The Annals of Applied Statistics
2021, Vol. 15, No. 1, 22–40
https://doi.org/10.1214/20-AOAS1367

LATE 19TH CENTURY NAVIGATIONAL UNCERTAINTIES AND THEIR INFLUENCE ON SEA SURFACE TEMPERATURE ESTIMATES

BY CHENGUANG DAI[1,*], DUO CHAN[2,‡], PETER HUYBERS[2,§] AND NATESH PILLAI[1,†]

[1]*Department of Statistics, Harvard University,* [*]*chenguangdai@g.harvard.edu;* [†]*pillai@fas.harvard.edu*
[2]*Department of Earth and Planetary Sciences, Harvard University,* [‡]*duochan@g.harvard.edu;* [§]*phuybers@fas.harvard.edu*

Accurate estimates of historical changes in sea surface temperatures (SSTs) and their uncertainties are important for documenting and understanding historical changes in climate. A source of uncertainty that has not previously been quantified in historical SST estimates stems from position errors. A Bayesian inference framework is proposed for quantifying errors in reported positions and their implications on SST estimates. The analysis framework is applied to data from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS3.0) in 1885, a time when astronomical and chronometer estimation of position was common but predated the use of radio signals. Focus is upon a subset of 943 ship tracks from ICOADS3.0 that report their position every two hours to a precision of 0.01° longitude and latitude. These data are interpreted as positions determined by dead reckoning that are periodically updated by celestial correction techniques. The posterior medians of uncertainties in celestial correction are 33.1 km (0.30° on the equator) in longitude and 24.4 km (0.22°) in latitude, respectively. Celestial navigation uncertainties being smaller in latitude than longitude is qualitatively consistent with the relative difficulty of obtaining astronomical estimates. The posterior medians for two-hourly dead reckoning uncertainties are 19.2% for ship speed and 13.2° for ship heading, leading to random position uncertainties with median 0.18° (20 km on the equator) in longitude and 0.15° (17 km) in latitude. Reported ship tracks also contain systematic position uncertainties relating to precursor dead-reckoning positions not being updated after obtaining celestial position estimates, indicating that more accurate positions can be provided for SST observations. Finally, we translate position errors into SST uncertainties by sampling an ensemble of SSTs from the Multiscale Ultrahigh Resolution Sea Surface Temperature (MURSST) data set. Evolving technology for determining ship position, heterogeneous reporting and archiving of position information, and seasonal and spatial changes in navigational uncertainty and SST gradients together imply that accounting for positional error in SST estimates over the span of the instrumental record will require substantial additional effort.

1. Introduction. Accurate estimates of past sea surface temperatures (SSTs) are important for assessing historical climate states (Morice et al. (2012)), detecting and attributing changes in climate (Chan and Wu (2015)), and computing climate sensitivity (Gregory et al. (2002)). SST datasets are also used as boundary conditions to run general circulation models (Folland (2005), Sobel (2007)), and are assimilated as part of generating atmospheric reanalysis datasets (Dee et al. (2011)). SST datasets are, however, known to have substantial errors (Kent et al. (2017)), especially prior to the systematic satellite, drifters and moored buoy temperatures that became routinely available in the 1980s (Kennedy et al. (2011a)). For SST, quantified errors include those associated with random errors of individual measurements (Ingleby (2010), Kennedy et al. (2011a), Kent and Challenor (2006)), systematic errors associated with different measurement methods (Huang et al. (2017), Kennedy et al. (2011b)),

Received October 2019; revised June 2020.
Key words and phrases. State-space model, hierarchical model, position error, navigational uncertainty, sea surface temperature uncertainty.

offsets among different groups of observers (Chan et al. (2019)) as well as those associated with individual ships (Kennedy et al. (2011a), Kennedy, Smith and Rayner (2012)). Another important source of uncertainty involves mapping noisy and often sparse observations to infill unobserved locations (Kennedy (1930)).

Despite quantification of many contributors to SST uncertainties, we are unaware of previous studies having quantitatively assessed navigational uncertainties associated with historical ocean observations. That is, errors in position associated with incorrectly recording or transcribing locations have been recognized (Woodruff et al. (1998)), as have errors introduced by rounding of positions (Kent, Challenor and Taylor (1999)), but the magnitude of navigational uncertainties prior to the widespread deployment of radio navigation in the 1930s (Fried (1977)) appears not to have been quantified. Prior to radio navigation, ship position in the open ocean was mainly estimated by dead reckoning and celestial techniques (Bowditch (1906)). Dead reckoning involves updating ship position using estimates of heading and distance. Celestial navigation involves estimating latitude from the zenith angle associated with various celestial bodies, including the sun, moon and stars. Longitude may be inferred using a chronometer method whereby the difference between a local apparent time and the time at some known longitude are determined from a clock carried onboard or some other method, such as the phase of Jupiter's moons. Dead reckoning can potentially introduce both systematic and random uncertainties, whereas celestial correction is assumed free of systematic uncertainties.

Position errors have implications on the accuracy of mapped SSTs. For example, if SST measurements are binned into gridboxes, misspecification of the appropriate box will influence the mean and higher-order moments (Director and Bornn (2015)). Cervone and Pillai (2015) have shown that incorporating position uncertainties when averaging land-station data within gridboxes, which have typically been assumed to reside at the center of the gridbox, is important for valid inference of land surface temperatures, and we expect that the additional uncertainties over the sea associated with ship positions are no less important. In the following, we propose a Bayesian model to quantify position errors for various ship tracks through estimating navigational uncertainties in dead reckoning and celestial correction. We then translate position errors into SST uncertainties by sampling high-resolution SSTs using posterior samples of ship positions.

2. Data description. The ship data used in this study are from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS3.0) (Freeman et al. (2017)), which is the most comprehensive available historical dataset of ship-based measurements from the eighteenth century to the present. We use data from 1885 to demonstrate a Bayesian framework for estimating position errors associated with historical ship tracks. Individual ship tracks are identified using ICOADS identification (ID) information, and tracks with missing or nonunique IDs are excluded. Ship tracks traversing across open ocean are separated into shorter segments whenever they are close to islands. Only ship tracks that have their positions reported at a resolution of every four hours or better and to a precision of better than one degree longitude and latitude are retained. These high-resolution ship tracks are focussed on because it is otherwise difficult to identify positions errors and to distinguish between contributions from dead-reckoning and celestial navigation.

The highest resolution data comes from the U.S. Marine Meteorological Journals Collection, which was a program sponsored by the U.S. Navy's Hydrographic Office, that enlisted the help of commercial vessels in compiling meteorological data. Reports are primarily, albeit not exclusively, from U.S. vessels and are provided every two hours at a resolution of 0.01° longitude and latitude. In total there are 1,341 of these two-hourly ship tracks. 943 of these tracks, referred to as HQ2 tracks, are characterized by stable velocities that are episodically

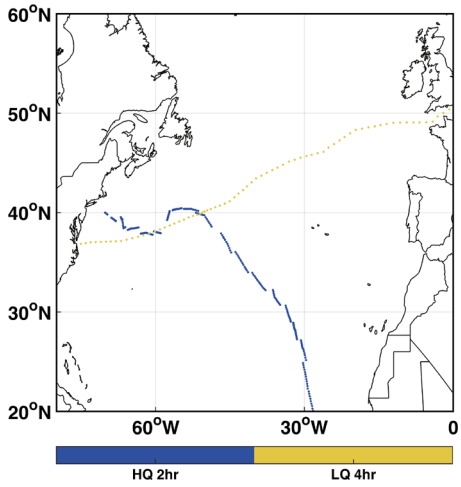

FIG. 1. *Two types of ship tracks. The blue track is the high-quality two-hourly ship track No.*30, *and the yellow track is the low-quality four-hourly ship track No.*108. *Jumps in position are seen in the two-hourly track, whereas the four-hourly track appears overly smooth.*

punctuated by jumps in position (see Figure 1 and Table 1). Jumps in otherwise smooth ship tracks typically occur at midnight and are consistent with navigation using dead reckoning that is updated by a celestial positioning technique (Bowditch (1906)). Ship tracks generally follow well-established trade routes that tend to be meridional in the tropics and zonal in the midlatitudes with the highest data density in the Atlantic, the Eastern Pacific and the Southern Indian Ocean (see Figure 2).

The remaining 398 two-hourly tracks show static positions followed by jumps averaging 84.6 km in two hours which is unphysical for a ship under sail. We are unaware of metadata indicating how these positions were prescribed, and thus exclude these tracks from our present analysis.

There also exists a separate collection of 576 ship tracks that report position every four hours to a precision of 0.01° longitude and latitude. These four-hourly tracks, referred to as LQ4 tracks, primarily track zonally between Europe and North America and meridionally between Europe and South America. Unlike HQ2 tracks, LQ4 tracks appear overly smooth, showing no discontinuities as would be expected from celestial navigational updates. We assume that the position reports of these tracks have been manually interpolated, implying that they contain less useful information for purposes of inferring navigational uncertainties. Although LQ4 tracks are unreliable for inferring underlying navigational uncertainties, these more smoothly-varying tracks are more generally representative of position data available in ICOADS, and we develop a methodology for exploring their position uncertainties that leverages results obtained from HQ2 tracks.

TABLE 1
Empirical speed and the jumping distance

	Empirical speed (km/hr)			Jumping distance (km)		
Quantiles	25%	50%	75%	25%	50%	75%
HQ2	6.4	10.4	14.6	15.0	22.8	37.0
LQ4	15.8	18.3	19.6	–	–	–

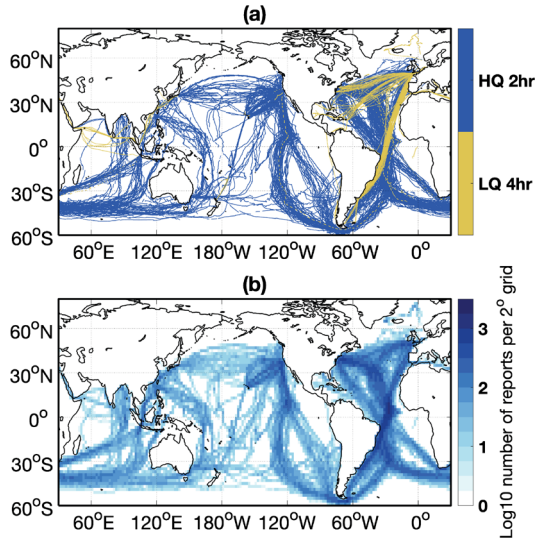

FIG. 2. *Ship tracks used in this study.* (*a*) *Tracks are divided into high-quality, two-hourly ship tracks* (*HQ*2) *and low-quality, four-hourly ship tracks* (*LQ*4). (*b*) *Density of ship position reports at* 2° *resolution.*

It is necessary to define when celestial updates occur for HQ2 tracks. Celestial navigational updates are presumed to occur when the ship track jumps. Using the speed and heading from neighboring ship positions, we predict the next position, and a jump is identified when the predicted and reported positions differ by at least 7 km in either longitude or latitude. 7 km is chosen on account of its being the 80th percentile of latitudinal differences between predicted and reported ship positions. Note that longitude and latitude are treated independently because their respective methods of celestial positioning are distinct. When several jumps are identified in a single day, only the largest jump is selected.

3. Bayesian model. The proposed Bayesian model for estimating position errors contains three stages. First, position errors, uncertainties in ship speed and heading are inferred for each HQ2 track using a state-space time series model. Second, navigational uncertainties are synthesized across different HQ2 tracks using a Bayesian hierarchical model. Finally, uncertainties are modeled for LQ4 data using a forward navigation model based upon results obtained from HQ2 data. Stages two and three utilize the posterior samples obtained from prior stages. All the models described below are fitted using RStan (Stan Development Team (2019)).

We note that ideally, these stages would be integrated into a single, inclusive, hierarchical Bayesian model. However, as we will see, for each ship track the number of parameters in the model is approximately two times the length of the track. Based on our experiences, it can take up to six hours using RStan to fit a single ship track with approximate 400 reported positions. Therefore, given the amount of data as well as the model complexity, we proceed with the multistage approach mentioned above to bypass the computational difficulty of implementing a full Bayesian procedure.

3.1. *State-space model for HQ*2 *tracks.* The proposed model utilizes the reported HQ2 ship positions to empirically calculate ship speed and heading at two-hourly time steps. Ship position and heading are in radians and speed is in km/hr. Let (q_t^x, q_t^y) be the displacements (km) that the ship travels from the starting position to the *reported* position at time step t. Correspondingly, let (p_t^x, p_t^y) be the displacements from the starting position to the *true* position. q_t^x, p_t^x (q_t^y, p_t^y) are positive if the current ship position is to the east (north) of the

TABLE 2
Definition of parameters

Parameter	Definition
(ϕ_t, ψ_t)	Reported ship position in longitude and latitude.
(q_t^x, q_t^y)	Displacement from the starting position to the reported position.
(p_t^x, p_t^y)	Displacement from the starting position to the true position.
$s_t, \hat{s}_t, \mu_s$	True ship speed, empirical ship speed, mean of ship speed.
$\theta_t, \hat{\theta}_t, \beta$	True ship heading, empirical ship heading and its systematic bias.
σ_s, σ_θ	Evolutionary uncertainties in ship speed and heading.
τ_x, τ_y	Track-based uncertainties in celestial correction.
τ_s, τ_θ	Track-based uncertainties in dead reckoning.
$\mu_{\tau_x}, \mu_{\tau_y}$	Population median of uncertainties in celestial correction.
$\mu_{\tau_s}, \mu_{\tau_\theta}$	Population median of uncertainties in dead reckoning.

The subscripts or superscripts x, y, s, θ refer to the longitudinal direction, the latitudinal direction, ship speed and ship heading, respectively.

starting point. q_t^x, q_t^y are calculated as accumulated sums following equation (3.1), where ϕ denotes the reported longitude, ψ denotes the reported latitude, r_a denotes the radius of earth, and the subscripts denote the time step. The cosine term accounts for changes in distance-longitude scaling with latitude. Definitions of parameters used in the model are listed in Table 2:

$$q_t^x = \sum_{i=1}^{t} r_a(\phi_i - \phi_{i-1}) \cos\left(\frac{\psi_i + \psi_{i-1}}{2}\right), \qquad q_t^y = \sum_{i=1}^{t} r_a(\psi_i - \psi_{i-1}). \tag{3.1}$$

3.1.1. *Transition model.* For ship speed s_t, we assume a customary centered AR model described as below:

$$s_t = \mu_s + \alpha_s(s_{t-1} - \mu_s) + \epsilon_t^s. \tag{3.2}$$

$\alpha_s \in (0, 1)$ denotes the drift parameter, and μ_s denotes the unknown *track-based population* mean of s_t. Conditioning on α_s, μ_s and s_{t-1}, we assume that ϵ_t^s follows a truncated normal distribution with mean 0, variance σ_s^2 and lower truncation at $-\mu_s - \alpha_s(s_{t-1} - \mu_s)$, so as to guarantee that s_t is nonnegative.

For ship heading θ_t, we assume a simple random walk model. That is,

$$\theta_t = \theta_{t-1} + \epsilon_t^\theta, \quad \epsilon_t^\theta \sim N(0, \sigma_\theta^2). \tag{3.3}$$

Both ϵ_t^s and ϵ_t^θ are assumed to be independent across time. The transition model is essentially parametrized in terms of s_t and θ_t, while the aggregated displacements p_t^x and p_t^y are deterministic functions of s_t and θ_t. That is,

$$p_t^x = p_{t-1}^x + 2s_t \cos\theta_t, \qquad p_t^y = p_{t-1}^y + 2s_t \sin\theta_t, \tag{3.4}$$

which captures the physical navigation process by using ship speed and ship heading to project its next position (the same as dead reckoning navigation). We multiply s_t by 2 in the model because we are modeling two-hourly ship tracks, whereas s_t is in unit of km/hr.

3.1.2. *Observation model.* Suppose there are in total m celestial updates along the ship track. Let $\mathcal{C} = \{t_1, \ldots, t_m\}$ be the set of the corresponding time steps.

For $t \notin \mathcal{C}$ (e.g., position B in Figure 3), we note that the reported ship trajectory is systematically biased, thus the ship position contains accumulated position errors. Therefore,

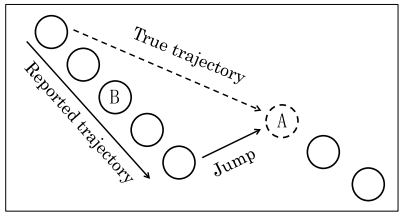

FIG. 3. *A cartoon illustration of a celestial update along HQ2 tracks. Reported ship positions are represented by circles, and the dashed circle denotes a celestial correction. The reported ship trajectory is indicated by the solid arrow, but it is systematically biased relative to the true ship trajectory indicated by the dashed arrow.*

rather than basing the model on position information, we use the empirical ship speed $\hat{s}_t$ and ship heading $\hat{\theta}_t$, which are not expected to have persistence insomuch as dead-reckoning is utilized for navigation. Mathematically, for $t \in (t_k, t_{k+1})$, we assume

$$\hat{s}_t \sim N\big(s_t, (\tau_s s_t)^2\big), \qquad \hat{\theta}_t \sim N\big(\theta_t + \beta_k, \tau_\theta^2\big), \tag{3.5}$$

in which τ_s and τ_θ denote the track-based uncertainties in dead reckoning. It is worthwhile to note that in early navigations, ship speed was usually measured by a chip log that measures the length of a rope released as a ship travels, or by a patent log that counts the numbers of turns a turbine rotates when water passes though (Bowditch (1906)). Since both methods tend to have larger measurement errors at higher ship speed, we model the uncertainty in terms of the relative ship speed instead of the absolute ship speed. Besides, we introduce β_k to model the systematic bias in ship heading between two celestial updates at time step t_k and t_{k+1}. The prior on β_k is Unif$(-\pi, \pi]$.

For $t \in \mathcal{C}$ (e.g., position A in Figure 3), empirical ship speed and heading are ignored, as they are very inaccurate because of the large jumps (see Table 1), whereas the reported ship positions are assumed to contain only celestial observational errors. Thus, for $t \in \mathcal{C}$, we assume

$$q_t^x \sim N\big(p_t^x, (\tau_x \cos\psi_t)^2\big), \qquad q_t^y \sim N\big(p_t^y, {\tau_y}^2\big). \tag{3.6}$$

q_t^x and q_t^y denote the observed aggregated displacements. ψ_t denotes the reported latitude. τ_x and τ_y denote the track-based uncertainties in celestial correction in the longitudinal and latitudinal direction, respectively.

3.2. *Synthesizing information across different HQ2 tracks.* The state-space model discussed in Section 3.1 describes a single HQ2 track. We assume consistent levels of accuracy in celestial correction and dead reckoning over HQ2 tracks, permitting for borrowing information on the uncertainty parameters across different tracks. Let $\{\tau_x^{(j)}, \tau_y^{(j)}, \tau_s^{(j)}, \tau_\theta^{(j)}\}$ be the uncertainty parameters associated with HQ2 track j. We assume a hierarchical structure on the uncertainty parameters,

$$\log \tau_x^{(j)} \sim N\big(\log \mu_{\tau_x}, \gamma_{\tau_x}^2\big), \qquad \log \tau_y^{(j)} \sim N\big(\log \mu_{\tau_y}, \gamma_{\tau_y}^2\big). \tag{3.7}$$

$\log \mu_{\tau_x}$ and $\log \mu_{\tau_y}$ denote the respective population medians of $\log \tau_x^{(j)}$ and $\log \tau_y^{(j)}$. Since the log transformation is monotone, μ_{τ_x} and μ_{τ_y} are the respective population medians of $\tau_x^{(j)}$ and $\tau_y^{(j)}$. $\tau_s^{(j)}$, $\tau_\theta^{(j)}$ are modeled in the same fashion. Conditioning on the population-level parameters, track-based navigational parameters are assumed to be independent.

To remedy the computational challenge, we consider a second-stage model, where we synthesize the uncertainty information across tracks based on the posterior samples obtained by fitting each HQ2 track. The variability of the navigational parameters across different tracks are summarized in Figure 12 in the Appendix.

Given n posterior samples $\{\tau_x^{(j,i)}, \tau_y^{(j,i)}\}_{i\in\{1,\cdots,n\}}$ of HQ2 track j, we assume a hierarchical model,

$$\begin{aligned} \log \tau_x^{(j,i)} &\sim N\big(\log \mu_{\tau_x}^{(j)}, \eta_{\tau_x}^{(j)^2}\big), \qquad \log \mu_{\tau_x}^{(j)} \sim N\big(\log \mu_{\tau_x}, \gamma_{\tau_x}^2\big), \\ \log \tau_y^{(j,i)} &\sim N\big(\log \mu_{\tau_y}^{(j)}, \eta_{\tau_y}^{(j)^2}\big), \qquad \log \mu_{\tau_y}^{(j)} \sim N\big(\log \mu_{\tau_y}, \gamma_{\tau_y}^2\big), \end{aligned} \tag{3.8}$$

in which $\mu_{\tau_x}^{(j)}$, $\mu_{\tau_y}^{(j)}$ and $\eta_{\tau_x}^{(j)}$, $\eta_{\tau_y}^{(j)}$ denote the track-level medians and standard deviations for HQ2 track j. Conditioning on the track-level parameters $\mu_{\tau_x}^{(j)}$, $\eta_{\tau_x}^{(j)}$ and $\mu_{\tau_y}^{(j)}$, $\eta_{\tau_y}^{(j)}$, we assume $\tau_x^{(j,i)}$ and $\tau_y^{(j,i)}$ are independent across the sample index i. Conditioning on the population-level parameters μ_{τ_x}, γ_{τ_x} and μ_{τ_y}, γ_{τ_y}, we assume $\mu_{\tau_x}^{(j)}$, $\eta_{\tau_x}^{(j)}$ and $\mu_{\tau_y}^{(j)}$, $\eta_{\tau_y}^{(j)}$ are independent across the track index j. $\{\tau_s^{(j,i)}, \tau_\theta^{(j,i)}\}$ are modeled in the same fashion.

3.3. *Forward navigation model for LQ4 tracks.* The forward navigation model aims at representing dead reckoning and celestial correction contributions to position errors for LQ4 tracks. We assume that the population-level uncertainties in ship speed and heading for LQ4 tracks are consistent with those for HQ2 tracks. Speed and heading can then be represented as

$$\hat{s}_t = s_t\big(1 + e_t^s\big), \qquad \hat{\theta}_t = \theta_t + e_t^\theta, \tag{3.9}$$

where e_t^s, e_t^θ are assumed to be i.i.d. over time. We impose the following hierarchical priors on e_t^s, e_t^θ:

$$\begin{aligned} e_t^s &\sim N\big(0, \tau_s^2\big), \qquad \log \tau_s \sim N\big(\log \hat{\mu}_{\tau_s}, \hat{\gamma}_{\tau_s}^2\big), \\ e_t^\theta &\sim N\big(0, \tau_\theta^2\big), \qquad \log \tau_\theta \sim N\big(\log \hat{\mu}_{\tau_\theta}, \hat{\gamma}_{\tau_\theta}^2\big), \end{aligned} \tag{3.10}$$

in which $\hat{\mu}_{\tau_s}$, $\hat{\mu}_{\tau_\theta}$, $\hat{\gamma}_{\tau_s}$ and $\hat{\gamma}_{\tau_\theta}$ are the empirical posterior means of μ_{τ_s}, μ_{τ_θ}, γ_{τ_s} and γ_{τ_θ}, respectively, obtained by fitting the hierarchical model discussed in Section 3.2. τ_x and τ_y, which calibrate the uncertainties in celestial correction, are incorporated into the data likelihood as in equation (3.6), and their priors are specified similarly as in equation (3.10).

For celestial navigation, we denote p as the probability that a celestial observation is employed. In order to illustrate the sensitivity of our results to whether celestial observations are taken—and because we have no direct evidence of whether celestial observations are employed on any given day—we provide results assuming observations are taken every night ($p = 1$), half of the time ($p = 0.5$) and never ($p = 0$).

4. Results.

4.1. *High-quality two-hourly ship tracks.* A single track, HQ2 track No.30, is described for purposes of illustrating the results. HQ2 track No.30 moved from west to east, where blue dots in Figure 4 represent the reported ship positions, black dots represent the posterior mean positions and ellipses indicate posterior one-standard deviation uncertainties. The trajectory of the posterior mean tends to diverge from the reported ship positions when celestial navigation updates are large because, unlike reported positions, posterior means take into account not only information from preceding celestial updates but also later ones. Figure 5 demonstrates the global pattern of random position uncertainties for HQ2 tracks.

For celestial correction, the population median of the uncertainty in the longitudinal direction is approximately 33.1 km or 0.30° on the equator, while the population median of the uncertainty in the latitudinal direction is smaller, approximately 24.4 km or 0.22° (see Table 3). A smaller uncertainty estimate in latitude accords with expectation because celestial

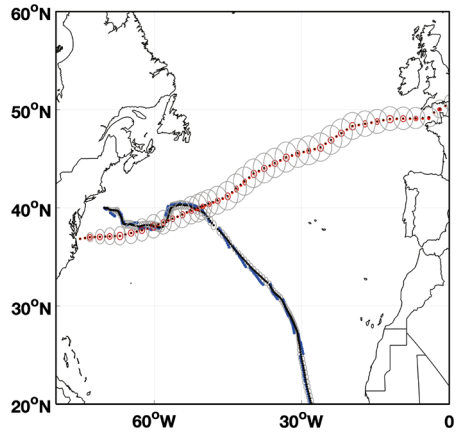

FIG. 4. *The posterior distributions of ship positions. For HQ2 track No.30, blue dots are reported ship positions, black dots represent the posterior means and circles show posterior uncertainties (one standard deviation). For LQ4 track No.108, yellow dots are reported ship positions, red dots and circles are the posterior means and uncertainties assuming a celestial correction happens every midnight* ($p = 1$)*, and black dots and gray circles are the posterior means and uncertainties assuming there are no celestial corrections* ($p = 0$).

correction in the longitudinal direction is subject to errors in both celestial observations and chronometers, whereas celestial correction in the latitudinal direction is free of chronometer errors (Bowditch (1906)). The population median of the uncertainty in the relative ship speed $\hat{s}_t/s_t$ is approximately 19.2%, which is possibly attributed to the less reliable instruments, for example, chip logs and patent logs, used in early navigations. The population median of the uncertainty in ship heading is approximately 0.23 radian or 13.2°. Among all HQ2 tracks, the

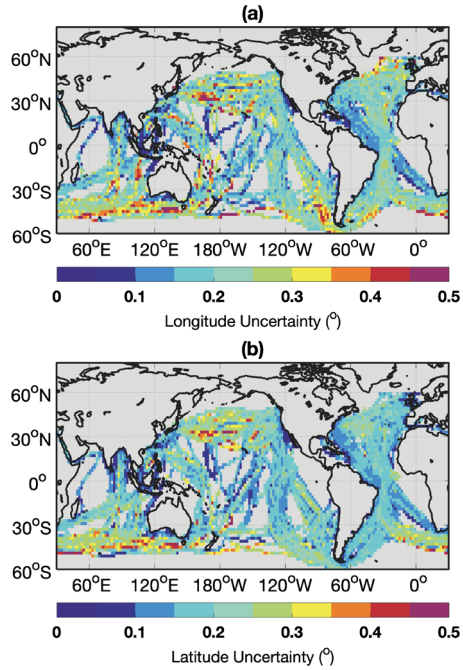

FIG. 5. *Global pattern of random position uncertainties for HQ2 tracks. Individual panels are:* (*a*) *longitude and* (*b*) *latitude. Track-based estimates of random position uncertainties are gridded to* 2° *resolution for visualization. Binning is accomplished by averaging random position uncertainties within a gridbox in quadrature.*

TABLE 3
The population median of navigational parameters

Quantiles	5%	25%	50%	75%	95%	std
μ_{τ_x} (km)	31.6	32.4	33.1	33.7	34.6	0.92
μ_{τ_y} (km)	23.2	23.9	24.4	24.9	25.6	0.74
μ_{τ_s} (%)	18.3	18.8	19.2	19.6	20.1	0.5
μ_{τ_θ} (rad)	0.21	0.22	0.23	0.24	0.25	0.01

μ_{τ_x} and μ_{τ_y} denote the population median of the uncertainty in celestial correction along the longitudinal and the latitudinal directions, respectively. μ_{τ_s} and μ_{τ_θ} denote the population median of the uncertainty in the relative ship speed $\hat{s}_t/s_t$ and ship heading, respectively.

25%, 50% and 75% quantiles of the absolute difference between the reported ship heading and the posterior mean heading are 2.9°, 7.9° and 20.4°, respectively.

4.2. *Low-quality four-hourly ship tracks*. As noted, we are unable to infer when LQ4 tracks have their positions updated by celestial observations and, therefore, explore three scenarios wherein celestial positions are made every midnight, with 0.5 probability, or never. Figure 4 shows the posterior distributions of LQ4 track No.108 under the best-case and worst-case scenarios. Under the best-case scenario, all the midnight positions along the track are considered being celestially corrected, and LQ4 track exhibits small Brownian-bridge uncertainty structures between consecutive midnight positions. Under the worst-case scenario, we assume no celestial corrections, and LQ4 track exhibits a Brownian-bridge uncertainty structure that spans the departure and arrival points.

Table 4 summarizes position errors of all LQ4 tracks under the three scenarios. The overall position uncertainty (MSE) combines the random position uncertainty (standard deviation) and the systematic position uncertainty (bias) using the bias-variance decomposition. On average, random position uncertainties of LQ4 tracks are 0.50° (56 km on the equator) in longitude and 0.44° (49 km) in latitude under the worst-case scenario, and 0.16° (18 km) in longitude and 0.13° (14 km) in latitude under the best-case scenario. The half-probability scenario is similar to, albeit of course slightly more uncertain than, the best-case scenario. The random position uncertainty of LQ4 tracks under the best-case scenario appears to be smaller than that of HQ2 tracks, where the latter have only 87% of nights associated with a celestial correction. Because there are no apparent jumps, LQ4 tracks are inferred to have smaller systematic uncertainty than estimated for HQ2 tracks.

The Brownian-bridge uncertainty structure implies larger errors associated with longer journeys, and being further away from departure and arrival points such that positions in the interior of oceans are generally more uncertain (see Figure 6). Note that position uncertainties depend not only on the distance from coasts or islands but also on directions that ships are heading, which determines the relative magnitude of the uncertainties in longitude and latitude.

4.3. *SST uncertainties*. To quantify uncertainties in SSTs associated with errors in position, we sample a high-resolution SST dataset with position errors that mimic those expected from our analysis of HQ2 tracks. We use the Multiscale Ultrahigh Resolution Sea Surface Temperature dataset (MURSST) (Chin, Vazquez-Cuervo and Armstrong (2017)) that incorporates infrared and microwave satellite retrievals and observations from ships and buoys. Although the data is obviously more recent than the 1885 ship tracks that we analyze, MURSST has the advantage of having a 0.01° spatial resolution that is comparable to the

TABLE 4
Random and systematic position uncertainties

Random position uncertainty				
Quantiles	25%	50%	75%	mean
HQ2 longitude	0.11° (12 km)	0.18° (20 km)	0.23° (26 km)	0.18° (20 km)
HQ2 latitude	0.10° (11 km)	0.15° (17 km)	0.20° (22 km)	0.16° (18 km)
LQ4 longitude (1.0)	0.11° (12 km)	0.16° (18 km)	0.20° (22 km)	0.16° (18 km)
LQ4 longitude (0.5)	0.16° (18 km)	0.22° (24 km)	0.28° (31 km)	0.23° (26 km)
LQ4 longitude (0.0)	0.35° (39 km)	0.50° (56 km)	0.66° (73 km)	0.50° (56 km)
LQ4 latitude (1.0)	0.09° (10 km)	0.13° (14 km)	0.17° (19 km)	0.13° (14 km)
LQ4 latitude (0.5)	0.13° (14 km)	0.18° (20 km)	0.24° (27 km)	0.19° (21 km)
LQ4 latitude (0.0)	0.29° (32 km)	0.44° (49 km)	0.60° (67 km)	0.44° (49 km)
Systematic position uncertainty				
Quantiles	25%	50%	75%	mean
HQ2 longitude	0.06° (6.7 km)	0.16° (18 km)	0.34° (38 km)	0.24° (27 km)
HQ2 latitude	0.05° (5.6 km)	0.14° (16 km)	0.30° (33 km)	0.21° (23 km)
LQ4 longitude (1.0)	0.00° (0.0 km)	0.00° (0.0 km)	0.01° (1.1 km)	0.00° (0.0 km)
LQ4 longitude (0.5)	0.00° (0.0 km)	0.01° (1.1 km)	0.01° (1.1 km)	0.01° (1.1 km)
LQ4 longitude (0.0)	0.01° (1.1 km)	0.03° (3.3 km)	0.06° (6.7 km)	0.05° (5.6 km)
LQ4 latitude (1.0)	0.00° (0.0 km)	0.00° (0.0 km)	0.01° (1.1 km)	0.00° (0.0 km)
LQ4 latitude (0.5)	0.00° (0.0 km)	0.01° (1.1 km)	0.01° (1.1 km)	0.01° (1.1 km)
LQ4 latitude (0.0)	0.01° (1.1 km)	0.03° (3.3 km)	0.06° (6.7 km)	0.05° (5.6 km)
Overall position uncertainty				
Quantiles	25%	50%	75%	mean
HQ2 longitude	0.16° (18 km)	0.26° (29 km)	0.42° (47 km)	0.32° (36 km)
HQ2 latitude	0.14° (16 km)	0.23° (26 km)	0.37° (41 km)	0.28° (31 km)
LQ4 longitude (1.0)	0.12° (13 km)	0.16° (18 km)	0.20° (22 km)	0.16° (18 km)
LQ4 longitude (0.5)	0.16° (18 km)	0.22° (24 km)	0.28° (31 km)	0.22° (24 km)
LQ4 longitude (0.0)	0.36° (40 km)	0.51° (57 km)	0.68° (75 km)	0.51° (57 km)
LQ4 latitude (1.0)	0.10° (11 km)	0.14° (16 km)	0.17° (19 km)	0.14° (16 km)
LQ4 latitude (0.5)	0.13° (14 km)	0.18° (20 km)	0.24° (27 km)	0.19° (21 km)
LQ4 latitude (0.0)	0.30° (33 km)	0.46° (51 km)	0.62° (69 km)	0.46° (51 km)

1.0, 0.5 and 0.0 correspond to the the best-case scenario, the random-guess scenario and the worst-case scenario for LQ4 tracks, respectively.

HQ2 ship-track precision. Estimated SST uncertainties are still meaningful because the basic SST patterns—including those related to equator-to-pole temperature gradients, boundary currents, gyres, and upwelling regions—are stable features of the ocean circulation (Wunsch (2004)).

SSTs in MURSST are repeatedly sampled in order to estimate uncertainties. For each posterior ship trajectory, we sample SSTs at the realized positions in MURSST on the corresponding high-resolution monthly climatology over 2003–2018. In total we obtain 1,000 posterior ship trajectories, and their corresponding SSTs are sampled. The uncertainties in SSTs are estimated by taking the standard deviation across these 1,000 samples. For purposes of visual display, the uncertainties are regridded at 2° resolution.

Position errors induce uncertainties in SST in regions where position errors are large and SST gradients are strong. On average, position errors in HQ2 tracks translate into 0.11°C SST

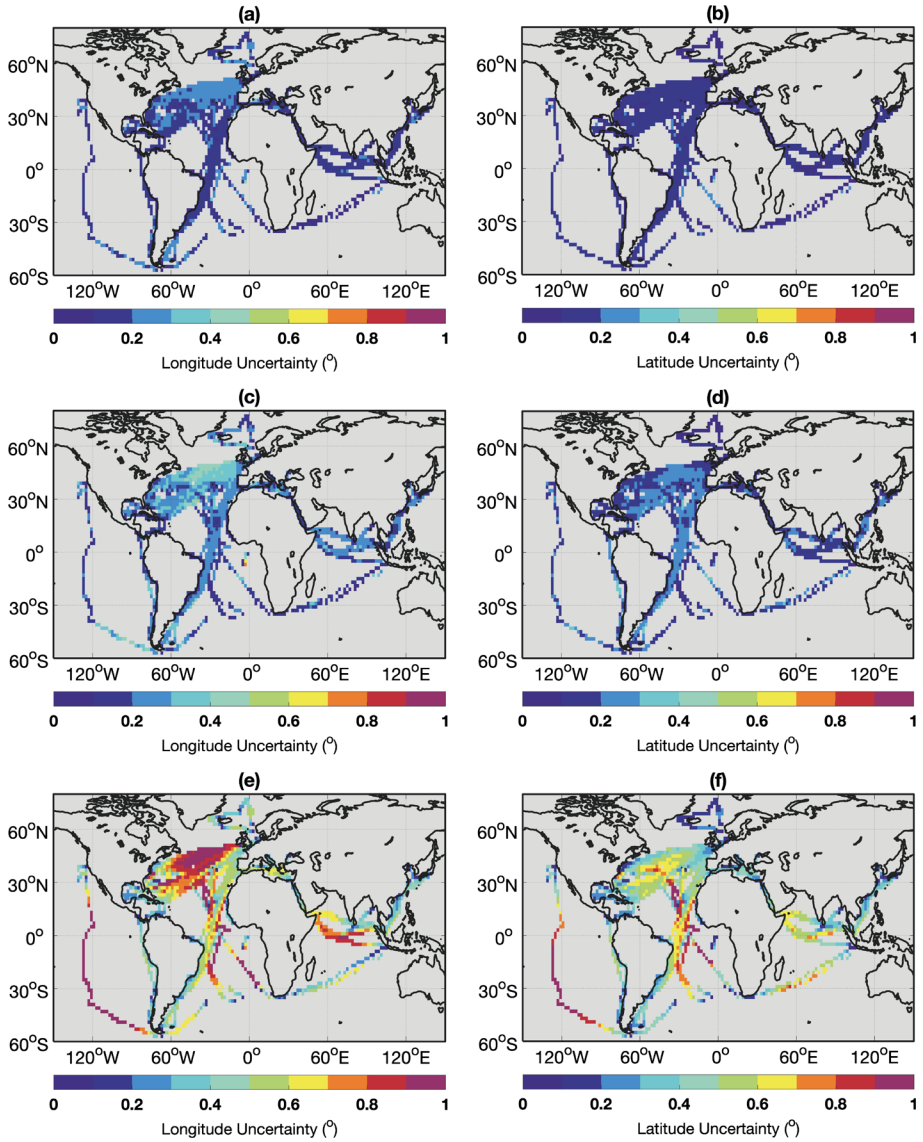

FIG. 6. *Global pattern of random position uncertainties for LQ4 tracks. Left panels show longitude uncertainties for*: (*a*) *best-case*, (*b*) *random-guess and* (*c*) *worst-case scenarios. Right panels are as the left but for latitude uncertainties. In the best-case scenario, the celestial correction happens at each midnight. In the random-guess scenario, midnight positions have a probability of* 0.5 *to be corrected, whereas, in the worst-case scenario, celestial corrections never happen. Maps are shifted to center on the Atlantic for the purpose of visualization. Procedures of generating maps from individual measurements are as per Figure* 5.

uncertainties (see Table 5), but can be as large as 0.33°C in the Northwest Atlantic where the western boundary current detaches from the East Coast of the U.S. (see Figure 7). Larger SST uncertainties are also found in regions where other boundaries currents detach in the Northwest Pacific and Southwest Atlantic as well as in the vicinity of the Aghulas Current south of South Africa. For LQ4 tracks, SST uncertainties average 0.22°C and reach 0.94°C in the midlatitude Northwest Atlantic under the zero-celestial observation scenario, and are similar to HQ2 tracks in the other scenarios (see Table 5 and Figure 8).

5. Concluding remarks. We are unaware of previous efforts to quantify the position uncertainty of historical observations from ships. The magnitude of uncertainties of a given

TABLE 5
A comparison of SST uncertainty and SST offset

Global[a]	SST uncertainty (°C)				SST offset (°C)			
Quantiles	25%	50%	75%	mean	25%	50%	75%	mean
HQ2	0.03	0.06	0.12	0.11	−0.05	0.01	0.07	0.02
LQ4 (1.0)	0.02	0.05	0.09	0.08	−0.01	0.00	0.01	0.00
LQ4 (0.5)	0.03	0.07	0.12	0.10	−0.01	0.00	0.01	0.00
LQ4 (0.0)	0.08	0.16	0.27	0.22	−0.03	0.00	0.03	0.01
Regional[b]	SST uncertainty (°C)				SST offset (°C)			
Quantiles	25%	50%	75%	mean	25%	50%	75%	mean
HQ2	0.08	0.23	0.46	0.33	−0.22	−0.00	0.24	0.03
LQ4 (1.0)	0.11	0.26	0.47	0.32	−0.05	0.00	0.04	−0.01
LQ4 (0.5)	0.17	0.37	0.62	0.43	−0.07	−0.01	0.04	−0.01
LQ4 (0.0)	0.42	0.83	1.29	0.94	−0.30	−0.03	0.25	−0.02

[a]The summary statistics are calculated using SSTs at all positions along ship tracks.

[b]The summary statistics are calculated using SSTs restricted in the regions over the Gulf Stream (280–320°E, 40–50°N and 280–300°E, 35–40°N).

quantity associated with position uncertainties will depend on the local gradient of that quantity, which we estimated for SSTs is on the order of 0.1°C over the globe but can be up to 0.3°C regionally (Figures 7, 8 and Table 5). Expected SST errors from position uncertainties are small, as compared with other error sources. For example, individual SST measurements have random measurement errors of approximately 1.0°C (Kent and Challenor (2006)) and are subject to biases that average approximately 0.4°C cool when measured using can-

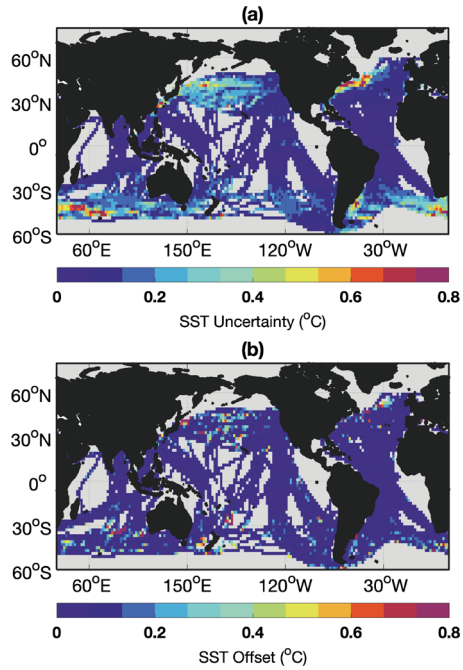

FIG. 7. *SST uncertainties associated with position errors for HQ2 tracks. Individual panels are*: (*a*) *random SST uncertainties and* (*b*) *systematic SST offsets. Results are binned to* 2° *grids for visualization.*

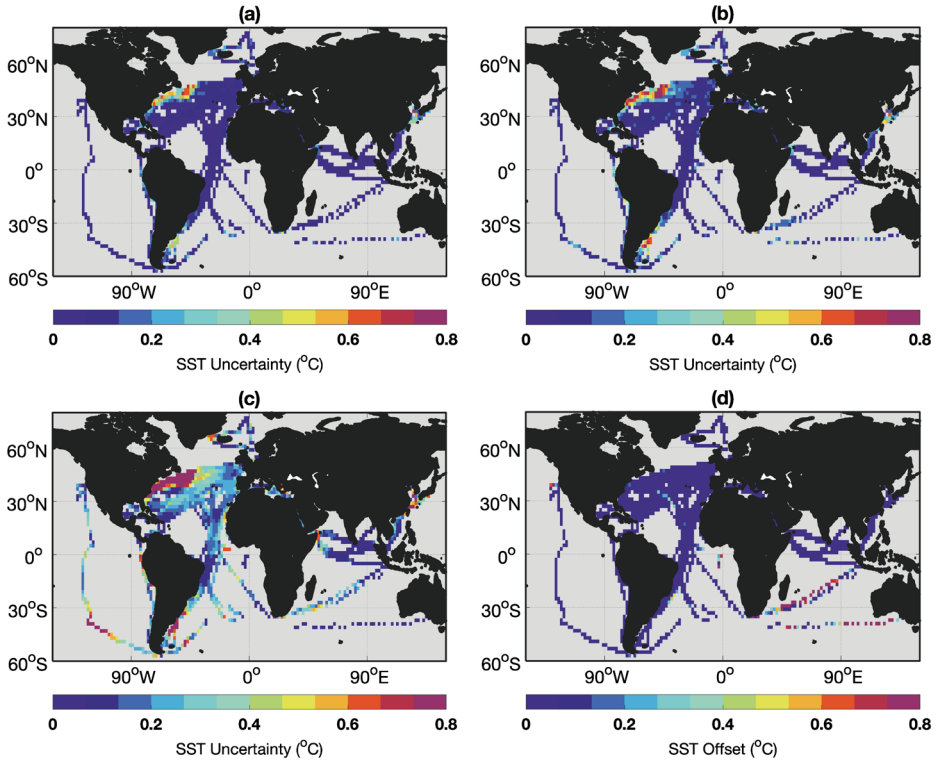

FIG. 8. *SST uncertainties associated with position errors for LQ4 tracks. Panels* (*a*), (*b*) *and* (*c*) *demonstrate random SST uncertainties under the best-case scenario, the random-guess scenario and the worst-case scenario, respectively. Panel* (*d*) *shows SST offsets under the best-case scenario. Results are binned to* 2° *grids for visualization.*

vas buckets and, approximately, 0.2°C warm when collected using the engine-room-intake method (Kennedy et al. (2011b)). Recent studies also identified significant offsets up to 0.5°C associated with national groups by comparing collocated measurements (Chan and Huybers (2019)).

We note, however, that position uncertainties are distinct from other errors in that they will influence all observations made from the same journey, leading to correlated errors along ship tracks. Moreover, position uncertainties will affect the estimation of biases and random SST errors since many existing approaches require pairing collocated measurements (Kent and Challenor (2006), Chan and Huybers (2019)). As noted, position errors can also lead to changes in tail behavior and high-order moments in grid-box averaged SSTs (Director and Bornn (2015)). We suspect that similar concerns arise with respect to other mapping procedures that interpolate using weighted averages of observations, depending upon uncertain positions.

The HQ2 data provides sufficiently frequent and precise observations to characterize uncertainties but may not be indicative of the overall accuracy of SST positions in 1885. We speculate the vessels enrolled in the U.S. meteorological program represent a subset of ships wherein a higher priority was placed upon navigation. Data that is reported with lower resolution and without distinct indications of celestial navigational updates may reflect cruises wherein navigation was a lower priority, or less feasible, given limitations with regard to expertise, equipment or labor. Thus, estimated position errors may not reflect the overall uncertainty of position data in 1885. There will also be heterogeneity in reports among ships for which we have not fully accounted. Some cruises, presumably, had lower need of precise navigation; for example, a zonal cross-Atlantic cruise would have less need of determining

longitude for purposes of ensuring landfall than a cruise with a meridional heading whose intended port was an island. In addition, there are also possibilities that longitude and latitude are not celestially corrected simultaneously for earlier navigators because of less widely deployed ship-board chronometers (Bowditch (1906)).

We focus on a single year in developing and testing our procedure, but it would be useful to extend the analysis over a longer time horizon and to a greater fraction of the data. In 1885, around 85% of observations are associated with ship tracks. Furthermore, Carella, Kent and Berry (2017) have provided estimates of additional data belonging to individual ship tracks, bringing the percentage of observations associated with ship tracks in 1885 to 90%. In more data-rich intervals, however, distinguishing individual ship track becomes more difficult such that between 1900–1940 only 60% of observations, on average, are associated with tracks.

There is presumably a trend toward increasing accuracy of position with time, given technological improvements in marine navigation. This implies that errors and modification of SST distributions introduced through positional error will decrease through time, possibly having consequence for trend estimates, especially those in the vicinity of sharp SST gradients. Position error may limit the spatial resolution over which trends can accurately be determined. It would be useful to estimate uncertainties for a gridded SST product with global coverage that, in addition to accounting for observational SST errors (Kennedy et al. (2011b)) and correcting for biases (Kent et al. (2017), Chan et al. (2019)), also accounts for position errors.

APPENDIX

To evaluate the fitness of the proposed models, we check the following aspects, including the posterior predictive distributions of HQ2 tracks and the posterior distributions of the navigational parameters.

A.1. Posterior predictive check of HQ2 tracks. The posterior predictive check is a self-consistency check in the sense that any replicated data simulated from the posterior predictive distribution should look similar to the observed data.

To generate posterior predictive samples of each HQ2 track, we first sample a set of navigational parameters τ_x, τ_y, τ_s, τ_θ from their posterior samples. Recall that $\mathcal{C} = \{t_1, \ldots, t_m\}$ is the set of time steps when celestial updates happen. For $t \in \mathcal{C}$, that is, the reported ship position contains only celestial observational errors, we sample p_t^x, p_t^y from their posterior samples and sample q_t^x, q_t^y from

$$q_t^x \sim N\big(p_t^x, (\tau_x \cos\psi_t)^2\big), \qquad q_t^y \sim N\big(p_t^y, {\tau_y}^2\big). \tag{A.1}$$

Otherwise, if $t \in (t_k, t_{k+1})$, we follow the dead reckoning navigation, and sample s_t, θ_t, β_k from their posterior samples and sample $\hat{s}_t$, $\hat{\theta}_t$ from

$$\hat{s}_t \sim N\big(s_t, (\tau_s s_t)^2\big), \qquad \hat{\theta}_t \sim N\big(\theta_t + \beta_k, \tau_\theta^2\big). \tag{A.2}$$

Then we generate q_t^x, q_t^y as follows:

$$q_t^x = q_{t-1}^x + 2\hat{s}_t \cos(\hat{\theta}_t), \qquad q_t^y = q_{t-1}^y + 2\hat{s}_t \sin(\hat{\theta}_t). \tag{A.3}$$

Figure 9 shows the posterior predictive distributions of HQ2 track No.30, generated based on 1,000 posterior predictive samples. We see that the empirical means of the posterior predictive samples imitate the observed data. As expected, the position uncertainty follows a quasi-daily pattern.

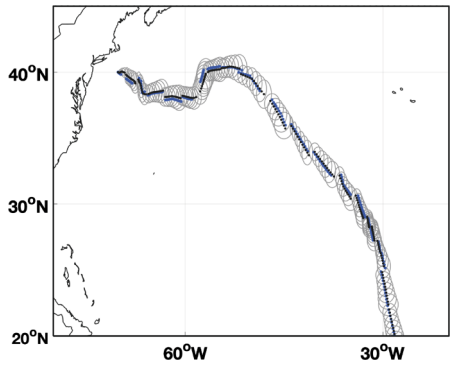

FIG. 9. *Posterior predictive distributions of HQ2 track No.30. The blue dots represent the reported ship positions, the black dots represent the posterior predictive means and the circles calibrate the posterior predictive uncertainties (one standard deviation).*

A.2. Check of navigational uncertainties via a linearized model. The estimates of the navigational parameters play a crucial role in the downstream analysis. We propose an independent linearized model to directly estimate these navigational parameters using HQ2 tracks, and compare the obtained estimates to the results in Table 3.

We first linearize the dead reckoning process by Taylor expansion and omit the second order terms. For $t \notin \mathcal{C}$, we have

$$\begin{aligned}\hat{\lambda}_t^x &= \hat{\lambda}_{t-1}^x + 2s_t(1+e_t^s)\cos(\theta_t + e_t^\theta)\\ &\approx \hat{\lambda}_{t-1}^x + 2s_t\cos\theta_t + 2s_t\cos\theta_t e_t^s - 2s_t\sin\theta_t e_t^\theta,\\ \hat{\lambda}_t^y &= \hat{\lambda}_{t-1}^y + 2s_t(1+e_t^s)\sin(\theta_t + e_t^\theta)\\ &\approx \hat{\lambda}_{t-1}^y + 2s_t\sin\theta_t + 2s_t\sin\theta_t e_t^s + 2s_t\cos\theta_t e_t^\theta,\end{aligned} \tag{A.4}$$

where $\hat{\lambda}_t^x$ and $\hat{\lambda}_t^y$, in units of km, are the reported zonal and meridional displacements of a ship to its position since the last celestial correction. e_t^s and e_t^θ denote the errors in the relative ship speed and ship heading. We assume that e_t^s, e_t^θ follow $N(0,\tau_s^2)$ and $N(0,\tau_\theta^2)$, respectively, where τ_s, τ_θ calibrate the uncertainties in the relative ship speed and ship heading. e_t^s, e_t^θ are assumed to be independent across time. For $t \in \mathcal{C}$, we have

$$\hat{\lambda}_t^x \sim N\big(0,(\tau_x\cos\psi_t)^2\big), \qquad \hat{\lambda}_t^y \sim N\big(0,\tau_y^2\big), \tag{A.5}$$

where τ_x and τ_y calibrate the uncertainties in celestial correction. Combining linearized dead reckoning and celestial correction, we can approximate the variances of jumps in the longitudinal and latitudinal direction by

$$\begin{aligned}\mathrm{Var}(J^x) &= \tau_s^2\Delta x^2 + \tau_\theta^2\Delta y^2 + 2(\tau_x\cos\psi)^2,\\ \mathrm{Var}(J^y) &= \tau_s^2\Delta y^2 + \tau_\theta^2\Delta x^2 + 2\tau_y^2.\end{aligned} \tag{A.6}$$

J^x, J^y denote the jumping distances in the longitudinal and latitudinal direction, and Δx^2, Δy^2 denote the sum of squared distances between consecutive reports from the last celestial update to the position right before the next celestial update. We drop the dependence on t for notational convenience.

We use 20,694 midnight jumps identified from 943 HQ2 tracks to estimate the navigational parameters τ_x, τ_y, τ_s, τ_θ. All the jumps are binned by 20 km × 20 km grids. Within each bin, we obtain the sample variances, $\widehat{V}_x$ and $\widehat{V}_y$, of the jumping distances in the longitudinal and

TABLE 6
Posterior distributions of the navigational parameters based on the linearized model

Quantiles	5%	25%	50%	75%	95%	std
τ_x (km)	30.91	31.12	31.27	31.44	31.67	0.23
τ_y (km)	24.17	24.37	24.50	24.64	24.85	0.21
τ_s (%)	15.07	15.25	15.38	15.49	15.65	0.17
τ_θ (rad)	0.058	0.061	0.063	0.064	0.067	0.003

τ_x and τ_y denote the uncertainty in celestial correction along the longitudinal and the latitudinal direction, respectively. τ_s and τ_θ denote the uncertainty in the relative ship speed and ship heading.

latitudinal direction. Approximately, the sample variances in each bin follow

$$(n-1)\frac{\widehat{V}_x}{\mathrm{Var}(J^x)} \sim \chi^2_{n-1}, \qquad (n-1)\frac{\widehat{V}_y}{\mathrm{Var}(J^y)} \sim \chi^2_{n-1}, \tag{A.7}$$

where n is the sample size in that bin. We set up standard noninformative priors on all the navigational parameters and combine the likelihood specified in equation (A.7) to obtain the posterior distributions.

The results are summarized in Table 6. For dead reckoning, due to the linearization, we obtain smaller estimates of the uncertainties in the relative ship speed and ship heading, compared to the results in Table 3. For celestial correction, we see that the two methods produce consistent estimates of the uncertainties.

A.3. Additional checks. Figure 10 shows the posterior distribution of ship speed s_t for HQ2 track No.30. The blue line represents the empirical ship speed, the black line represents the posterior median and the gray shadow calibrates the first and the third quantiles of the posterior distribution. We note that the unphysical blue peaks correspond to the large celestial updates along the ship trajectory, and the proposed state-space time series model in Section 3.1 helps reasonably smooth out s_t.

Figure 11 demonstrates the distributions of the navigational parameters σ_s and σ_θ[1] across HQ2 tracks, which calibrate the evolutionary uncertainties in ship speed and heading, respec-

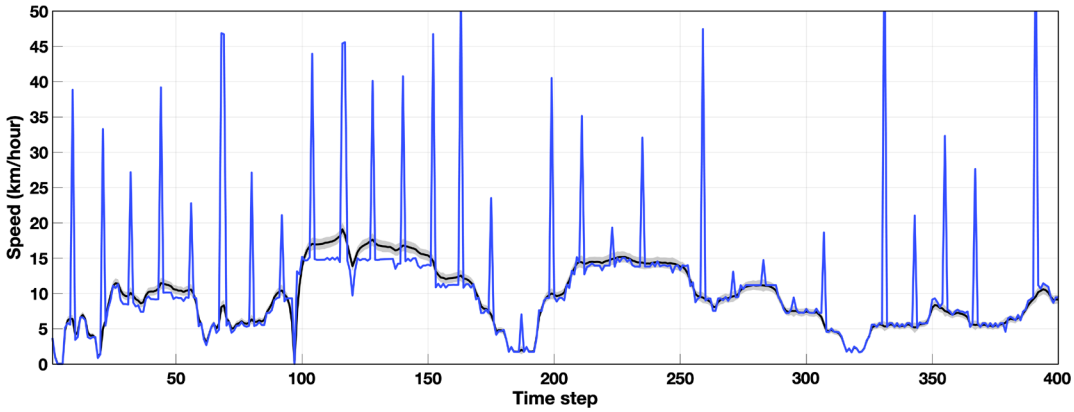

FIG. 10. *The posterior distribution of ship speed s_t for HQ2 track No.30. The blue line represents the empirical ship speed, the black line represents the posterior median, and the gray shadow calibrates the first and the third quantiles of the posterior distribution. The blue peaks correspond to the celestial corrections along the ship trajectory.*

[1]See the definitions of σ_s and σ_θ in equations (3.2), (3.3) and the discussion therein.

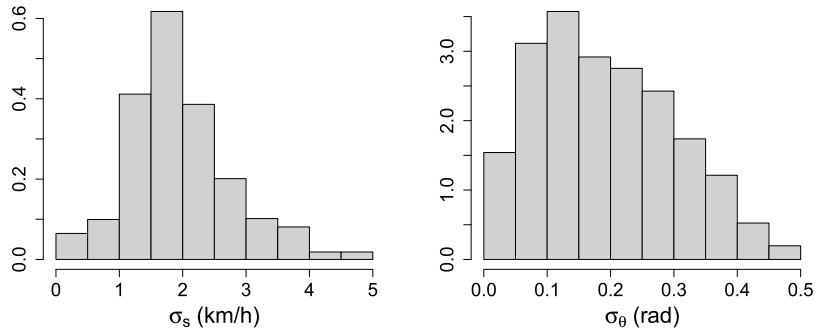

FIG. 11. *The histograms of the posterior means of the navigational parameters σ_s and σ_θ across HQ2 tracks, which calibrate the evolutionary uncertainties in ship speed and heading, respectively.*

tively. We see that σ_s is small enough so that the lower truncation on ϵ_t^s has little effect. In addition, although θ_t is a bounded quantity, σ_θ is also sufficiently small so that there are no unexpected consequences to assume a normal distribution for ϵ_t^θ.

We also summarize the variability of the navigational parameters across HQ2 tracks in Figure 12. In particular, we show the histograms of the posterior means of the navigational parameters τ_x, τ_y, τ_s and τ_θ for HQ2 tracks.

Acknowledgments. Natesh Pillai and Chenguang Dai were supported by the Office of Naval Research. Peter Huybers and Duo Chan were supported by the Harvard Global Institute. Chenguang Dai and Duo Chan contributed equally to this work.

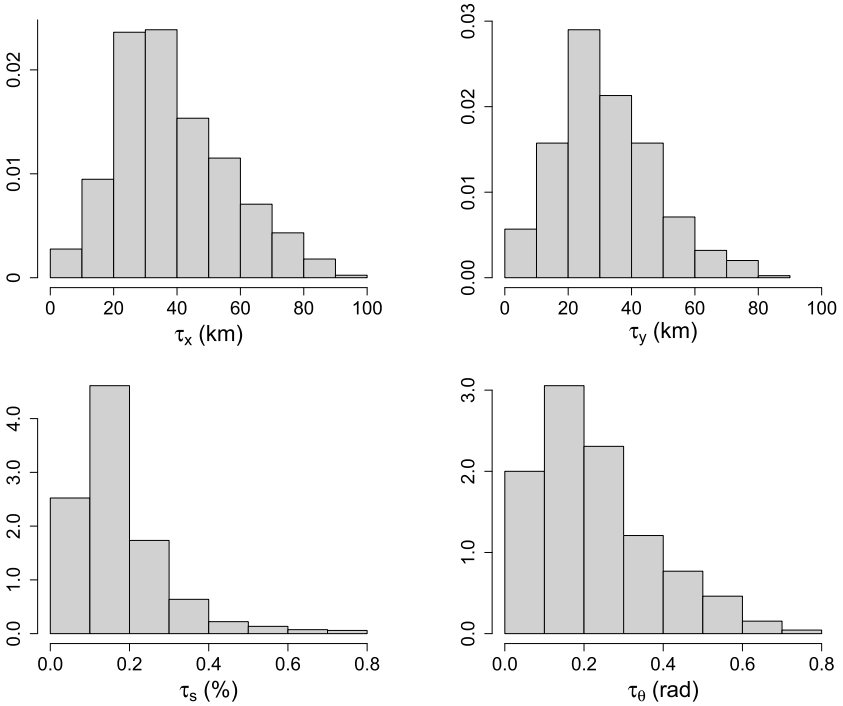

FIG. 12. *The histograms of the posterior means of the navigational parameters across HQ2 tracks. τ_x and τ_y calibrate the uncertainties in celestial correction along the longitudinal and the latitudinal direction, respectively. τ_s and τ_θ calibrate the uncertainties in the relative ship speed $\hat{s}_t/s_t$ and ship heading, respectively.*

REFERENCES

BOWDITCH, J. I. (1906). *American Practical Navigator*. US Government Printing Office. Number 9.

CARELLA, G., KENT, E. C. and BERRY, D. I. (2017). A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.* **37** 2233–2247.

CERVONE, D. and PILLAI, N. S. (2015). Gaussian process regression with location errors. arXiv:1506.08256.

CHAN, D. and HUYBERS, P. (2019). Systematic differences in bucket sea surface temperature measurements amongst nations identified using a linear-mixed-effect method. *J. Climate*.

CHAN, D. and WU, Q. (2015). Attributing observed SST trends and subcontinental land warming to anthropogenic forcing during 1979–2005. *J. Climate* **28** 3152–3170.

CHAN, D., KENT, E. C., BERRY, D. I. and HUYBERS, P. (2019). Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature* **571** 393–397. https://doi.org/10.1038/s41586-019-1349-2

CHIN, T. M., VAZQUEZ-CUERVO, J. and ARMSTRONG, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.* **200** 154–169.

DEE, D. P., UPPALA, S. M., SIMMONS, A. J., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M. A., BALSAMO, G. et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137** 553–597.

DIRECTOR, H. and BORNN, L. (2015). Connecting point-level and gridded moments in the analysis of climate data. *J. Climate* **28** 3496–3510.

FOLLAND, C. (2005). Assessing bias corrections in historical sea surface temperature using a climate model. *Int. J. Climatol.* **25** 895–911.

FREEMAN, E., WOODRUFF, S. D., WORLEY, S. J., LUBKER, S. J., KENT, E. C., ANGEL, W. E., BERRY, D. I., BROHAN, P., EASTMAN, R. et al. (2017). ICOADS Release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.* **37** 2211–2232.

FRIED, W. R. (1977). A comparative performance analysis of modern ground-based, air-based, and satellite-based radio navigation systems. *Navigation* **24** 48–58.

GREGORY, J. M., STOUFFER, R. J., RAPER, S. C. B., STOTT, P. A. and RAYNER, N. A. (2002). An observationally based estimate of the climate sensitivity. *J. Climate* **15** 3117–3121.

HUANG, B., THORNE, P. W., BANZON, V. F., BOYER, T., CHEPURIN, G., LAWRIMORE, J. H., MENNE, M. J., SMITH, T. M., VOSE, R. S. et al. (2017). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate* **30** 8179–8205.

INGLEBY, B. (2010). Factors affecting ship and buoy data quality: A data assimilation perspective. *J. Atmos. Ocean. Technol.* **27** 1476–1489.

KENNEDY, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **52** 1–32.

KENNEDY, J. J., SMITH, R. O. and RAYNER, N. A. (2012). Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies. *Remote Sens. Environ.* **116** 79–92.

KENNEDY, J. J., RAYNER, N. A., SMITH, R. O., PARKER, D. E. and SAUNBY, M. (2011a). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res., Atmos.* **116**.

KENNEDY, J. J., RAYNER, N. A., SMITH, R. O., PARKER, D. E. and SAUNBY, M. (2011b). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res., Atmos.* **116**.

KENT, E. C. and CHALLENOR, P. G. (2006). Toward estimating climatic trends in SST. Part II: Random errors. *J. Atmos. Ocean. Technol.* **23** 476–486.

KENT, E. C., CHALLENOR, P. G. and TAYLOR, P. K. (1999). A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *J. Atmos. Ocean. Technol.* **16** 905–914.

KENT, E. C., KENNEDY, J. J., SMITH, T. M., HIRAHARA, S., HUANG, B., KAPLAN, A., PARKER, D. E., ATKINSON, C. P., BERRY, D. I. et al. (2017). A call for new approaches to quantifying biases in observations of sea surface temperature. *Bull. Am. Meteorol. Soc.* **98** 1601–1616.

MORICE, C. P., KENNEDY, J. J., RAYNER, N. A. and JONES, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res., Atmos.* **117**.

SOBEL, D. (2007). *Longitude*: *The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. Bloomsbury Publishing USA.

STAN DEVELOPMENT TEAM (2019). RStan: The R interface to Stan. R package version 2.19.2.

WOODRUFF, S. D., DIAZ, H. F., ELMS, J. D. and WORLEY, S. J. (1998). COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth* **23** 517–526.

WUNSCH, C. (2004). Gulf Stream safe if wind blows and Earth turns. *Nature* **428** 601. https://doi.org/10.1038/428601c