class, is discarded. These two selection algorithms amount to discarding 13.1% of the QC Argos locations for the SPURS data set.

3.2. Revisiting the Kriging Method

3.2.1. Kriging Equations and the Variogram

Kriging is a general term in geostatistics which designates methods of prediction of an unknown field variable which is function of one or more coordinates. Given a set of observations, and possibly their associated observation errors, these methods utilize knowledge of some statistics of the field variable, such as the cross-covariance function, or the variogram function which is the variance of the difference of the field variable along its coordinates. In oceanic and atmospheric sciences, methods of interpolation, or of mapping, based on modeled covariance functions are traditionally called objective or optimal analyses [*Bretherton et al.*, 1976]. Under the condition of second-order stationarity, covariance and variogram functions are

linearly related so that using one function or the other are equivalent. Yet it is clear that the covariance of drifter displacements displays both geographic and spatial variability, indicating an underlying nonstationary process. Therefore, the use of the variogram for interpolating trajectories is advocated. One-dimensional Kriging of drifter coordinates along trajectories with a model of the variogram was first implemented by *Hansen and Herman* [1989], and later refined by *Hansen and Poulain* [1996] to incorporate observational errors in the method. Thus, in order to implement Kriging, one needs to specify the autocovariances of the location errors, as well as to define, or model and fit to data, the variogram function. Details of the method can be found in the aforementioned papers.

3.2.2. Implementations at 6 h Intervals by the GDP/DAC

The DAC applies Kriging with a number of settings inherited from *Hansen and Herman* [1989] and *Hansen and Poulain* [1996]. Drifter locations are interpolated at 6 h intervals at times t_0=00:00, 06:00, 12:00, and 18:00 GMT, using 10 data points, the total of the five nearest points in time both before and after t_0, regardless of the nature of the data (GPS or Argos). For the error covariances, the values used for all observations are $\sigma^2_\Phi = 6.7\times10^{-6}$ for longitude and $\sigma^2_\Theta = 2.9\times10^{-6}$ for latitude, corresponding approximately to the standard deviation squared of location errors for fixed drifters observed by *Hansen and Herman* [1989]. *Hansen and Poulain* [1996] used a model of the structure function which is simply half the variogram function,

$$S(\tau) = \alpha|\tau|^\beta, \tag{3}$$

where τ is the time difference between location observations. Numerical values for the parameters α and β are obtained by fitting this model to empirical estimates of the structure function obtained from the QC nonuniform observations at intervals $\tau = 3$ and 6 days (see *Hansen and Poulain* [1996] for details). Thus, it is expected that the model is representative of oceanic processes at those time scales. In addition, because the results of Kriging were found to be relatively insensitive to α, its value is adjusted to make the Kriging variance error converge to an observed mean squared interpolation error which is obtained by a cross-validation procedure. For the current DAC implementation, the values of β have been calculated for six oceanic regions (North and South Atlantic, North and South Pacific, and Indian Ocean) and are kept constant with time, while the values of α are repetitively calculated for the same regions and in 6 month intervals (January to June and July to December).

The resulting product consists of 6-hourly longitude and latitude values along drifter trajectories, together with uncertainties taken as the square root of their respective Kriging variances. Drifter horizontal velocities are also provided, calculated at 6 h intervals by central differences. These data are available at http://www.aoml.noaa.gov/phod/dac/dacdata.php.

3.2.3. Revisiting Kriging for SPURS Data Set

We revisit Kriging based on the structure function to apply it to the edited QC Argos nonuniform locations in order to obtain hourly locations along the trajectories. We choose to use four data points for each interpolation time, the sum of two preceding and two following observations. Experimentation showed that using more observations for Kriging led to slightly better error statistics but to the detriment of velocity variance (discussed in the next section). For each of the observations used, we specify the longitude and latitude error variances as the squared standard deviations estimated from comparing Argos and GPS locations as a function of class (listed in Table 2). Finally, we use our own estimates of the parameters of model (3) by fitting the model to estimates of the structure function from the SPURS data for time differences shorter than 3 days, aiming to optimize Kriging for shorter time scales compared to the DAC implementation (see Appendix B).

3.3. Spherical Linear Interpolation

A relatively simple interpolation method is linear interpolation in time of both longitude and latitude coordinates. For this method, the underlying drifter trajectory model is a polynomial of order 1, with unknown intercept and slope which are determined using two data points located before and after the interpolating time. An immediate problem with interpolating longitude and latitude linearly, and separately, in time is that the resulting horizontal speed of the drifter is not constant, and therefore the kinetic energy of the particle is not conserved.

A more appropriate trajectory connecting two points in space is the path of the free, unforced particle—a geodesic. In 2-D or 3-D Euclidean space, this is simply a constant speed straight line, on the sphere, this is a constant speed great circle, and on the rotating earth, this is a constant speed inertial circle [see *Early*, 2012].

While the inertial circle may be the most appropriate trajectory for modeling ocean drifters, the great circle trajectory is more physically intuitive, easier to implement, and is nearly the same on the time scales considered here. Thus, we implement the constant speed great circle trajectory using a method which we call "spherical linear interpolation" that ensures that the interpolating path between two observations is along a great circle parameterized with constant angular speed, and therefore the horizontal speed of the drifters remains constant along the path. The method is due to *Shoemake* [1985] and involves the use of quaternions, which are hypercomplex numbers. Despite the unfamiliar aspect of quaternions, their simplicity of use for our current interpolating problem makes them attractive.

We quantify the differences between conducting linear interpolation along great circles with the quaternion method, and conducting the linear interpolation independently for longitude and latitude. For the SPURS drifter data set, the maximum difference between the two methods is 74 m, yet 96% of the differences are less than 1 m. We still prefer to conduct the linear interpolation on the sphere in order to ensure that the interpolated trajectory lies on a great circle, which can be important in the case of large interpolating gaps.

We next would like to obtain the variance of the longitude and latitude estimates for the spherical linear interpolation. For simplicity, we approximate this variance by the variance obtained by simply propagating errors using the formula for linear interpolation, and using the Argos class variances (Table 2). This variance is minimum at the exact mid-temporal distance pair points. Variance estimates are used to assess uncertainties for the interpolation method, arising from the observation errors.

3.4. Weighted Maximum Likelihood Estimation of Linear Trajectory Model With Prescribed Error PDF

Finally, we extend the linear interpolation by locally modeling the drifter trajectory as a linear function of time, but here we do not limit ourselves to using just one observation before and one observation after the interpolation time. The model around the interpolation time t_0 is a linear expansion

$$\phi(t;\boldsymbol{\beta}^{\phi}) = \beta_0^{\phi} + \beta_1^{\phi}(t - t_0), \tag{4}$$

$$\theta(t;\boldsymbol{\beta}^{\theta}) = \beta_0^{\theta} + \beta_1^{\theta}(t - t_0), \tag{5}$$

for which we seek that the set of parameters $\boldsymbol{\beta} = [\boldsymbol{\beta}^{\phi}, \boldsymbol{\beta}^{\theta}] = [\beta_0^{\phi}, \beta_1^{\phi}, \beta_0^{\theta}, \beta_1^{\theta}]$ maximizes a weighted probability of the observed data, or weighted likelihood,

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{N} \left\{ p\left[\Phi_k, \Theta_k, \phi(t_k;\boldsymbol{\beta}^{\phi}), \theta(t_k;\boldsymbol{\beta}^{\theta})\right] \right\}^{w_k}, \tag{6}$$

where p is the PDF describing whether a location observation (Φ_k, Θ_k) will yield the true location $(\phi(t_k), \theta(t_k))$. This function is raised to a power w_k to achieve a temporal weighting as shown below. The method of maximum likelihood is used to obtain values of model parameters that define PDFs that are most likely to have resulted in the observed data.

Maximizing the likelihood function (6) is equivalent to maximizing its natural logarithm, leading to a weighted log likelihood function:

$$l(\boldsymbol{\beta}) = \sum_{k=1}^{N} w_k \ln p\left[\Phi_k, \Theta_k, \phi\left(t_k;\boldsymbol{\beta}^{\phi}\right), \theta\left(t_k;\boldsymbol{\beta}^{\theta}\right)\right]. \tag{7}$$

The weighted maximum likelihood estimate (WMLE) $\widehat{\boldsymbol{\beta}}$ is such that

$$l(\widehat{\boldsymbol{\beta}}) = \max_{\boldsymbol{\beta}} \sum_{k=1}^{N} w_k \ln p\left[\Phi_k, \Theta_k, \phi\left(t_k;\boldsymbol{\beta}^{\phi}\right), \theta\left(t_k;\boldsymbol{\beta}^{\theta}\right)\right], \tag{8}$$

which means that $\widehat{\boldsymbol{\beta}}$ is the set of parameter values that maximizes (7).

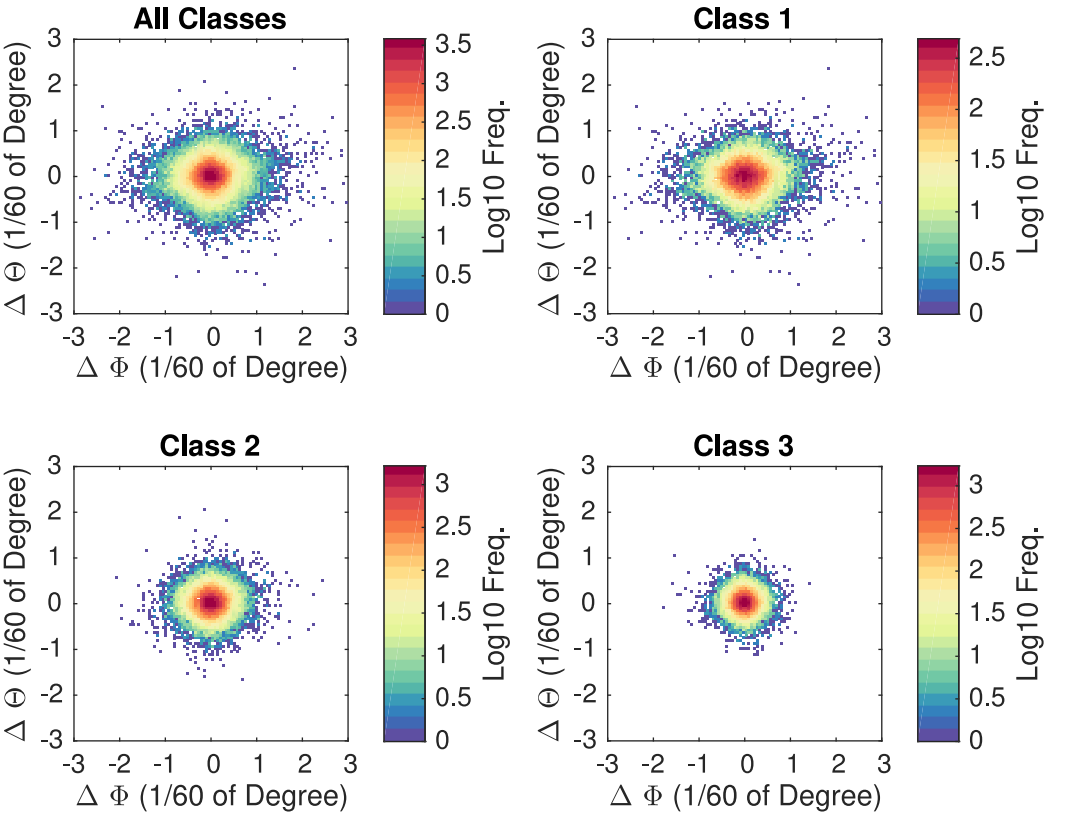

Figure 6. Two-dimensional histogram of Argos longitude errors ($\Delta\phi$) and Argos latitude errors ($\Delta\theta$) for all Argos location classes, and for each Argos class, in 10^{-3} degree bins. Absence of data is indicated by white shading. Note the different scales for each color bar.

From the assessment of Argos location errors using GPS data (section 2.3), we know that the probability distribution of Argos longitudes and latitudes can be described in term of the longitude differences $\delta\phi=\Phi_k-\phi(t_k;\boldsymbol{\beta}^\phi)$ and latitude differences $\delta\theta=\Theta_k-\theta(t_k;\boldsymbol{\beta}^\theta)$ between observations and true locations, which are here modeled using the linear expansion (4) about the interpolation time t_0. In addition, we see that the errors in longitude and latitude are independent (Figure 6), which implies that the joint distribution of longitude and latitude is the product of their marginal distributions, i.e.,

$$p\left[\Phi_k,\Theta_k,\phi\left(t_k;\boldsymbol{\beta}^\phi\right),\theta\left(t_k;\boldsymbol{\beta}^\theta\right)\right]=p\left[\Phi_k,\phi\left(t_k;\boldsymbol{\beta}^\phi\right)\right]p\left[\Theta_k,\theta\left(t_k;\boldsymbol{\beta}^\theta\right)\right]. \tag{9}$$

If the observations are normally distributed, we can substitute expression (1) into (7) with $z=\delta\phi$ and $z=\delta\theta$, and with (9), one obtains

$$\begin{aligned} l(\boldsymbol{\beta})=&\sum_{k=1}^{N}\left\{-w_k\ln\left(\sigma_{\Phi_k}\sqrt{2\pi}\right)-w_k\ln\left(\sigma_{\Theta_k}\sqrt{2\pi}\right)\right\}\\ &+\sum_{k=1}^{N}\left\{-\frac{w_k}{2}\frac{\left[\Phi_k-\phi(t_k;\boldsymbol{\beta}^\phi)-\mu_{\Phi_k}\right]^2}{\sigma_{\Phi_k}^2}-\frac{w_k}{2}\frac{\left[\Theta_k-\theta(t_k;\boldsymbol{\beta}^\theta)-\mu_{\Theta_k}\right]^2}{\sigma_{\Theta_k}^2}\right\}, \end{aligned} \tag{10}$$

where the $\mu_{\Phi_k},\mu_{\Theta_k}$ and $\sigma_{\Phi_k},\sigma_{\Theta_k}$ are the means and standard deviations of the normal PDFs of longitude or latitude observations with index *k*. Equation (10) is, in this case, the expression to maximize with respect to $\boldsymbol{\beta}$. While the first term of the right side is a constant, the second term can be maximized by varying $\boldsymbol{\beta}$, or equivalently, its negative can be minimized. Thus, it is apparent that the WMLE of $\boldsymbol{\beta}$ for normally distributed errors is a weighted least squares solution for $\boldsymbol{\beta}$.

As seen earlier, a better model than the normal PDF can be used for the Argos observations. In particular, observation outliers are more likely to occur than would be expected based on a normal PDF, and these data would then be given inadequate weights during the estimation [*Press et al.*, 1988, Chapter 17.1]. Instead of the normal PDF, we can use the *t* PDF model (2), for which the log likelihood becomes

$$l(\boldsymbol{\beta})=\sum_{k=1}^{N}\left\{-w_k\ln\left[\frac{\sigma_{\Phi_k}\sqrt{\nu_{\Phi_k}\pi}\,\Gamma\left(\frac{\nu_{\Phi_k}}{2}\right)}{\Gamma\left(\frac{\nu_{\Phi_k}+1}{2}\right)}\right]-w_k\ln\left[\frac{\sigma_{\Theta_k}\sqrt{\nu_{\Theta_k}\pi}\,\Gamma\left(\frac{\nu_{\Theta_k}}{2}\right)}{\Gamma\left(\frac{\nu_{\Theta_k}+1}{2}\right)}\right]\right\}$$
$$+\sum_{k=1}^{N}\left\{-w_k\frac{\nu_{\Phi_k}+1}{2}\ln\left[1+\frac{1}{\nu_{\Phi_k}}\left(\frac{\Phi_k-\phi(t_k;\boldsymbol{\beta}^{\phi})-\mu_{\Phi_k}}{\sigma_{\Phi_k}}\right)^2\right]-w_k\frac{\nu_{\Theta_k}+1}{2}\ln\left[1+\frac{1}{\nu_{\Theta_k}}\left(\frac{\Theta_k-\theta(t_k;\boldsymbol{\beta}^{\theta})-\mu_{\Theta_k}}{\sigma_{\Theta_k}}\right)^2\right]\right\}, \tag{11}$$

where $\mu_{\Phi_k}, \mu_{\Theta_k}, \sigma_{\Phi_k}, \sigma_{\Theta_k}$, and $\nu_{\Phi_k}, \nu_{\Theta_k}$ are the location, scale, and shape parameters of the *t* PDF of longitude or latitude observations with index *k*. As before, the first summation of (11) is a constant, but the second summation can be maximized with respect to $\boldsymbol{\beta}$.

We proceed to find the WMLE of $\boldsymbol{\beta}$ for model (4) at hourly time steps in order to interpolate the Argos longitude and latitude observations. We implement two versions of such a WMLE method, one assuming normal PDFs of the observations and one assuming *t* PDFs. Thus, finding the WMLE for the log likelihoods (10) and (11) consists in each case, and at each time step, of solving a set of four equations $[\partial l(\boldsymbol{\beta})/\partial\beta_j^{\phi,\theta}=0, j=0,1]$, or alternatively, of using a numerical algorithm of optimization to find a minimum of the negative of (10) or (11), a choice we make here because no analytical solution appears to exist for the *t* PDF case. We use an unconstrained minimization algorithm—the trust-region algorithm (as implemented by the function fminunc of the optimization toolbox of MATLAB)—for which we provide the analytical expression of the gradients of (10) or (11). As initial values for the algorithm, we use the nonweighted least squares solution of fitting the *N* data points to the linear model (4).

For the choice of the number of observations used to form the log likelihood functions, experimentation leads us to choose two location data points before and two after the hourly times (i.e., $N=4$), permitting also a fair comparison with Kriging, which uses the same number of observations. For each estimation, the values of the parameters of the PDFs of the observations are dictated by their classes, and we use the values from our Argos class analyses (section 2.3 and Tables 2 and 3). Regarding the choice of temporal weights w_k, experimentation reveals that for a uniform weighting scheme, the WMLE is nearly constant for a given set of *N* observations. Thus, in order for the solution to gradually evolve with hourly times, we make the weights w_k a function of absolute temporal distance $|t_k-t_0|$ from the interpolating time t_0. For this, we choose an inverse distance weighting scheme as

$$w_k=w(|t_k-t_0|)=\frac{C}{t_{00}+|t_k-t_0|}, \tag{12}$$

where t_{00} is a fixed parameter and *C* is a normalizing constant so that $\sum_{k=1}^{N} w_k=N$. This normalization recovers a nonweighted version of the log likelihood (7) in the case $w_k\equiv 1$ for all *k*. The constant t_{00} is set equal to 20 min to prevent the weights from reaching infinite value as t_k approaches t_0. This time interval is consistent with the Argos data selection algorithm described in section 3.1.

To obtain confidence intervals for the WMLE, we could use asymptotic theory, which tells us that as the number of observation data points *N* becomes large, the difference between the true parameters and the WMLE ($\widehat{\beta}_j-\beta_j, j=1,2$) is normally distributed with a variance that can be computed from the curvature of the log likelihood function at its maximum [*Dzhaparidze and Kotz*, 1986]. Yet we have at our disposal few observations (i.e., $N=4$) to obtain our estimates, so we rely instead on a bootstrapping method [*Efron and Gong*, 1983]. We reestimate $\boldsymbol{\beta}$ at each hourly time step by using three observations out of four, which leads to four alternative WMLEs from which we compute the variances $s_j=\mathrm{Var}[\widehat{\beta}_j]$. As we will show, we find that the bootstrap method performs well, and thus the resulting *t* value $t_j=\frac{\widehat{\beta}_j-\beta_j}{s_j/\sqrt{N}}$ should be distributed as a Student's *t* PDF with $N-1$ degrees of freedom. We therefore derive 95% confidence intervals, at each hourly time step for locations and velocities, as

$$\widehat{\beta}_j\pm t(0.025,N-1)\frac{s}{\sqrt{N}}, \tag{13}$$

where $t(0.025,N-1)=3.1824$ is the value for the 2.5th percentile of the Student's *t* cumulative distribution function with *N* – 1 degrees of freedom.

4. Assessment of Methods

4.1. Locations

In this section, we present an overall comparison of the four different interpolation methods: updated Kriging (section 3.2), spherical linear interpolation (section 3.3), the WMLE method with normal PDF, and the WMLE method with *t* PDF (section 3.4). For all methods, we define interpolation errors as the difference between the smoothed GPS locations interpolated linearly to hourly time steps, our ground truth, and the interpolated positions (Figure 2). We consider only 82% of the calculated errors (347,583 data points), hereafter called "verifiable," for which a smoothed GPS location is available within an hour before, and within an hour after, the interpolating time.

Kernel estimates of the distribution of the longitude differences ($\Delta\phi$), latitude differences ($\Delta\theta$), and angular separation ($\Delta\lambda$) are shown in Figure 7. Also shown are the curves corresponding to fitting these distribution to PDF models, and the fitting parameters are listed in Tables 6–8. These results allow us to answer the following two questions. How do the interpolation errors compare to the Argos errors (for noninterpolated locations)? Which method returns the smallest errors?

4.1.1. How do the Interpolation Errors Compare to the Argos Errors (for Noninterpolated Locations)?

All methods return a mode value for $\Delta\lambda$ which is smaller than for Argos errors for all classes (by 23 or 33 m) (Tables 5 and 6), which suggests that interpolation reduce errors by incorporating information from multiple observations. The 50th and 68th percentiles are also smaller after interpolation, except the 68th percentile for the spherical linear interpolation. In contrast, the 95th percentile is dramatically increased (by hundreds of meters) after interpolation. The fitted scales of exponential PDFs are also significantly larger after interpolation.

For the mean coordinates errors (μ, Table 7), interpolating does not qualitatively change what was found for nonuniform Argos locations for all classes (μ_{Φ_k} and μ_{Θ_k} in Table 2): the mean longitude errors are less than 10 m, but the latitude errors are comparatively larger. All methods, however, increase the standard deviations of longitude and latitude errors by more than 100 m (Tables 2 and 6). For each method, the probability that the errors will fall within one standard deviation of the mean error are also increased compared to Argos (all classes), and therefore interpolating exacerbates the nonnormal character of the error distributions. This is further confirmed by alternatively examining the *t* PDF descriptions of the errors (Tables 3 and 8). We find that all interpolation methods increase the importance of the tails of the error distributions, and narrow the central peaks; the shape parameter ν estimates are all significantly smaller, and the scale parameter σ estimates are also all significantly smaller, compared to the corresponding parameters for the nonuniform Argos errors. The only exception is for the spherical linear interpolation method for the scale parameter for longitude errors, but the value fitted (equivalent to 304 m) is within a few meters of the corresponding Argos all classes value (297 m for all classes). The overall impact of the changing of the shapes of the distribution can be measured for each method by evaluating the error values that comprise 68% of the errors based on the *t* PDF models. Compared to the same value for Argos (all classes), Kriging reduces these values for both coordinates errors (equivalent 329 and 294 m for longitude and latitude, respectively, compared to 346 and 335 m before interpolation) which is not surprising since Kriging is a smoother and thus should reduce errors by averaging. All other methods increase this value for longitude errors but decrease this value for latitude errors, albeit by moderate amounts (no more than 35 m).

4.1.2. Which Method Returns the Smallest Errors?

The two WMLE methods exhibit the smallest mode value for $\Delta\lambda$, 201 m (Table 6). The Kriging and spherical linear methods have a mode value for $\Delta\lambda$ of 211 m. We can reduce further this value to 201 m by increasing the number of observations used for each hourly interpolation, but doing so is to the detriment of the high-frequency variance, so it is not favored (see section 3.2.2). Yet we find that Kriging returns the smallest values for typical percentiles (listed in Table 6). For the 50th and 68th percentiles, the WMLE method with *t* location-scale PDF comes second best, followed by the WMLE method with normal PDF, and the spherical linear method comes last. The sample means of the $\Delta\lambda$, which are also estimates of the scale parameter of an exponential PDF model, place Kriging first with a distance equivalence of 490 m, the spherical linear and WMLE with *t* PDF second with 545 m, and the WMLE with normal PDF last with 552 m. These values are statistically different according to the 95% confidence intervals of the ML estimates (not listed).

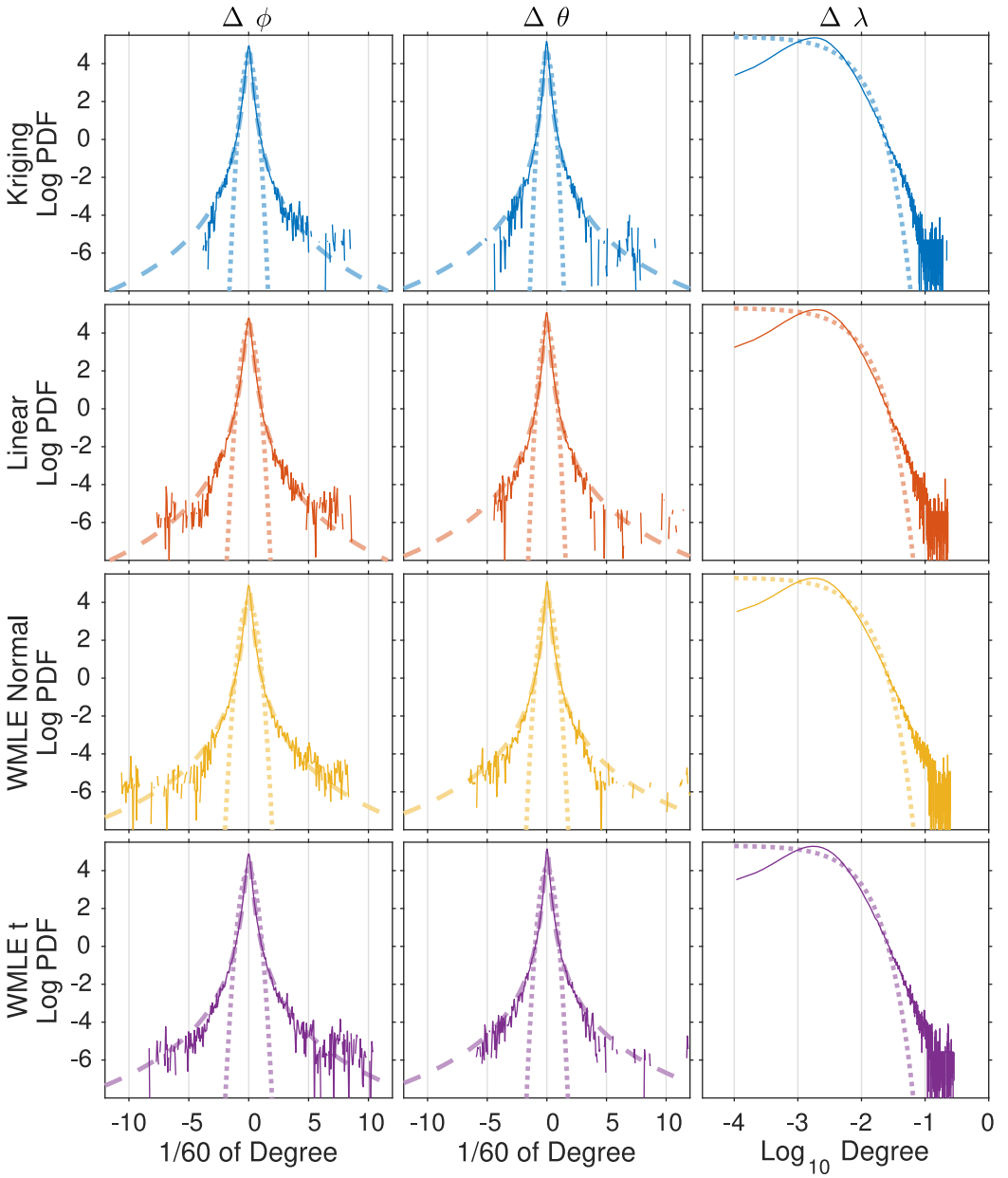

Figure 7. (left column) Longitude error ($\Delta\phi$), (middle column) latitude error ($\Delta\theta$), and (right column) angular separation ($\Delta\lambda$) error distributions for each of the methods of interpolation, from top to bottom: Kriging, spherical linear, WMLE with Normal PDF and WMLE with *t* location-scale PDF. In each plot, the thin solid curve is a kernel estimate. For longitude and latitude errors, the dotted heavy curve is a fit to a normal PDF, and the dashed heavy curve is a fit to a *t* location-scale PDF. For the $\Delta\lambda$ distributions, the dotted lines are fit to exponential PDFs. The parameters of the fits are listed in Tables 7 and 8.

We now consider the distributions of the longitude errors ($\Delta\phi$) and latitude errors ($\Delta\theta$), and their statistics corresponding to a normal PDF model (Figure 7 and Table 7). In terms of returning the smallest standard deviation values for both coordinate's errors, Kriging comes first, the spherical linear method comes second, and the WMLE method with *t* PDF and with normal PDF come third and fourth, respectively. We compute next the observed probabilities of $\Delta\phi$ and $\Delta\theta$ to fall within ±1 and ±1.96 standard deviations from their means. All methods return similar probabilities, larger than 84%, much more than for normally distributed data (68%). However, all methods return similar probabilities of falling within ±1.96 standard deviation, close to the expectation for normally distributed data (95.8–96.6% compared to 95%). Thus, the errors after interpolation are still nonnormal, and we alternatively describe and compare them by fitting them to *t* PDF models (Figure 7 and Table 8). It is not straightforward to compare these next results because we have now

Table 6. Angular Separation Errors $\Delta\lambda$ Statistics for Each of the Four Interpolation Methods

$\Delta\lambda$	Kriging	Spherical Linear	WMLE With Normal PDF	WMLE With *t* PDF
10^2 mode	0.190° (211 m)	0.190° (211 m)	0.181° (201 m)	0.181° (201 m)
10^2 50th percentile	0.303° (337 m)	0.343° (381 m)	0.325° (362 m)	0.322° (358 m)
10^2 68th percentile	0.441° (490 m)	0.499° (555 m)	0.483° (537 m)	0.479° (533 m)
10^2 95th percentile	1.233° (1371 m)	1.340° (1490 m)	1.416° (1575 m)	1.397° (1553 m)
Exp. Fit: $10^2\times\sigma$ (sample mean)	0.440° (490 m)	0.490° (545 m)	0.497° (552 m)	0.490° (545 m)

Table 7. Results of Fits of Longitude and Latitude Errors of Each of the Four Interpolation Methods to Normal PDFs[a]

	Kriging	Spherical Linear	WMLE With Normal PDF	WMLE With *t* PDF
Lon. Err. $\Delta\phi$				
$10^2\times\mu$ (sample mean)	−0.003° (−3 m)	−0.003° (−4 m)	−0.006° (−6 m)	−0.005° (−5 m)
$10^2\times\sigma$ (sample std)	0.543° (549 m)	0.618° (624 m)	0.670° (677 m)	0.663° (670 m)
Prob. within $\mu\pm\sigma$ (68%)	84.2%	84.1%	87.0%	86.7%
Prob. within $\mu\pm1.96\sigma$ (95%)	95.8%	96.0%	96.5%	96.4%
Lat. Err. $\Delta\theta$				
$10^2\times\mu$ (sample mean)	0.003° (4 m)	0.004° (5 m)	0.050° (55 m)	0.051° (56 m)
$10^2\times\sigma$ (sample std)	0.479° (532 m)	0.528° (587 m)	0.596° (662 m)	0.587° (653 m)
Prob. within $\mu\pm\sigma$ (68%)	85.4%	86.1%	87.5%	88.1%
Prob. within $\mu\pm1.96\sigma$ (95%)	95.8%	96.3%	96.3%	96.6%

[a]Longitude parameters are converted to distance using the 24.7°N median latitude of the data. The observed probability of the errors to fall within ±1 and ±1.96 the standard deviation from the mean are also listed, which should be 68% and 95%, respectively, for a normal PDF.

two parameters (σ and ν) to compare between four methods. Thus, we calculate from the analytical expression of the *t* PDF the error values ($\delta\phi$ and $\delta\theta$ in Table 8) for each method that would comprise 68% and 95% of the distributions. For the 68% interval for $\Delta\phi$, Kriging comes first by exhibiting the smallest value, the WMLE with normal PDF comes second, the WMLE with *t* PDF comes third, and the spherical linear method comes last. The order differs for the 68% interval for $\Delta\theta$: Kriging comes first again but the WMLE with *t* PDF comes second, the spherical linear comes third, and the WMLE with normal PDF comes last.

In summary, we assess that Kriging performs the best in terms of location errors, and that the WMLE methods generally perform second best.

4.2. Velocity

In this section, we assess velocities obtained from each of the four interpolation methods. The spherical linear and WMLE methods return direct estimates of velocities. In contrast, Kriging estimates locations only, hence we calculate velocities by central differences of the kriged locations. We compare the hourly

Table 8. Results of Fits of Longitude and Latitude Errors of Each of the Four Interpolation Methods to *t* Location-Scale PDFs[a]

	Kriging	Spherical Linear	WMLE With Normal PDF	WMLE With *t* PDF
Lon. Err. $\Delta\phi$				
$10^2\times\mu$	0.000° (0 m)	−0.002° (−2 m)	−0.005° (−5 m)	−0.005° (−5 m)
$10^2\times\sigma$	0.257° (259 m)	0.301° (304 m)	0.268° (270 m)	0.270° (273 m)
ν	2.32	2.44	2.11	2.11
Prob $(\delta\phi < \Delta\phi-\mu < \delta\phi)=0.68$				
$10^2\times\delta\phi$	0.3260° (329 m)	0.370° (381 m)	0.3480° (352 m)	0.3510° (355 m)
Prob $(\delta\phi < \Delta\phi-\mu < \delta\phi)=0.95$				
$10^2\times\delta\phi$	0.976° (986 m)	1.098° (1109 m)	1.099° (1110 m)	1.1060° (1117 m)
Lat. Err. $\Delta\theta$				
$10^2\times\mu$	−0.015° (−17 m)	−0.013° (−14 m)	0.035° (39 m)	0.035° (39 m)
$10^2\times\sigma$	0.202° (224 m)	0.219° (243 m)	0.215° (239 m)	0.207° (230 m)
ν	2.08	2.11	1.90	1.91
Prob $(\delta\theta < \Delta\theta-\mu < \delta\theta)=0.68$				
$10^2\times\delta\theta$	0.264° (294 m)	0.285° (317 m)	0.289° (321 m)	0.278° (309 m)
Prob $(\delta\theta < \Delta\theta-\mu < \delta\theta)=0.95$				
$10^2\times\delta\theta$	0.839° (933 m)	0.901° (1002 m)	0.9730° (1082 m)	0.936° (1041 m)

[a]Longitude parameters are converted to distance using the 24.7°N median latitude of the data. Also listed are the calculated error values around the mean that define 68% and 95% of the fitted *t* location-scale PDFs.

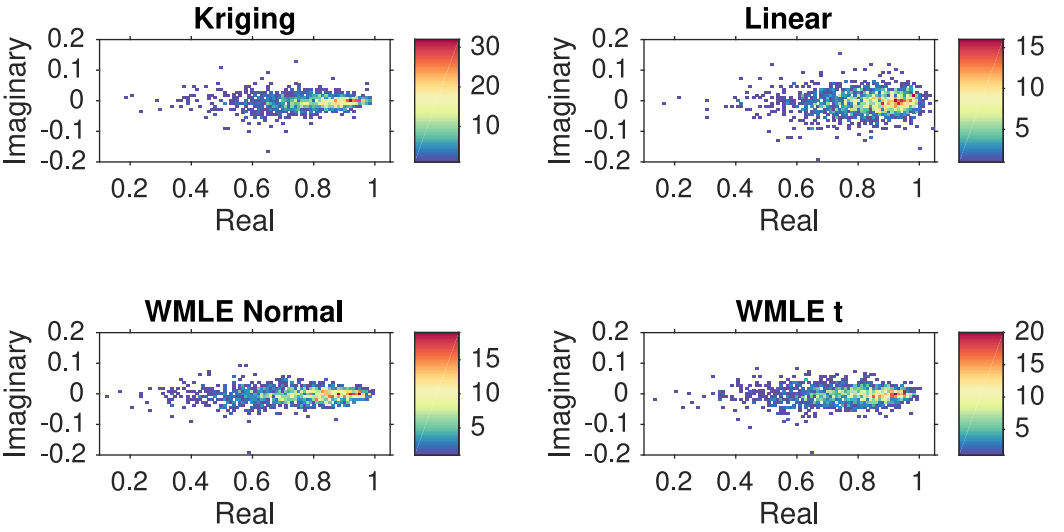

Figure 8. Two-dimensional histograms (real versus imaginary parts) of complex regression coefficient estimates between GPS-derived drifter velocities and drifter velocities, derived by each of the four interpolation methods: Kriging, spherical linear, WMLE with normal and with *t* location-scale distribution.

velocities from the interpolation methods to the GPS-derived velocities after linearly interpolating them onto hourly times (Figure 2).

4.2.1. Vector Cross Covariance

We consider the complex-valued time series $z_1(t)=u_1(t)+iv_1(t)$ where $u_1(t)$ and $v_1(t)$ are the zonal and meridional components of the GPS-derived velocities, respectively, and the velocity time series derived from any of the interpolation methods, hereafter $z_2(t)=u_2(t)+iv_2(t)$, where $u_2(t)$ and $v_2(t)$ are the zonal and meridional components. Let $z^m(t)$ be the result of the complex linear regression of $z_2(t)$ onto $z_1(t)$,

$$z^m(t)=\frac{R_{12}(0)}{R_{11}(0)}z_1(t)=\frac{|R_{12}(0)|}{R_{11}(0)}e^{i\mathrm{Arg}[R_{12}(0)]}z_1(t), \tag{14}$$

where $R_{12}(0)=E[z_1^*z_2]=E[(u_1-iv_1)(u_2+iv_2)]$ is the complex-valued cross covariance at zero lag, and $R_{11}(0)$ the autocovariance of z_1. The ratio $|R_{12}(0)|/R_{11}(0)$ indicates if $z_2(t)$ underestimates or overestimates the magnitude of $z_1(t)$. $\mathrm{Arg}[R_{12}(0)]$ indicates an average angular offset of $z_2(t)$ compared to the orientation of $z_1(t)$. A perfect interpolation method would return a regression coefficient with real component 1 and imaginary component 0. To focus on relatively short time scales of variance, we calculate the regression coefficients for 1724 10-day nonoverlapping segment trajectories, keeping all data, verifiable or not. We then compute 2-D histograms of the regression coefficients in the complex plane (Figure 8).

Kriging returns the most favorable distribution of regression coefficients with the smallest spread along both real and imaginary axes. The coefficient distributions for the WMLE methods indicate that these methods return velocities which are better aligned with the GPS velocities compared to the spherical linear method (less spread along the imaginary axis for the 2-D distribution) even if these methods appear to underestimate the GPS velocity magnitude (more spread along the real axis toward zero). Between the WMLE methods, the method with normal PDF returns velocities which are better aligned with GPS velocities than the method with *t* PDF, but underestimates more often the velocity magnitude. These results do not qualitatively change when the calculations are conducted on time series five days or 20 days long.

4.2.2. Rotary Spectral Analyses

We now assess the ability of each method to represent velocity variance as a function of frequency by comparing spectral analyses between the hourly velocities derived from the QC GPS locations, and the velocities derived from the four interpolation methods. We divide the drifter velocity time series into 395 40-day nonoverlapping segments, keeping all data points. To reduce spectral leakage, we multiply each time series segment by a single Slepian taper with a time-bandwidth product equal to 4 [*Slepian*, 1978], calculate the Fourier transform, and average across all estimates to obtain rotary autospectra (Figure 9) and cross-spectra estimates from which we also compute coherence squared (Figure 10), and coherence phase (not shown).

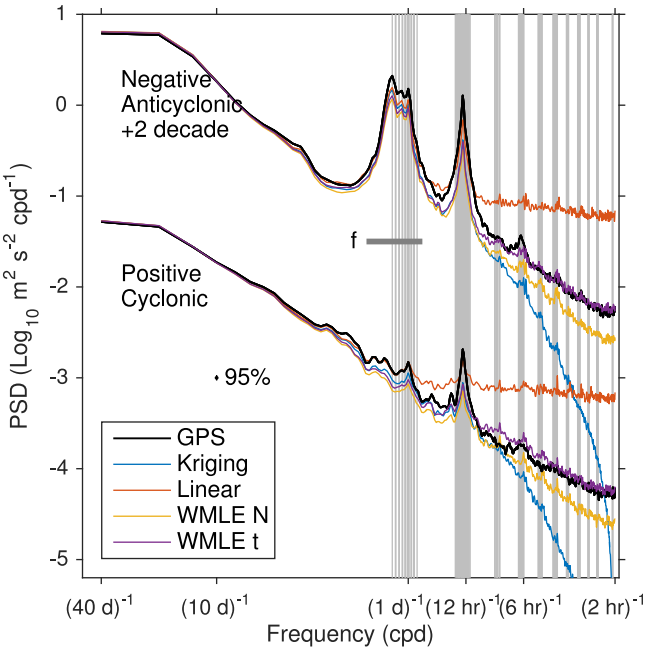

Figure 9. Rotary velocity spectra estimates for GPS-derived velocities and velocities derived from each of the four interpolation methods as indicated in the legend. The negative, anticyclonic, spectra are offset vertically by two decades for legibility. The thin vertical gray lines correspond to tidal frequencies. The horizontal bar marked by *f* indicates the range of inertial frequencies from the latitudinal distribution of the data.

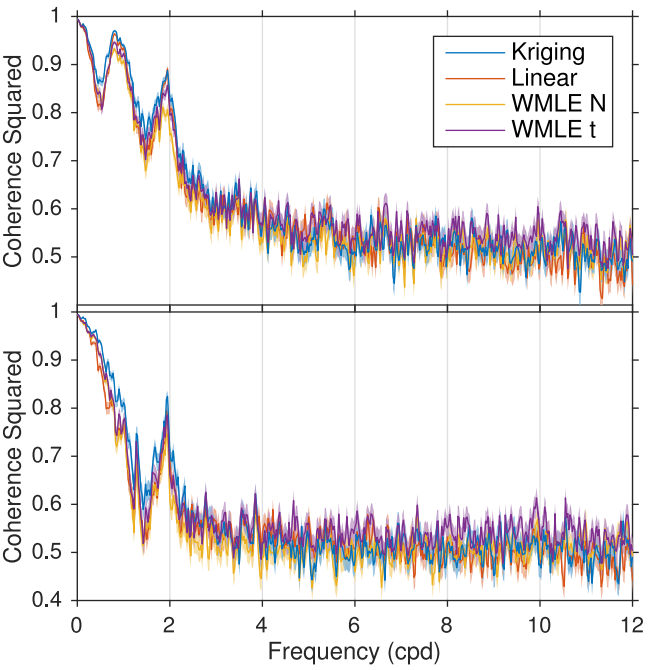

Figure 10. Rotary coherence squared estimates for (top) negative anticyclonic frequencies and (bottom) positive cyclonic frequencies between GPS-derived velocities and velocities derived from each of the four interpolation methods, as indicated in the legend. The shadings indicate plus or minus one approximated standard deviation of the estimates following *Carter* [1987].

Here the coherence squared indicates the fraction of the variance of the GPS-derived velocity that the estimated velocity by each of the methods is able to capture, as a function of frequency. The coherence phase indicates any geometrical or temporal angular difference between velocities, also as a function of frequency. The standard deviations of the coherence squared estimates are approximated using the expression for a large number of segments given by *Carter* [1987].

Compared to the GPS velocity spectra in both cyclonic and anticyclonic domains, all methods accurately capture the power level at relatively low frequencies (frequencies less than 1/10 cpd), but underestimate the power at intermediate frequencies between 1/10 cpd and 1/12 cph (Figure 9). At frequencies higher than approximately 1/12 cph, the velocity spectra from the spherical linear interpolation become dramatically higher and flatter than the GPS spectra, showing that this method introduces spurious noise at these frequencies. In the same high frequency range, Kriging dramatically underestimate the GPS spectra, which is expected since Kriging velocities are calculated by central differences (see also Figure 3). In contrast, the WMLE methods reproduce approximately the slopes and power levels of the GPS spectra, with the WMLE method with *t* PDF performing the best. Note that the ability of the WMLE methods to reach a power level similar to the GPS velocities is achieved by selecting an appropriate number of data points for the method (using more than 2 data points on each side of interpolating times reduces the velocity variance at high frequencies).

For frequencies less than 1/10 cpd, all methods exhibit near-perfect coherence squared at levels of 0.98 and above (Figure 10), which is not surprising considering the relatively short Argos sampling interval for this data set. At intermediate frequencies, from 0.1 to between 2 and 4 cpd, the

coherence squared is generally higher for anticyclonic frequencies than for cyclonic frequencies, and the shape of the coherence squared follows roughly the shape of the rotary spectra, decreasing nonmonotonically from 0.98 to about 0.5, likely a result of a decreasing signal-to-noise ratio. Between 0.1 and 2–3 cpd, the coherence squared is slightly significantly higher for Kriging compared to other methods, likely because Kriging relies on a semivariogram model (and hence a model for velocities) specifically fitted to the data. The other methods fare approximately equivalently in this range of frequencies. Toward frequencies higher than 4 cpd, the WMLE method with *t* PDF generally exhibits the highest coherence squared, between 0.5 and 0.6, and often significantly. The coherence phase (not shown) reveals that none of the methods introduce distortions of the phase or temporal delays, at least up to frequency 4 cpd. At higher frequencies, the phase estimates are noisy and unreliable due to the relatively low coherence.

In summary, the WMLE method with *t* PDF exhibits the best performances toward high frequencies while the kriging method Kriging fares better at low frequencies.

4.3. Analyses of Predicted Variances

4.3.1. Location Variance

All four interpolation methods return location uncertainty, or error, in the form of estimated (predicted) variances for the interpolated coordinates. In this section, we quantify and evaluate the predicted variances to assess their usefulness. By considering the GPS locations minus the interpolated locations from each of the methods, we have calculated observed longitude and latitude errors at each interpolating time, continuously along trajectories. We consider 3366 10-day trajectory segments with 50% overlap in order to calculate the variance of the observed longitude and latitude error time series for each segment, which we compare to the mean of the estimated variances for longitude and latitude, from each method along these segments (Figure 11). For each calculation, we exclude the nonverifiable locations.

The results indicate that over the range of densest observed error variances, the Kriging variance almost always overestimates the true error variance. In contrast, over most observed values, both WMLE methods most often adequately estimate the error variance. Only toward the largest observed values do the WMLE methods underestimate the variance. The predicted variance from the spherical linear interpolation is not useful as it is independent of the observed variance. By design, the predicted variance from the spherical linear interpolation takes a limited range of values set by the linear combination of the error variance of each class.

In conclusion, only location variances obtained by bootstrapping for the WMLE methods are adequate measures of the true error variance for longitude and latitude.

4.3.2. Velocity Variance

For the WMLE methods, horizontal drifter velocity components are calculated from the estimates of the coordinates and their derivatives as

$$\widehat{u} = R\cos\left(\widehat{\theta}\right)\widehat{\frac{d\phi}{dt}} = R\cos\left(\widehat{\beta_0^\theta}\right)\widehat{\beta_1^\phi}, \tag{15}$$

$$\widehat{v} = R\widehat{\frac{d\theta}{dt}} = R\widehat{\beta_1^\theta}. \tag{16}$$

Thus, to calculate the predicted variance of $\widehat{v}$ is straightforward, but to calculate the predicted variance of $\widehat{u}$, we use a Taylor approximation, and neglect the covariance between the parameter estimates $\widehat{\beta_0^\theta}$ and $\widehat{\beta_1^\phi}$. We therefore obtain

$$\mathrm{Var}[\widehat{u}] \approx R^2\sin^2\left(\widehat{\beta_0^\theta}\right)\left(\beta_1^\phi\right)^2\mathrm{Var}[\widehat{\beta_0^\theta}] + R^2\cos^2\left(\widehat{\beta_0^\theta}\right)\mathrm{Var}[\widehat{\beta_1^\phi}], \tag{17}$$

$$\mathrm{Var}[\widehat{v}] = R^2\mathrm{Var}[\widehat{\beta_1^\theta}]. \tag{18}$$

We consider again 3366 10-day trajectory segments with 50% overlap as previously, and calculate the variance of the observed zonal and meridional velocity component errors. We compare these values to the mean of the predicted variances, $\mathrm{Var}[\widehat{u}]$ and $\mathrm{Var}[\widehat{v}]$, for the two WMLE methods. For each calculation, we exclude the nonverifiable locations (Figure 12). The results indicate that both WMLE methods typically underestimate the square root of observed variances for both velocity components, by typically 25%–50%,

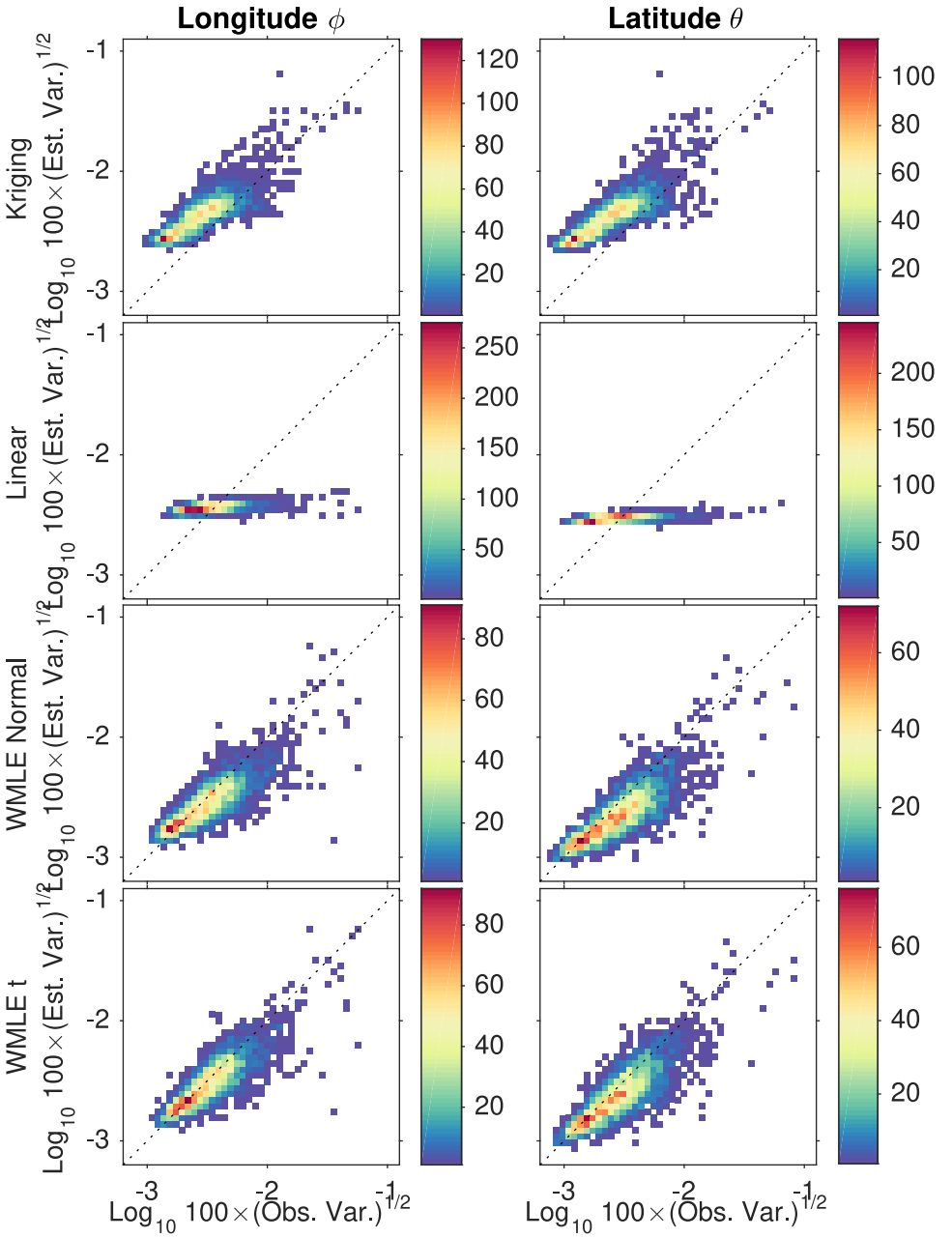

Figure 11. Two-dimensional histograms of the square root of estimated versus observed error variance for (left) longitude and (right) latitude for each method (rows). Note the different scales for each color bar. For reference, the dotted black lines are regression curves with slope 1 and intercept 0. The scales on these plots are such that −3 represents approximately 100 m error and −2 approximately 1 km error.

and up to 75% (and thus the variance by typically 44%–75%, up to 94%). The WMLE method with *t* PDF usually performs better by underestimating less often the observed velocity components variances.

5. Global Application to the Argos Data Set

For the interpolation of the global Argos data set, we choose to apply the WMLE interpolation method assuming a *t* PDF of the Argos locations. This choice is motivated by several factors. This method performs as well and sometimes better than the other methods when estimating locations. It performs the best in terms of velocity spectral level, and the estimated velocities exhibit coherence with GPS-derived velocities which is the best at high frequencies and almost as good as the best method at lower frequencies. It returns

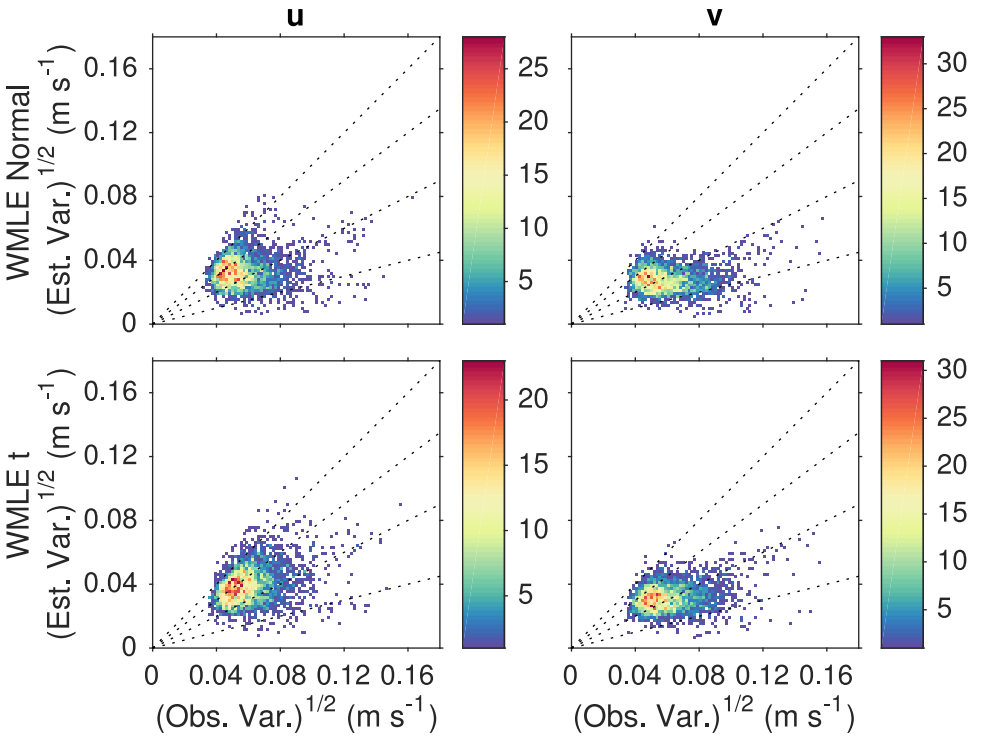

Figure 12. Two-dimensional histograms of the square root of estimated versus observed error variance for (left) zonal velocity and (right) meridional velocity for the two WMLE method (rows). Note the different scales for each color bar. For reference, the dotted black lines are regression curves with slope 1, 0.75, 0.5, and 0.25, and intercept 0.

a predicted variance of the location estimates which is commensurate with the observed error variance, and the best predicted variance of velocity components estimates of all methods, albeit an underestimation. Finally, we favor that this method relies on observed statistics of the data, rather than on assumptions about the underlying dynamics leading to the observations, as Kriging does by assuming a model for the structure function.

5.1. The Global Drifter Data Set Since 2005

Figure 13 shows the spatial density of 69,487,583 nonuniform Argos locations from 11,849 drifter trajectories since 1 January 2005 on a 1° × 1° grid. The geographical distribution is the result of the history of deployments, the efforts by the GDP to coordinate a regular and global sampling, the lifespan of drifters [*Lumpkin et al*., 2012], as well as the patterns of near-surface convergence and divergence zones associated with the oceanic general circulation and locally wind-driven currents.

Figure 14 displays combined illustrations of the distribution of temporal separation and distance separation between consecutive QC nonuniform Argos locations along drifter trajectories. The latitudinal and temporal distributions of those separations are mostly a consequence of the geometry and orbital characteristics of the Argos satellite array. As an example, 15 min is approximately the length of an Argos satellite pass during which a single location is captured, and 101.47 min is the orbital period of a single satellite. A drifter seen by a single satellite is likely seen again by the same satellite 101.47 min later and this explains the peak in the temporal separation distribution (Figure 14b). The

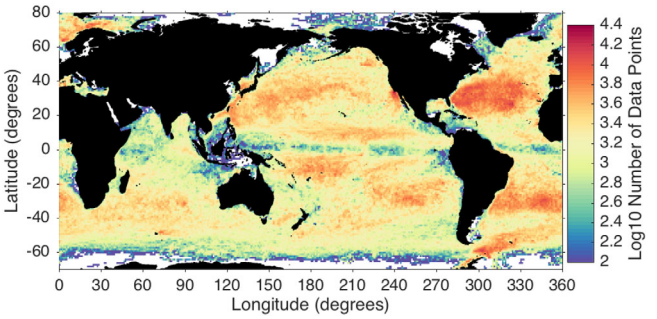

Figure 13. Geographical distribution of the QC nonuniform Argos drifter data between 1 January 2005 and 30 June 2015.

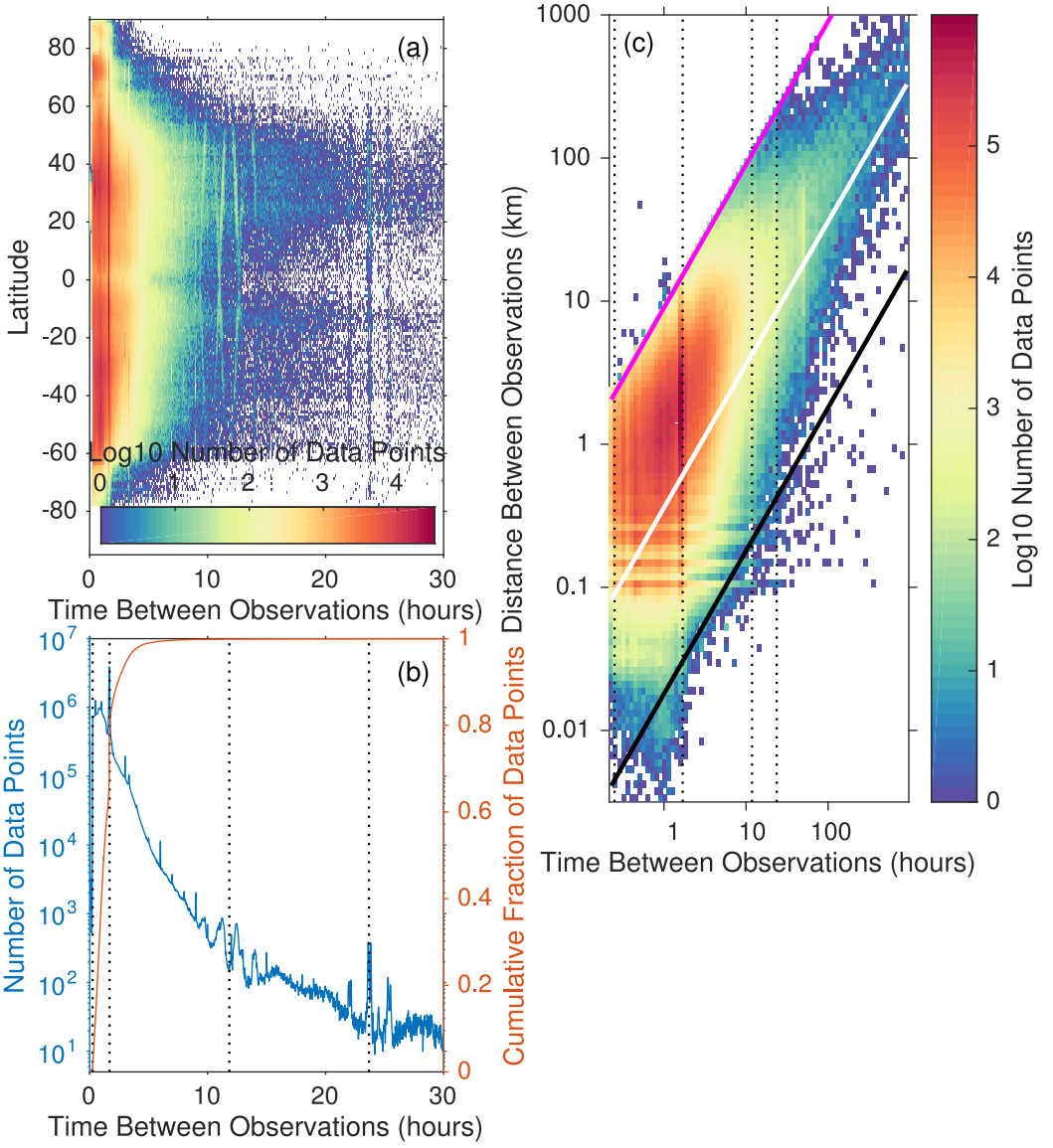

Figure 14. (a) Joint histogram of mean latitude and time separation of two consecutive QC nonuniform Argos locations along drifter trajectories between 1 January 2005 and 30 June 2015. (b) Distribution of time intervals between consecutive locations (left axis) and cumulative fraction distribution (right axis). The black vertical lines correspond to 15, 101.47, 7 × 101.47, and 14 × 101.47 min, where 101.47 min is the orbital period of polar orbiting Argos satellites. Fifteen minutes is the approximate length of a satellite pass. (c) Joint histogram of distance separation and time separation of two consecutive Argos locations along drifter trajectories. The magenta, white, and black lines correspond to 2.5, 0.1, and 0.005 m s^{-1}, respectively.

mean value of the temporal separation is 1.4 h, and the median value is 1.2 h. The joint distribution of temporal separation and distance separation suggests that there is a 2.5 m s^{-1} cap imposed on Argos platform velocities (Figure 14c).

5.2. Interpolation Implementation

We apply the WMLE interpolation method with *t* PDF to the global QC nonuniform Argos drifter data since 1 January 2005. Before running the interpolation on each individual trajectory, we reduce the Argos nonuniform location time series in 20 min windows in the same way as explained in section 3.1. At the time of this current study, the QC data set extends to 30 June 2015. In the historical data set, some drifter locations are of Argos location class 0, and for various reasons, a number of drifters do not have a record of Argos location class. For these observations, we assume that they are distributed like the overall distribution of all Argos locations for the SPURS data set (all classes in Table 2), and use those corresponding parameters of

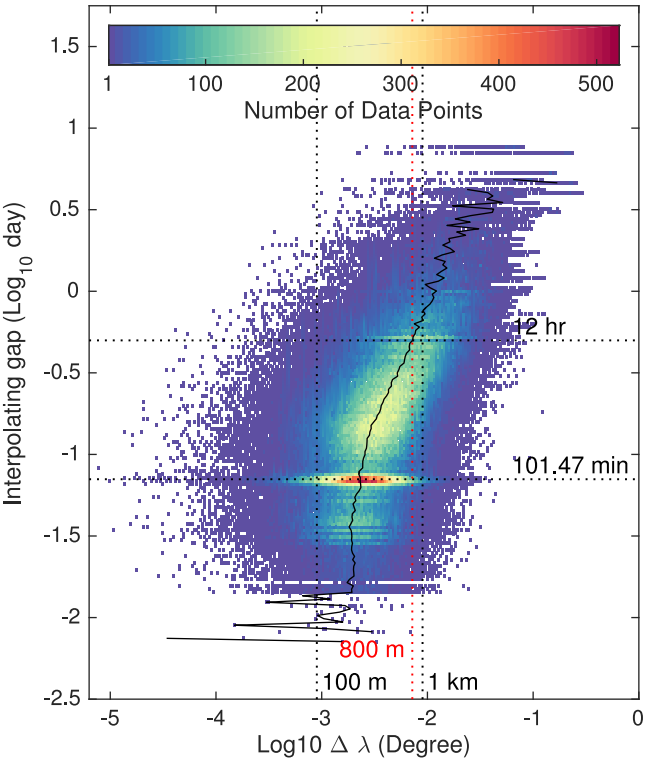

Figure 15. Two-dimensional histogram on decimal logarithmic scales of angular separation error ($\Delta\lambda$) and length of interpolating gap. The errors presented are for the WMLE method with *t* location-scale PDF for the SPURS drifter data. Horizontal dotted lines indicate 12 h and 101.47 min gaps. Vertical dotted lines indicate 100 m, 800 m, and 1 km errors. The black curve indicates the fiftieth percentiles of the error distributions for each interpolating gap bin. For 12 h gaps, the fiftieth percentile is at about 800 m.

the *t* PDF to compute the likelihood functions. We provide with this data set two pieces of information from the interpolation procedure. The first is the size of the interpolating gap, i.e., the difference between the closest posterior and anterior Argos locations used for the interpolation. The second piece of information is the square root of the mean squared temporal distances from the hourly interpolated location to the four adjacent Argos locations used for the interpolation. Note that at the beginning and end of trajectories only three locations are used for the interpolation. These pieces of information should be considered by users in order to assess the validity of the interpolation.

As an example of the impact of the interpolating gap, Figure 15 shows for the SPURS data the distribution of angular separation error for the WMLE method with *t* PDF as a function of the size of the interpolating gap. The error distribution is fairly unchanged for gaps shorter than 101.47 min (Argos orbital period), but progressively shifts toward larger values for increasingly longer gaps. In particular, we find that for gaps of 12 h and longer, the fiftieth percentile of the angular separation error is at least 800 m. This error value corresponds approximately to the 95th percentile of observed standard deviation of errors in both the zonal and meridional directions (see Figure 12). Thus, we choose to not interpolate trajectories for gaps longer than 12 h for the global data set.

We thus create a global drifter data set of locations and velocities at hourly intervals, along with their associated variance errors from 2005 to 2015, now freely available through the DAC website at http://www.aoml.noaa.gov/phod/dac/dacdata.php. Excluding drifters tracked by GPS, this data set consists of 11,670 individual trajectories, totaling more than 95.8 million on-the-hour estimates. For the GPS drifters of the GDP data set, we apply the LOWESS filter described in section 2.2 followed by linear interpolation to hourly times. Since 2005, 408 drifters were tracked by GPS, and the resulting interpolated data set consists of 2.3 million on-the-hour locations and velocities. Of these, 136 were also tracked by Argos and both types of trajectories are available in the new database.

5.3. Global Velocity Variance Analysis

In order to broadly assess the global potential of the new hourly data set, we conduct a frequency analysis of the drifter velocity variance of the interpolated drifter data set from Argos. For comparison, we also consider the standard DAC kriged data set of drifter locations at 6 h intervals and the associated velocities calculated by a central difference scheme. For this analysis, we purposefully discard drifters tracked by GPS to focus on the results from the interpolation of Argos locations. We also ignore the drogue status of the drifters, considering the trajectories from deployment to drifter death. Estimating how much of the variance is due to real oceanic motions versus wind slippage [*Niiler et al.*, 1995] is beyond the scope of this assessment and should be considered in future studies.

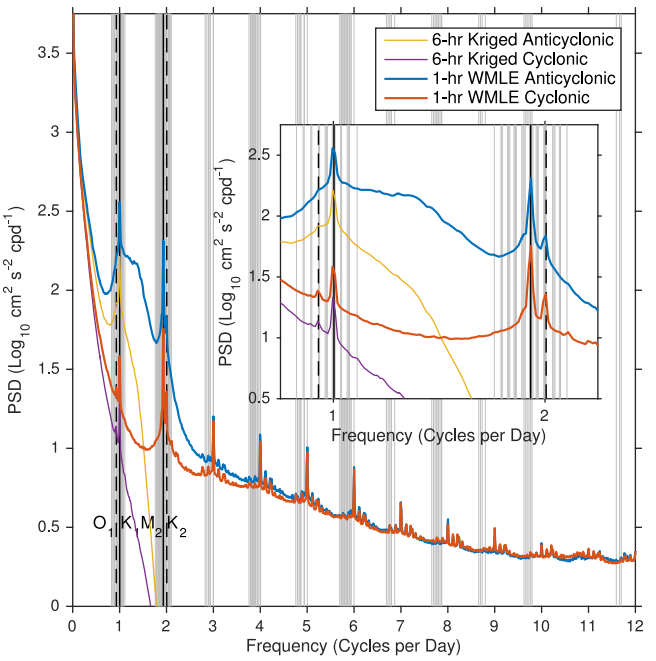

Figure 16. Globally averaged anticyclonic and cyclonic rotary spectra of drifter velocities from 2048 h (≈85 days) trajectory segments overlapping by 50%. Vertical gray lines correspond to astronomical tidal frequencies. The inset is a zoom for frequencies between 0.75 and 2.25 cpd.

Figure 16 shows globally averaged spectra obtained as follows. When long enough, we divide all available trajectories into trajectory segments of length 2048 h, overlapping by 50%, eventually obtaining 65,735 segments. From these, we consider the complex-valued time series of velocities which we multiply by a single Slepian taper with time-bandwidth product equal to 2 [*Slepian*, 1978] before taking the absolute FFT squared to obtain single rotary spectral estimates. We subsequently average anticyclonic frequency estimates (negative frequencies in the northern hemisphere, positive in the southern hemisphere) and all cyclonic frequency estimates (opposite signs) to obtain globally averaged anticyclonic and cyclonic spectra. We apply the same spectral estimation method to the 6-hourly kriged version of the same trajectories. It is beyond the scope of this study to analyze and interpret all aspects of this new global rotary spectra but we highlight a few points here.

As expected, over the common range of accessible frequencies, the rotary spectra from the hourly data surpass the spectra from the 6-hourly data, not only at near-inertial and tidal frequencies but also down to approximately 0.2 cpd, for both cyclonic and anticyclonic frequencies.

Up to approximately 5 cpd, the anticylonic spectrum is higher than the cyclonic spectrum, a consequence of the generally larger anticyclonic wind forcing [*Rio and Hernandez*, 2003; *Elipot and Gille*, 2009a] and preferential anticyclonic response of the upper ocean [*Gonella*, 1972; *Elipot and Gille*, 2009b].

The broad plateau of power between approximately 0.7 and 1.77 cpd corresponds primarily to wind-driven inertial oscillations in the ocean. The frequency distribution of this plateau is the result of the inertial frequency range (from 0 at the equator to 2 cpd at the poles) and of the latitudes of predominant wind forcing [*Elipot and Lumpkin*, 2008; *Chaigneau et al*., 2008]. As the inertial frequency goes to zero near the Equator, some of the power seen within the near-zero frequency peak corresponds to inertial oscillations blending with low-frequency motions. Note how the anticyclonic power plateau has no conspicuous counterpart in the cyclonic domain, indicating the ubiquitous circularity of inertial motions.

Many sharp peaks above the background spectrum stand out, some clearly located at tidal astronomical frequencies, notably the dominant diurnal and semidiurnal species O_1, K_1, M_2, and K_2 [*Poulain*, 2013; *Poulain and Centurioni*, 2015]. The anticyclonic peak for K_1 is clearly superimposed onto, and standing above, the inertial plateau. Many other peaks at higher frequencies also exist, some maybe associated with astronomical species, others with harmonics of the diurnal and semidiurnal peaks.

As an update of *Elipot and Lumpkin* [2008], we present in Figure 17 a latitude-frequency plot of the drifter velocity spectra in 1° latitude bins, averaged over all ocean basins, from both our new hourly data set (Figures 17a and 17b) and from the kriged 6-hourly data set (Figure 17c). Many of the same features highlighted in *Elipot and Lumpkin* [2008] can be seen, yet here more sharply because of the increased volume of data, the new method of interpolation, and more advanced spectral estimates. Note the sharply defined ridge of near-inertial energy as a function of latitude. Note also the spreading of energy toward higher frequency with increasing latitude. As seen previously in the global averaged spectra (Figure 16), spectral peaks near integer frequency numbers (in cpd) are conspicuous at almost all latitudes. A remarkable feature

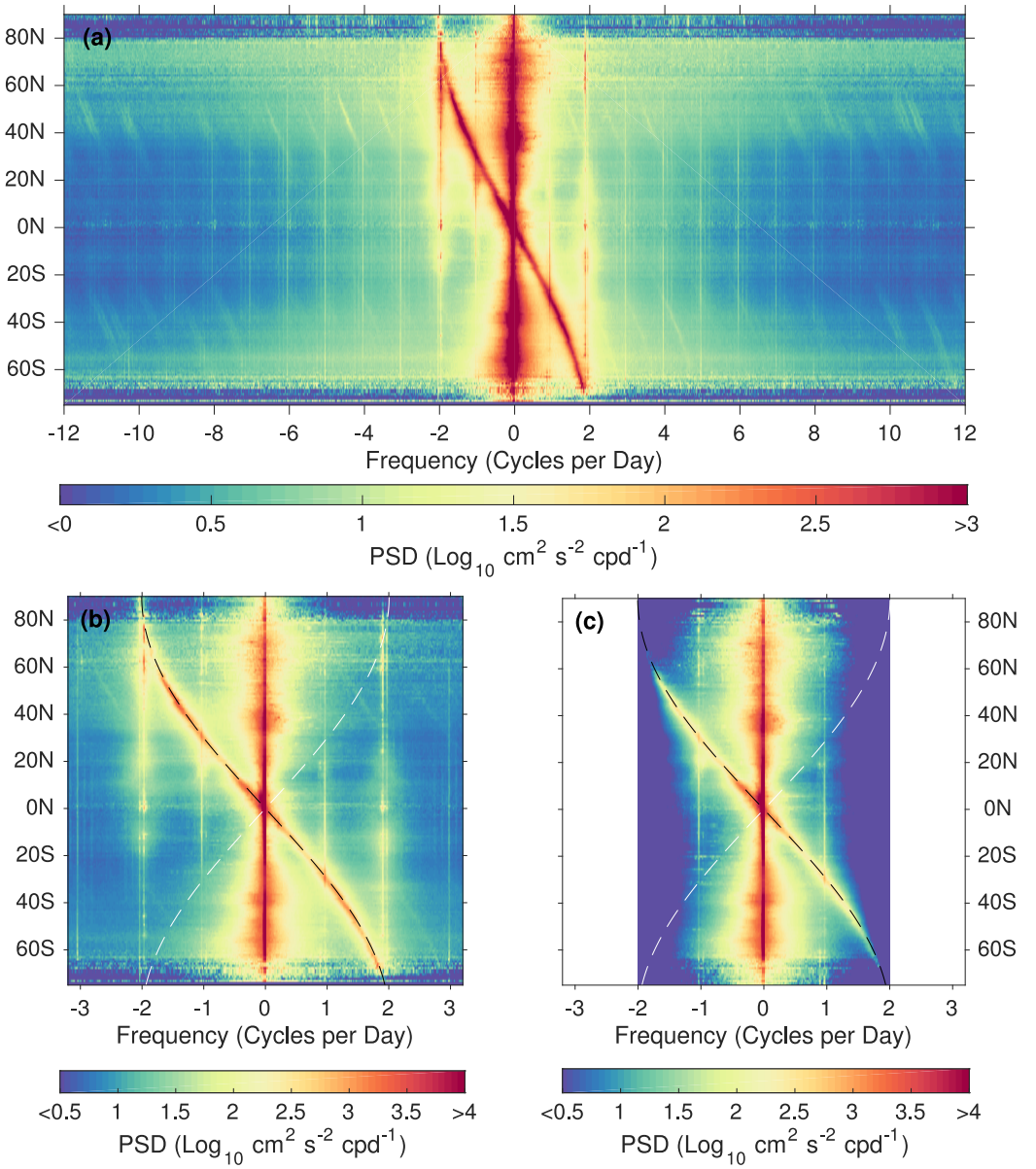

Figure 17. (a) Zonally averaged drifter rotary velocity spectra in 1° latitude bins from 2048 h (≈85 days) trajectory segments at 1 h resolution, overlapping by 50%, from the new global hourly product post-2005. (b) Same as in Figure 17a but zoomed in for frequencies in the range ±3.2 cpd. The black dashed curve indicates the inertial frequency ($-f/2\pi$ cpd) and the white dashed curve the Coriolis frequency ($f/2\pi$ cpd). Note that the color scales indicate different ranges of values and are both saturated. (c) Same as in Figure 17b but calculated from the post-2005 6-hourly Kriged global product for the same trajectories.

is the presence of replicates of the latitudinal-dependent inertial peak near frequencies $-f/2\pi \pm n$ where n is an integer. While the origins of these features are unknown at this stage, one can speculate that they may arise from wave triad interactions, or may indicate the departure from purely sinusoidal-shaped waves. Also, sampling artifacts may not be ruled out. To test to some extent the robustness of these features, we recompute the same spectra after applying the spherical linear interpolation method on the global data set (not shown). We find that the spectral peaks near integer frequency numbers persist, but the replicates of the near-inertial peak do not, most likely overwhelmed by the velocity noise introduced from the linear interpolation (see Figure 9). The study of the exact nature of these features will be the subject of future investigations. We calculate as a function of latitude the ratios of integrated hourly spectra to the 6-hourly spectra, for frequencies between 0 and 2 cpd, for positive and negative frequencies. We find that these

ratios (not shown) are generally larger than 1, and indicate that the new data set contains between 25% and 50% more variance than the 6-hourly kriged data set, and up to 100% more in the Southern Ocean.

6. Conclusions

We have reported in this paper the methods used to produce location and velocity estimates along the trajectories of GDP surface drifters, tracked by either Argos or by GPS. The resulting product is freely available via the DAC of the GDP (http://www.aoml.noaa.gov/phod/dac/dacdata.php). Preliminary analyses of the drifter velocity variance showed that this new data set is potentially a new tool for the study of relatively small-scale, and high-frequency, oceanic processes. In particular, velocity rotary spectra suggest that high-frequency tidal and internal wave motions are detectable globally.

A prerequisite for making the choice of interpolation method was to realize that the observed frequency distribution of the Argos locations is not well described by a normal PDF model, and this distribution varies significantly as a function of Argos location class, and between longitude and latitude coordinates. Our analyses of the Argos errors led us to devise an estimation method which considers that Argos location errors are distributed like *t* PDFs. This type of PDF has an extra parameter compared to a normal PDF, which can account for heavier tails and narrower central peaks. The underlying idea with this type of data modeling is that the estimation is less sensitive to outliers, and the method is deemed robust. For our interpolation model, we have chosen to model the drifter trajectories by an order one polynomial locally in time, and have applied a weighted maximum likelihood estimation method to obtain the coefficients of this polynomial.

An order one polynomial may not be high enough to model properly drifter trajectories when acceleration is important, but we have not found that using higher-order polynomials improved overall the results. Assessing the performance of the methods by comparing interpolated positions to GPS-derived positions, we have found that assuming either form of PDF of the data, normal or *t*, did not make a very large difference when considering the longitude and latitude errors, contrary to our expectations. It is likely that this does not make much of a difference because the error distributions are not dramatically different from normal distributions. In contrast, we have found that the method assuming *t* PDF of the data performed better when comparing the estimated drifter velocities to the GPS-derived drifter velocities. Variances of the trajectory parameters were obtained by bootstrapping and we have found that either assumption for the PDF of the data returned estimated coordinate variances commensurate with the observed variances. However, we found that bootstrapping assuming a *t* PDF of the data returned slightly better estimates of the velocity variances. Overall, these analyses led us to our final choice of interpolation method, the WMLE method with *t* PDF, for Argos data.

For GPS data, we have also chosen to model the drifter trajectories by a first-order polynomial, but we have estimated the polynomial coefficient at the original nonuniform GPS times by using the Locally Weighted Scatterplot Smoother, or LOWESS [*Cleveland*, 1979]. These parameters were subsequently linearly interpolated to uniform hourly steps.

Based on the analyses of 82 drifters released in the North Atlantic as part of the SPURS experiment, the interpolated locations from Argos by our chosen method can be expected to be most often absolutely accurate by 201 m, and 68% of these errors are expected to be 533 m or less. In the zonal direction, 68% of the errors are expected to be 355 m or less, and 309 m or less in the meridional direction. Our chosen method is not the method that performs the best for location errors: Kriging does. Yet Kriging is by definition an objective analysis method which requires one to specify a variogram model for the data, which can be obtained by fitting to the data first, as is done at the DAC for the 6 h product. For our purpose of producing a high-resolution data set, there are several issues with Kriging that ultimately led us to dismiss this method, beyond the fact that it did not result in the best estimate of velocities in our analyses. First, using a variogram model assumes that it is applicable to all time scales, which is certainly not the case in the ocean where linear and nonlinear superpositions of dynamical regimes occur. Second, fitting first a variogram model to the data before interpolating would require (as an example) to divide the drifter data into regular or irregular regions in which the model can be fitted, and ideally to take into account potential monthly, annual, and interannual variations of the parameters for this model. This would require extensive statistical

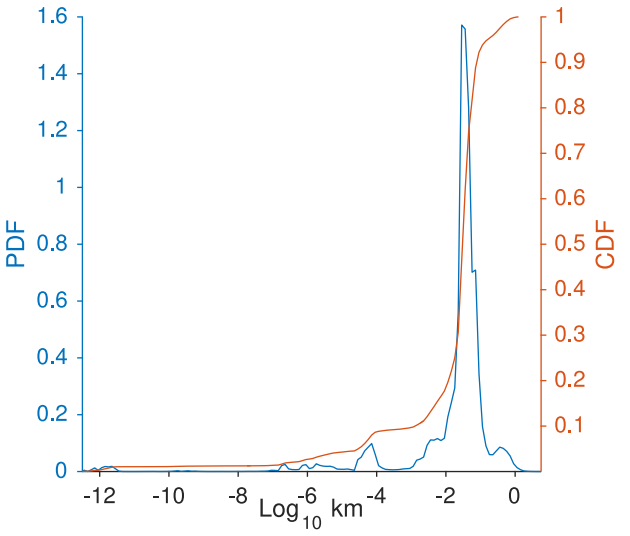

Figure 18. Kernel estimates of PDF and CDF of distances between QC GPS locations and smoothed GPS locations after applying a LOWESS filter (see Figure 2).

and dynamical analyses of the near-surface circulation from drifters, which is precisely what we are hoping to provide the tools for with this new global product.

Appendix A: Smoothing of Quality-Controlled GPS Locations

The quality-controlled GPS locations are processed by a 2-D implementation of a robust regression method called the Locally Weighted Scatterplot Smoother, or LOWESS [*Cleveland*, 1979], in order to produce smooth locations and horizontal velocities at nonuniform time steps (Figure 2). This processing consists of smoothing the GPS locations at their original nonuniform times by fitting iteratively (here 3 times) a polynomial of order 1, modeling two-dimensional (2-D) locations and their temporal derivatives in a local Cartesian plane of projection. We apply a weighted least squares method to fit the trajectory model to location data within a temporal window of variable bandwidth, or half-width h_k, around each original nonuniform time step t_k. We set the bandwidth $h_k=(t_{k+B}-t_{k-B})$ with $B=2$. For a given time t_k, the temporal window sets the weights for the data points at other times t_j as

$$K_{h_k}(t_j)=K\left(\frac{t_j-t_k}{h_k}\right), \tag{A1}$$

with the tricube kernel

$$K(t)=\frac{70}{81}(1-|t|^3)^3 I_{[-1,1]}(t), \tag{A2}$$

and

$$I(t)_{[-1,1]}=\begin{cases}1, & |t|\le 1\\ 0, & |t|>1.\end{cases} \tag{A3}$$

The purpose of the iteration is to calculate a new set of weights for the least squares estimation from the previous set, but scaled down at each pass by the median of the absolute distance residuals.

The probability and cumulative distribution functions of the nonzero distance differences between the QC GPS positions and the smoothed GPS positions are displayed in Figure 18. While the maximum difference is 6.8 km, 50% of differences are 30 m or less and 90% are 77 m or less. The differences may appear negligible, but the advantage of the method is that velocities, as well as locations, were estimated.

Appendix B: Estimates of the Structure Function From QC Location Data

In order to fit the structure function model (3) using SPURS drifter data, we first compute estimate $\widehat{S}$ of the structure function from longitude differences, latitude differences and angular separations, from either the QC nonuniform Argos or the QC nonuniform GPS locations. We consider trajectory segments of $N=170$ consecutive nonuniform Argos locations, and $N=170$ every other nonuniform GPS locations. For each of these segments (1950 from GPS and 1154 from Argos), we compute the $N!/2!(N-2)!=14,365$ possible half differences squared $(x_k-x_j)^2/2$ which we sort as a function of increasing time differences $\tau_{kj}=|t_k-t_j|$ for each segment. We then compute smooth estimates of S at discrete intervals τ, hourly intervals from 1 h to 6 days and then at daily intervals from 6 to 10 days. For these estimates, we use a Nadaraya-Watson kernel estimator with a 1-day bandwidth and an Epanechnikov kernel [e.g., *Fan and Gijbels*, 1996]. We then

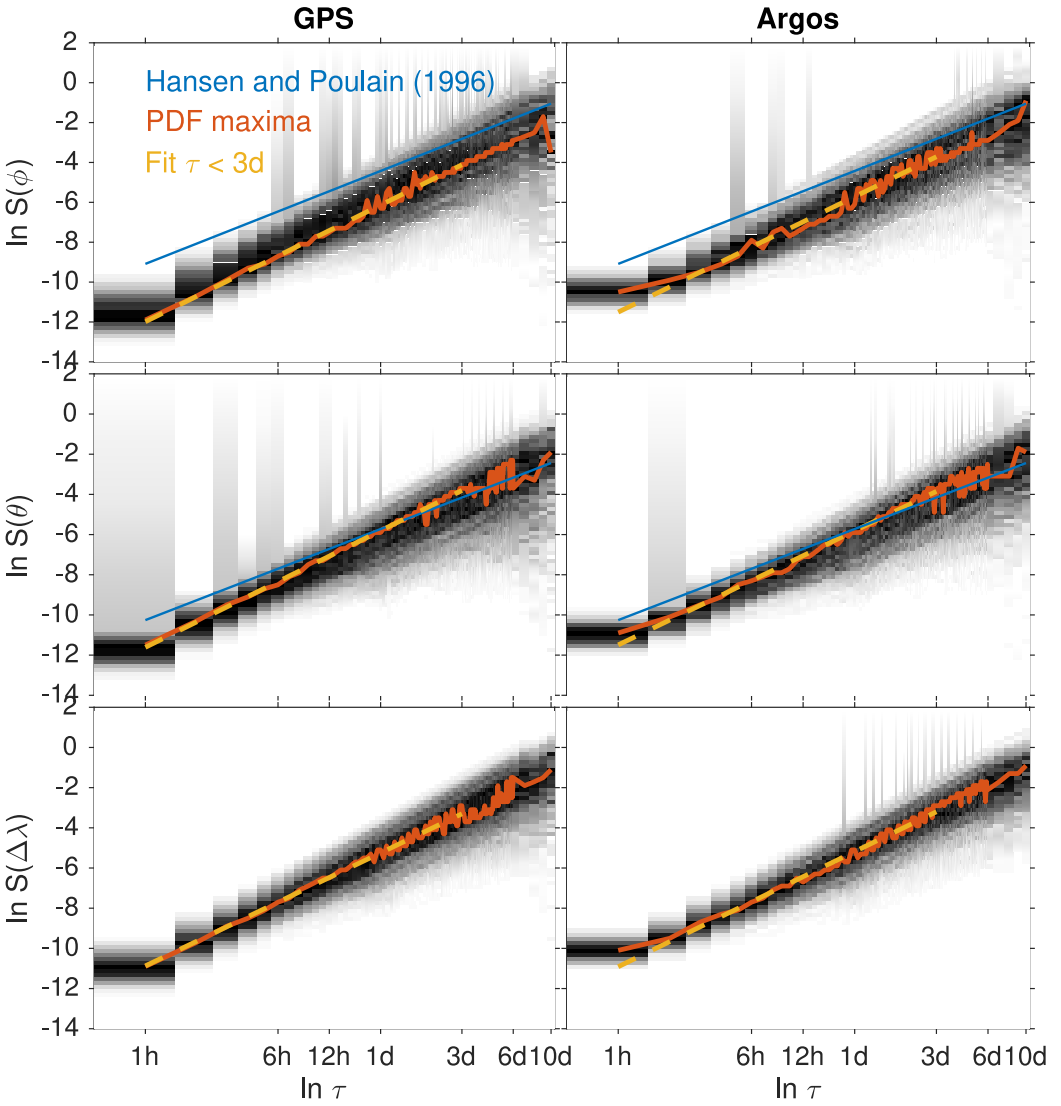

Figure 19. Two-dimensional histograms of variogram function S estimates for (top row) longitude ϕ, (middle row) latitude θ, and (bottom row) angular separation $\Delta\lambda$, for (left column) GPS data and (right column) Argos data. For each temporal bin, the gray shading represents the normalized histogram count from 1 (black) to zero (white). The orange curves connect the maximum in each bin. The straight curves corresponds to the variogram model $S(\tau)=\alpha\tau^{\beta}$ either fitted to the maxima in the range 1 h $<\tau<$ 3 day (dashed yellow lines), or from *Hansen and Poulain* [1996] for ϕ and θ (blue lines).

Table 9. Fitted Coefficients for the Variogram Model $S(\tau)=\alpha\tau^{\beta}$ of Figure 19 From GPS and Argos Data[a]

		$10^4\alpha$	β
Longitude	GPS	22	1.84
	Argos	33	1.82
	Hansen and Poulain [1996]	119; 99	1.47
Latitude	GPS	30	1.83
	Argos	30	1.79
	Hansen and Poulain [1996]	33; 32	1.43
Angular separation	GPS	53	1.77
	Argos	56	1.80

[a]The two values for α by *Hansen and Poulain* [1996] correspond to the updated coefficient by the DAC for the second half of 2012 and first half of 2013 for the North Atlantic.

consider all Nadaraya-Watson estimates at discrete time lags and form 2-D histograms of $\ln\widehat{S}$ and $\ln\tau$ (Figure 19).

For all cases, the 2-D histograms shows that the spread of $\ln\widehat{S}$ increases with $\ln\tau$. Yet there exists in general a peak of the $\ln\widehat{S}$ distribution for each $\ln\tau$ bin (shown as a jagged curve in each plot of Figure 19). These peaks are aligned along a near straight line in this 2-D logarithm space, compatible with model (3). Thus, we conduct a least squares fit for $\ln S(\tau)=\ln\alpha+\beta\ln\tau$ for all time differences shorter than 3 days

(results are listed in Table 9). Our results are larger than *Hansen and Poulain* [1996] and closer to 2, possibly because we focus on comparatively shorter time scales and the displacement of drifters may approach linear motion. Argos and GPS data return similar values for β, suggesting that to some extent both types of tracking systems capture similar drifter kinematics. For the α parameter from Argos data, we find commensurate values for longitude and latitude suggesting isotropic displacement variance but from GPS we find smaller values for longitude compared to latitudes. The method of *Hansen and Poulain* [1996] returns the opposite, that is, much larger values of α for longitude compared to latitude, maybe because of the coarser time scales.

We recalculated the structure function model parameters after discarding undrogued data points. We found that this lowers the values of the parameters, except for the β parameter for longitude which is increased. Yet, we find that the changes for β amounts to less than $\pm$10% of the values obtained with all data points, and the decreases in α amount to about 21% at the most for longitude from Argos data. Thus, we use for Kriging the parameter values for both drogued and undrogued data.

Acknowledgments

The authors collectively thank Mayra Pazos and Erik Valdez who conduct the quality control and data management of the DAC. This work was supported in part by the U.S. National Science Foundation through grants 1459482 (Elipot), 1031278 (Perez), and 1031002 (Lilly and Early). This research was also 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 #NA10OAR4320143. R. Lumpkin and R. Perez were funded by the Atlantic Oceanographic and Meteorological Laboratory and the Climate Observation Division, Climate Program Office, both of the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. The work of A.M. Sykulski was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme. The surface drifter data used for this study are available through the Data Assembly Center of the Global Drifter Program (http://www.aoml.noaa.gov/phod/dac/dacdata.php).

References

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom. Control*, *19*(6), 716–723.

Boyd, J. D., and D. J. Brightsmith (2013), Error properties of Argos satellite telemetry locations using least squares and Kalman filtering, *PLoS One*, *8*(5), e63051, doi:10.1371/journal.pone.0063051.

Bretherton, F., R. Davis, and C. Fandry (1976), A technique for objective analysis and design of oceanographic experiments applied to MODE-73, *Deep Sea Res.*, *23*, 559–582.

Carter, G. C. (1987), Coherence and time delay estimation, *IEEE Proc.*, *75*(2), 236–255, doi:10.1109/PROC.1987.13723.

Centurioni, L. R., V. Hörmann, Y. Chao, G. Reverdin, J. Font, and D.-K. Lee (2015), 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*(1), 96–105.

Chaigneau, A., O. Pizarro, and W. Rojas (2008), Global climatology of near-inertial current characteristics from Lagrangian observations, *Geophys. Res. Lett.*, *35*, L13603, doi:10.1029/2008GL034060.

Cleveland, W. (1979), Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, *74*, 829–836.

CLS (2011), *Argos User's Manual*. [Available at http://www.argos-system.org/manual/.]

Dzhaparidze, K., and S. Kotz (1986), *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*, Springer, N. Y.

Early, J. J. (2012), The forces of inertial oscillations, *Q. J. R. Meteorol. Soc.*, *138*(668), 1914–1922, doi:10.1002/qj.1917.

Efron, B., and G. Gong (1983), A leisurely look at the bootstrap, the jackknife, and cross-validation, *Am. Stat.*, *37*(1), 36–48.

Elipot, S., and S. T. Gille (2009a), Estimates of wind energy input to the Ekman layer from surface drifter data, *J. Geophys. Res.*, *114*, C06003, doi:10.1029/2008JC005170.

Elipot, S., and S. T. Gille (2009b), Ekman layers in the Southern Ocean: Spectral models and observations, vertical viscosity and boundary layer depth, *Ocean Sci.*, *5*(2), 115–139.

Elipot, S., and R. Lumpkin (2008), Spectral description of oceanic near-surface variability, *Geophys. Res. Lett.*, *35*, L05606, doi:10.1029/2007GL032874.

Elipot, S., R. Lumpkin, and G. Prieto (2010), Modification of inertial oscillations by the mesoscale eddy field, *J. Geophys. Res.*, *115*, C09010, doi:10.1029/2009JC005679.

Fan, J., and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*, vol. 66, CRC Press.

Gonella, J. (1972), A rotary-component method for analysing meteorological and oceanographic vector time series, *Deep Sea Res. Oceanogr. Abstr.*, *19*, 833–846.

Hansen, D. V., and A. Herman (1989), Temporal sampling requirements for surface drifting buoys in the Tropical Pacific, *J. Atmos. Oceanic Technol.*, *6*(4), 599–607.

Hansen, D. V., and P.-M. Poulain (1996), Quality control and interpolations of WOCE-TOGA drifter data, *J. Atmos. Oceanic Technol.*, *13*(4), 900–909.

Jackman, S. (2009), *Bayesian Analysis for the Social Sciences*, John Wiley, Hoboken, N. J., doi:10.1002/9780470686621.

Jonsen, I. D., J. M. Flemming, and R. A. Myers (2005), Robust state-space modeling of animal movement data, *Ecology*, *86*(11), 2874–2880.

Lopez, R., J.-P. Malarde, F. Royer, and P. Gaspar (2014), Improving Argos Doppler location using multiple-model Kalman filtering, *IEEE Trans. Geosci. Remote Sens.*, *52*(8), 4744–4755, doi:10.1109/TGRS.2013.2284293.

Lumpkin, R., and S. Elipot (2010), Surface drifter pair spreading in the North Atlantic, *J. Geophys. Res.*, *115*, C12017, doi:10.1029/2010JC006338.

Lumpkin, R., and G. C. Johnson (2013), Global ocean surface velocities from drifters: Mean, variance, ENSO response, and seasonal cycle, *J. Geophys. Res. Oceans*, *118*, 2992–3006, doi:10.1002/jgrc.20210.

Lumpkin, R., and M. Pazos (2007), Measuring surface currents with SVP drifters: The instrument, its data, and some recent results, in *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, pp. 39–67, Cambridge Univ. Press.

Lumpkin, R., N. Maximenko, and M. Pazos (2012), Evaluating where and why drifters die, *J. Atmos. Oceanic Technol.*, *29*(2), 300–308, doi:10.1175/JTECH-D-11-00100.1.

Lumpkin, R., S. A. Grodsky, L. Centurioni, M.-H. Rio, J. A. Carton, and D. Lee (2013), Removing spurious low-frequency variability in drifter velocities, *J. Atmos. Oceanic Technol.*, *30*(2), 353–360, doi:10.1175/JTECH-D-12-00139.1.

Massey, F. J., Jr. (1951), The Kolmogorov-Smirnov test for goodness of fit, *J. Am. Stat. Assoc.*, *46*(253), 68–78.

Maximenko, N., R. Lumpkin, and L. Centurioni (2013), Ocean surface circulation, in *Ocean Circulation and Climate: A 21st Century Perspective*, vol.103, pp. 283–284, Academic.

Niiler, P. P., A. L. Sybrandy, K. Bi, P. Poulain, and D. Bitterman (1995), Measurements of the water-following capability of holey-sock and TRISTAR drifters, *Deep Sea Res., Part I*, *42*, 1951–1964.

NSTB/WAAS T&E Team (2014), Global Positioning System (GPS) standard positioning service (sps) performance analysis report, technical report 86, William J. Hughes Tech. Cent., Atlantic City, N. J. [Available at http://www.gps.gov/systems/gps/performance/accuracy/.]

Poulain, P.-M. (2013), Tidal currents in the adriatic as measured by surface drifters, *J. Geophys. Res. Oceans*, *118*, 1434–1444, doi:10.1002/jgrc.20147.

Poulain, P.-M., and L. Centurioni (2015), Direct measurements of world ocean tidal currents with surface drifters, *J. Geophys. Res. Oceans*, *120*, 6986–7003, doi:10.1002/2015JC010818.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1988), *Numerical Recipes in Fortran, vol. 77*, Cambridge Univ. Press.

Rio, M. H., and F. Hernandez (2003), High-frequency response of wind-driven currents measured by drifting buoys and altimetry over the world ocean, *J. Geophys. Res.*, *108*(C8), 3283, doi:10.1029/2002JC001655.

Shoemake, K. (1985), Animating rotation with quaternion curves, *ACM SIGGRAPH Comput. Graph.*, *19*(3), 245–254, doi:10.1145/325165.325242.

Slepian, D. (1978), Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case, *Bell Syst. Tech. J.*, *57*(5), 1371–1430, doi:10.1002/j.1538-7305.1978.tb02104.x.