arXiv:1411.6683v2 [stat.AP] 28 Dec 2014

Bayesian Melding of the Dead–Reckoned Path and GPS Measurements for an Accurate and High–Resolution Path of Marine Mammals

Yang Liu*[1], Brian C. Battaile[2], James V. Zidek[1], and Andrew W. Trites[2]

[1]Department of Statistics, University of British Columbia
[2]Marine Mammal Research Unit, Fisheries Centre, AERL, University of British Columbia

December 30, 2014

Abstract

With recent advances in electrical engineering, devices attached to free–ranging marine mammals today can collect oceanographic data in remarkably high spatial–temporal resolution. However, those data cannot be fully utilized without a matching high–resolution, accurate path for the animal, something that is currently missing in this field. In this paper, we develop a Bayesian Melding approach based on a Brownian Bridge process to combine the fine-resolution but seriously biased Dead–Reckoned path and the precise but sparse GPS measurements; the result is an accurate and high–resolution estimated path together with credible bands as quantified uncertainty statements. We also exploit the properties of underlying processes and some approximations to the likelihood to dramatically reduce the computational burden of handling those big high–resolution data sets.

1 Introduction

The idea of using free–ranging marine animals as platforms to collect oceanographic data, such as temperature and salinity, can be traced back to the discussion in Evans (1970), but only recent advances in electrical engineering make the animal–borne sensors feasible. Miniaturized sensors (tags) can now be attached to the animal and relay data about the environment as well as an animal's movements, behavior, physiology, something which is usually referred to as "bio–logging" (Rutz and Hays, 2009). The oceanographic data collected in bio–logging has successfully filled in the some "blind-spots" in parts of the oceans where little or no other data are currently available (Boehme et al., 2010). For example, animal–borne sensors can collect data at high latitudes or during the winter, where few ship–based measurements are available. When compared to the Argo[1] floats, which can also work in

*Corresponding Author: yang.liu@stat.ubc.ca
[1]http://www.argo.ucsd.edu/About_Argo.html

those situations, the animal-borne sensors can take measurements in the ocean currents or upwelling zones, where the Argo floats will drift away. We also need have little concern that animal-borne sensors can crash into sea floor or ice, since the animals, unlike Argo floats, can "automatically" avoid those obstacles.

There are numerous examples showing how data collected in bio–logging can help improve our knowledge of the ocean's environment. Lydersen (2002) first deployed sensors on white whales to monitor the salinity and temperature structure of an ice–covered Arctic fjord. The data collected by Southern elephant seals helped to identify the Antarctic circumpolar current fronts in the South Atlantic (Boehme et al., 2008) and the seasonal evolution of the upper–ocean adjacent to the South Orkney Islands, Southern Ocean (Meredith et al., 2011). Isachsen et al. (2014) combined Argos floats data with the data collected by hooded seals and so revealed that the Nordic Seas are getting warmer and saltier. Besides the ocean hydrographic data, the originally poorly sampled bathymetry map of the Antarctic continental shelf was improved using data collected by elephant seals (Padman et al., 2010). More examples on the contributions of animal–borne sensors can be found in Boehme et al. (2010).

In most of the studies above, the data were first transmitted to a satellite by the sensor and then downloaded from the satellite for analysis. Due to limitations in the communication bandwidth and cost, a very small amount of data can be transferred in this work–flow and usually we can only obtain data with a low spatial–temporal resolution. For example in Boehme et al. (2008), only two temperature and salinity profiles[2] are obtained per day and the spatial resolution of these data are 20–50km. Such low resolution data markedly restricts the scope and contribution of the scientific findings from the studies with animal–borne sensors (Boehme et al., 2008).

Fortunately some species of animals carrying the sensors, like elephant seals or fur seals, have a relatively fixed "home" (breeding habitat) on an island or sea shore, which is usually the start and end points of their foraging trips. Most sensors deployed in bio–logging are actually attached to the animals on those islands (see e.g. Boehme et al., 2008; Nordstrom et al., 2013), where they can be also retrieved when the animals return. It is thus possible to circle around the bottle neck of satellite communication by storing the data locally on the sensor and then download data after they are retrieved. This approach can provide a nearly continuous record of the sea water temperature as well as variables related the animal's movement and behavior, e.g. diving depth, acceleration etc., with temporal resolution of 1 second or even higher (see e.g. Nordstrom et al., 2013; Wilson et al., 2008).

Such high–resolution bio–logging data has contributed much to our knowledge of animal's behavior and movement pattern (Dowd and Joy, 2011) and energy expenditure (Wilson et al., 2007), but few studies have taken advantage of this high–resolution record of water temperature to achieve more knowledge about the ocean environment or the relationship between the tracked animal and the environment (Nordstrom et al., 2013). The main reason here is the lack of an accurate and high–resolution path (in longitude and latitude) of the tagged animal so we cannot know where the temperature is measured. The same reason also limits our understanding of the relationship between the animal's habitat preference and the environment (Benoit-Bird et al., 2013a,b).

[2]Profile here means a sequence of measurements at different depths

Currently in bio–logging, an animal's location is usually determined by the direct measurements from Global Positioning System (GPS Hofmann-Wellenhof et al., 1993), which can only work when the sensor floats on the water surface and has direct lines of sight to four or more satellites. However, the marine mammals usually spend very little time on the sea surface and the GPS sensor only provides sparse and irregular observations of the animal's path. Some studies today impute the gaps between two GPS observations via linear interpolation (Benoit-Bird et al., 2013b; Nordstrom et al., 2013) or a statistical model fitted to the GPS observations. Examples of the latter approach include the correlated random walk (Jonsen et al., 2005) and continuous time correlated random walk (Johnson et al., 2008). Both the linear interpolation and the more sophisticated models impose certain parametric assumptions that may incur bias in the imputed path.

In order to fill in the gaps between GPS observations, a "Dead–Reckoning" (DR) tag, which usually consists of an accelerometer, a magnetometer, a time–depth–recorder (TDR) and other supporting components, is also deployed on the instrumented animals (see e.g. Nordstrom et al., 2013; Wilson et al., 2007). This DR tag can sample at infra–second frequencies, like 16Hz or 32Hz, so a detailed record of the animal's movement can thus be obtained. After retrieving the tag, we can download the data and use the so–called "Dead–Reckoning Algorithm" (DRA, Elkaim and Decker, 2006; Wilson et al., 2007) to reconstruct an estimated path from the DR tag data. The basic idea of DRA is as follows: estimate the animal's orientation (direction of velocity) from the accelerometer and magnetometer readings by solving the Wahba's problem (Wahba, 1965); next estimate the animal's speed (norm of the velocity) using the TDR data or by assuming a known constant; estimate the path by integrating the velocity from a known starting point. This estimated path is called the "Dead–Reckoned" (DR) path hereafter.

The temporal resolution of the DR path is decided by the sampling frequency of the DR tag, namely, 1/16 or 1/32 of a second and the spatial resolution of it can be as fine as a few meters. The DR path provides remarkably detailed information about the animal's movement, especially fine scale fluctuations that the GPS is not able to capture. However, the DR path can be seriously biased due to measurement error or systematic bias in the DR tag, assumptions and errors in the orientation and speed estimation, and discretization bias in the integration, etc (Wilson et al., 2007). A more detailed review of the DRA is provided in Appendix A. As shown later in our case study, the bias of the DR path can be as large as 100km at the end of a seven–day trip. Therefore, we must correct the bias in the DR path before applying it, to provide an in–situ record of the oceanographic variables as well as to address other biological or ecological questions.

Fortunately, extensive tests have demonstrated the unbiasedness and high precision of the GPS observations with standard errors around a couple of hundred meters (Bryant, 2007; Hazel, 2009) so they can be used to correct the DR path. The conventional correction method in Wilson et al. (2007) can be summarized as follows: denote the DR path (in one dimension) as $x_1, x_2, \ldots, x_T$ at times $t = 1, 2, \ldots, T$ and the GPS observations at times 1 and T are y_1, y_T respectively; assume without loss of generality, that $x_1 = y_1 = 0$ and that the corrected path $\hat{\eta}_t$ is calculated as,

$$\hat{\eta}_t = x_t + \frac{y_T - x_T}{T - 1}(t - 1), \tag{1}$$

which evenly distributes the bias $y_T - x_T$ evenly over the individual time points. The DR

path between two GPS observations is directly shifted to the locations indicated by the GPS observations, namely $\hat{\eta}_1 = y_1$ and $\hat{\eta}_T = y_T$. This procedure is repeated for all the sections separated by the GPS observations to correct the whole path. Wilson et al. (2007) did not provide any justification for this conventional correction method. Also, it fails to take account of the measurement error in the GPS observations, nor does it provide an uncertainty statement about the corrected path. According to Nordstrom et al. (2013), the bio–logging community has concern about the validity of the corrected path and few applications are developed based on it.

In this paper, we use the Bayesian Melding (BM) approach to develop a statistically rigorous method for correcting the bias in each of the geographical coordinates of the DR path. That approach was pioneered by Fuentes and Raftery (2005) to combine the direct observations of air–pollutant level from a sparse network of monitoring stations and the output from deterministic chemical transportation (computer) model outputs, at each pixel of a map based on known pollutant source and geophysical information. The BM approach was later adapted for use in a variety of different fields, such as modeling hurricane surface wind (Foley and Fuentes, 2008), ozone level (Liu et al., 2011), and wet deposition (Sahu, 2010), etc. All these applications have demonstrated the remarkable flexibility and effectiveness of the BM approach. When comparing our application to that in Fuentes and Raftery (2005), the GPS observations play the role of the station measurements while the DR path plays the role of the computer model output. Using the GPS to correct DR path can also be viewed as combining the location information from the GPS and DR path, which is the very strength of BM. For each coordinate the method provides Bayesian credible intervals to quantify the residual uncertainty about the corrected coordinates. The estimated path can then be found by combining in the obvious way, the separate curves for the two geographical coordinates.

Our BM model assumes the animal's path is a Brownian Bridge process and the GPS observations are unbiased observations of this true process with $i.i.d.$ normal measurement errors. The DR path is assumed to be the sum of the true process, a systematic error component that is modeled by a parametric function, and a random error component that is modeled by a Brownian motion process. As discussed in the sequel, these model choices are supported by the biological and ecological literature.

All components in our BM model are Gaussian and linear, which makes inference conceptually easy. However in seeking to retain the benefits of the high–resolution DR path we encounter a big data problem. For example, a typical one–week foraging trip in our case study, with 16Hz sampling frequency, resulted in $T = 7 \times 24 \times 3600 \times 16 = 9,676,800$ time points. The large T rules out the decomposition and storage of the covariance matrix of the T–dimensional true process on most computers. But we show how that high computational burden we can be dramatically reduced by coupling the conditional independence properties of the Brownian Bridge and Brownian Motion processes with some judicious approximations to the likelihood. The computations needed to implement the resulting BM model can be completed on a regular laptop in under five minutes.

To illustration the capability of our BM approach, we apply it to data from two foraging trips of northern fur seals in the summer of 2009 in the Bering sea, observed as part of the Bering Sea Integrated Research Program (BSIERP) (Benoit-Bird et al., 2013b; Nordstrom et al., 2013). Cross–validation studies carried out for model selection show our method to be superior to the conventional bias correction method in Wilson et al. (2007).

The paper is organized as follows. Section 2 introduces our BM model. Section 3 describes how model inference can be carried out efficiently. The real data application together with cross–validation studies is found in Section 4. Section 5 concludes this paper and discusses some future work.

2 Bayesian Melding model

In this section, we adopt the BM approach to combine the information from the accurate but sparse GPS observations and biased but dense DRA results. For simplicity, the two dimensions of the path (latitude and longitude) are dealt with separately. Thus abstractly we are considering a one dimensional path over time which we denote by $\eta(t)$ at discrete time points $t = 1, 2, \ldots, T$. The time unit plays no essential role in the mathematical development our theory. Moreover the approach works just as well with unequally spaced time points, in other words, for arbitrary $t_1, t_2, \ldots, t_T$. But for expository simplicity we work with $1 : T$ as the real bio–logging data are equally spaced. As in the previous BM literature, we take $\eta(t)$ to be a Gaussian process, with

$$\boldsymbol{\eta}(1:T) \sim N(\mathbf{f}(1:T), \mathbf{R}(1:T, 1:T)), \tag{2}$$

where the $\mathbf{f}(\cdot)$ denotes the process mean function and $\mathbf{R}$, its covariance matrix. The notation $\mathbf{f}(1:T)$ stands for the vector $\{f(1), f(2), \ldots, f(T)\}^T$ while $\mathbf{R}(1:T, 1:T)$ is a $T \times T$ covariance matrix with $R(t, t') = \text{Cov}\,(\eta(t), \eta(t'))$. Throughout this paper, bold faced characters are used exclusively to represent vectors or matrices.

Various choices are available for this Gaussian process. A common one (Fuentes and Raftery, 2005; Sacks et al., 1989) assumes that $\mathbf{f}$ to be a simple parametric model, e.g., a constant or a linear function of the covariates, and $R(t, t') = \sigma^2 \rho(|t - t'|)$, where $\rho(\cdot)$ is an isotropic correlation function from a class such as the Matérn or power exponential. However, this popular stationary Gaussian process is not suitable for our application. As noted above, high–resolution data can only be obtained by retrieving the tag, which means that the start and end points of the path are fixed (usually the same location), which is illustrated in Figure 1. Apart for the start and end points, the path is unknown, and hence random in our Bayesian framework. Its variation is relatively large in the middle and small when close to the known start and end points. These features of the path inspire us to model it with a Brownian Bridge process, whose mean and covariance functions are:

$$\begin{aligned} f(t) =& A + (B - A)\frac{t-1}{T-1} \\ R(s,t) =& \sigma_H^2 \frac{(\min(s,t) - 1)(T - \max(s,t))}{(T-1)} \end{aligned}$$

where $\eta(1) = A$ and $\eta(T) = B$ are the known start and end points of the path, while σ_H^2 is the variance parameter. Notice that $R(1, \cdot) = R(\cdot, T) = 0$, in accordance with the known start and end points $\eta(1)$ and $\eta(T)$. Also, $R(t, t)$ increases with t when $t < (T - 1)/2$ and decreases with t for $t > (T - 1)/2$, reflecting the fact that the variation of the path is large in the middle. Another noteworthy property of our covariance matrix $\mathbf{R}$ is its form as the product of a scalar σ^2 and a matrix, the latter depending only on the time points. To clearly represent the parameters of the Brownian Bridge process, we introduce the notation

$$\text{BB}(A, B, T_S, T_E, \sigma^2) \tag{3}$$

for a Brownian Bridge process, which starts from A at time T_S and ends in B at time T_E with a variance parameter σ^2, namely $f(t) = A + (B - A)\frac{t-T_S}{T_E-T_S}$ and covariance function $R(s,t) = \sigma_H^2 \frac{(\min(s,t)-T_S)(T_E-\max(s,t))}{(T_E-T_S)}$.

Our choice of the Brownian Bridge model is well supported in the biology and ecology literature. According to Humphries et al. (2010), marine mammals tend to exhibit Brownian–like movements in environments with abundant food resources, such as the ocean around Bogoslof island where our case study is centered, one believed to be just such a resourceful environment (Benoit-Bird et al., 2013a). Also, a Brownian Bridge model was proposed Horne et al. (2007) to model the habitat use for a wide range of animal species. This model is well accepted by the biology and ecology field and further improved by Sawyer et al. (2009), Kranstauber et al. (2012), and Kranstauber et al. (2014). Many other examples of modeling an animal's path with Brownian Bridge processes can be found in the references of the above papers.

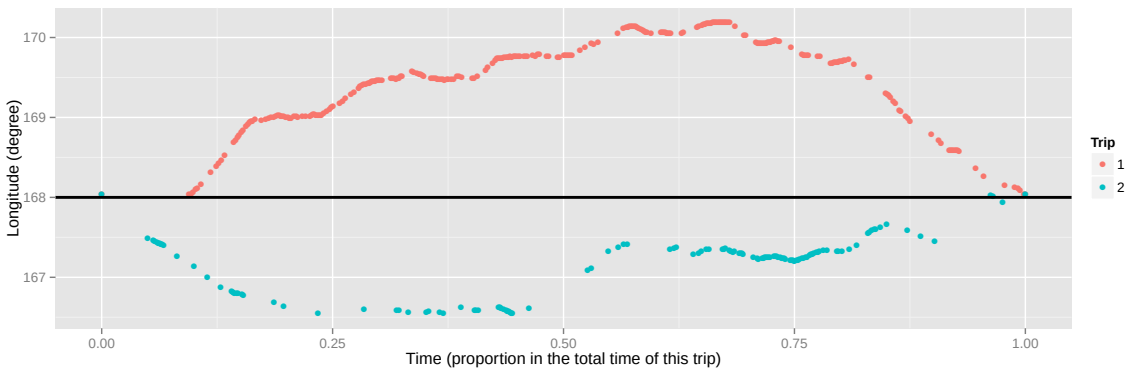

Figure 1: The longitude in the GPS observations of the two trips in our case study of northern fur seals. Both trips started and ended at Bogoslof Island and the horizontal line indicates the longitude $-168.035°E$ of the island. The time unit in this figure is the proportion of the total time of this trip.

The GPS observations of the locations are denoted by $Y(t_k), k = 1, 2, \ldots, K, t_1 = 1, t_K = T, t_k \in \{2, \ldots, T-1\}, k = 2, \ldots, K-2$, which are unbiased observations of the true location:

$$Y(t_k)|\eta(t_k) \overset{\text{iid}}{\sim} N(\eta(t_k), \sigma_G^2), \tag{4}$$

for $k = 2, \ldots, K-2$. The known start and end points assumption implies that $Y(t_1) = \eta(t_1)$, and $Y(t_K) = \eta(t_K)$ are known.

Next, we use $X(t), t = 1, 2, \ldots, T$ to denote the DR path without any error correction. To incorporate the bias of the DR path, we assume:

$$X(t) = \eta(t) + h(t) + \xi(t), \tag{5}$$

where $h(t)$ is a parametric bias term designed to capture trend if any in the DRA path while $\xi(t)$ denotes another Gaussian process independent of $\eta(t)$ that captures any irregular

components in the deviation of the DR path from the truth:

$$\boldsymbol{\xi}(1:T) \sim N(\mathbf{0}, \mathbf{C}(1:T, 1:T)).$$

For the parametric bias component $h(t)$, we have considered various models, e.g. $h(t) = \sum_{i=1}^{Q} \beta_i t^{i-1}$. The residual bias $\xi(t)$ is assumed to be a Brownian motion process (random walk of order 1) whose covariance function is therefore

$$C(s,t) = \sigma_D^2(\min(s,t) - 1).$$

We believe the Brownian motion process to be a reasonable approximation to the gradually accumulated error in the DRA. If we assume the error in the velocity estimates from the DRA at each time point is *i.i.d.* normal, the error in the integrated path is then a Brownian motion.

The final ingredients in our BM model are the prior distributions of the parameters. For notational simplicity, all densities are denoted by square brackets $[\ldots]$ throughout this report. For σ_G^2, we assume a known constant based on the previous extensive tests of the Fastloc® GPS device. The priors of the other two variance parameters are chosen to be the reference priors, $[\sigma_H^2] \propto \frac{1}{\sigma_H^2}$ and $[\sigma_D^2] \propto \frac{1}{\sigma_D^2}$, which is a non–informative prior on the log scale ($[\log(\sigma_H^2)] \propto 1$). For $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_Q\}^T$, a non–informative flat prior $[\boldsymbol{\beta}] \propto 1$ is used. All these parameters are assumed to be independent of each other.

For expository simplicity in describing the joint distribution of all the data and parameters, the following notations are introduced:

- The unknown part of the true path is denoted by $\boldsymbol{\eta} = \boldsymbol{\eta}(2:(T-1))$, a $T-2$ dimensional vector.
- GPS observations of the unknown part of the path are denoted by $\boldsymbol{Y} = \{Y(t_2), Y(t_3), \ldots, Y(t_{k-1})\}^T$, a $K-2$ dimensional vector.
- The DR path $\mathbf{X} = \{\mathbf{X}(2:(T-1)), X(T) - Y(T)\}^T$, a vector of dimension $T-1$;
- For the two unknown variance parameters $\boldsymbol{\phi} = \{\sigma_H^2, \sigma_D^2\}^T$.

The joint likelihood of our model is

$$[\mathbf{X}, \mathbf{Y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\phi}] = [\boldsymbol{\phi}][\boldsymbol{\beta}][\boldsymbol{\eta}|\boldsymbol{\phi}][\mathbf{Y}|\boldsymbol{\eta}][\mathbf{X}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\eta}]. \tag{6}$$

To obtain an estimate of the animal's true path and its uncertainty, we seek the posterior distribution

$$[\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{X}, \mathbf{Y}] = \int \underbrace{[\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{X}, \mathbf{Y}, \boldsymbol{\phi}]}_{(1)} \times \underbrace{[\boldsymbol{\phi}|\mathbf{X}, \mathbf{Y}]}_{(2)} \, \mathbf{d}\boldsymbol{\phi}. \tag{7}$$

Here we also include the $\boldsymbol{\beta}$ term, which can be used to assess the bias of the DRA. The posterior mean, denoted by $\tilde{\eta}(t)$, can be an estimate of the animal's path and the posterior standard error, denoted by $\tilde{\sigma}(t)$ provides an uncertainty statement about the estimated path. The point–wise 95% credible interval for $\eta(t)$ can be constructed as

$$[\tilde{\eta}(t) - 1.96\tilde{\sigma}(t), \tilde{\eta}(t) + 1.96\tilde{\sigma}(t)]. \tag{8}$$

3 Model Inference

To calculate the posterior (7), we first fix the variance parameters $\boldsymbol{\phi}$ and calculate part (1) in Equation (7) and then integrate over the posterior of $\boldsymbol{\phi}$. The first part of this section introduces how the components of (7) can be efficiently evaluated and the second part describes how we approximate the integral. A comparison of our BM approach to the conventional approach is included in Section 3.3.

3.1 Evaluating components of the posterior

For notational simplicity, we use $\langle\cdot|\cdot\rangle$ to denote $[\cdot|\cdot,\boldsymbol{\phi}]$, namely $\langle\boldsymbol{\eta}|\mathbf{X},\mathbf{Y}\rangle = [\boldsymbol{\eta}|\mathbf{X},\mathbf{Y},\boldsymbol{\phi}]$. As we specify our model in a Gaussian and linear fashion, it is straightforward to show that $\langle\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{X},\mathbf{Y}\rangle$ is a multivariate Gaussian density,

$$\langle\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{X},\mathbf{Y}\rangle \propto \exp\left\{-\frac{1}{2}\left[(\boldsymbol{\zeta}-\mathbf{M}_1^{-1}\mathbf{M}_2)^T\mathbf{M}_1(\boldsymbol{\zeta}-\mathbf{M}_1^{-1}\mathbf{M}_2)\right]\right\} \tag{9}$$

where $\boldsymbol{\zeta} = \{\boldsymbol{\beta}^T,\boldsymbol{\eta}^T\}^T$ and $\mathbf{M}_1,\mathbf{M}_2$ are derived in Appendix B.1. Although the multivariate Gaussian posterior makes the inference conceptually easy, calculating its posterior mean $\mathbf{M}_1^{-1}\mathbf{M}_2$ and covariance matrix $\mathbf{M}_1^{-1}$ actually involves a matrix decomposition with computational complexity of order $O(T^3)$, which is a tremendous computational burden when T is large. It is possible to avoid the $O(T^3)$ matrix decomposition with certain sparse matrix techniques together with the Sherman–Morrison–Woodbury formula (Henderson and Searle, 1981), but those techniques still require the storage of some huge matrices and complicated matrix calculations. This pushes us to further reduce the complexity of (9).

It is easily seen that we have more information (data) about $\boldsymbol{\eta}_G \triangleq \boldsymbol{\eta}(t_{1:K})$ where the GPS observations are available than where they are not. For $\boldsymbol{\eta}(1:T\setminus t_{1:K})$, we only have the DR path. So our first step is to break $\boldsymbol{\eta}$ into two sets, that is

$$\langle\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{X},\mathbf{Y}\rangle = \langle\boldsymbol{\eta}(1:T\setminus t_{1:K})|\boldsymbol{\beta},\boldsymbol{\eta}_G,\mathbf{X},\mathbf{Y}\rangle\langle\boldsymbol{\beta},\boldsymbol{\eta}_G,\mathbf{X},\mathbf{Y}\rangle. \tag{10}$$

We can then use the Markovian property of the Brownian Bridge process (see e.g. Stirzaker and Grimmett, 2001) to simplify (10) as:

$$\langle\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{X},\mathbf{Y},\boldsymbol{\phi}\rangle = \left\{\prod_{k=1}^{K-1}\langle\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\eta(t_k),\eta(t_{k+1}),\boldsymbol{\beta},\mathbf{X},\mathbf{Y}\rangle\right\}\times\langle\boldsymbol{\beta},\boldsymbol{\eta}_G|\mathbf{X},\mathbf{Y}\rangle. \tag{11}$$

In this way, we partition the long $\boldsymbol{\eta}$ series into small pieces separated by the GPS observations. We also exploit the Markovian property of the Brownian Motion and find the kth term in the first part of (11) can be simplified as

$$\begin{aligned}\langle\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\eta(t_k),\eta(t_{k+1}),\boldsymbol{\beta},\mathbf{X},\mathbf{Y}\rangle = \\ \langle\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\eta(t_k),\eta(t_{k+1}),\boldsymbol{\beta},\mathbf{X}(t_k:t_{k+1})\rangle.\end{aligned} \tag{12}$$

All the derivations for (11) and (12) are provided in Appendix B.2. In (12), the posterior of $\eta(t)$ between two GPS points can be evaluated only with the corresponding DR path together with the posterior distribution of the two GPS points and $\boldsymbol{\beta}$. This remarkably simple change greatly reduces the memory cost when computing the posterior of the long sequence and

enables us to easily parallelize the whole calculation. Moreover, both Brownian Bridge and Brownian Motion processes conditioned on two end points are Brownian Bridge processes, such that,

$$\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\eta(t_k),\eta(t_{k+1}) \sim \mathrm{BB}(\eta(t_k),\eta(t_{k+1}),t_k,t_{k+1},\sigma_H^2)$$
$$\boldsymbol{\xi}(t_k+1:t_{k+1}-1)|\xi(t_k),\xi(t_{k+1}) \sim \mathrm{BB}(\xi(t_k),\xi(t_{k+1}),t_k,t_{k+1},\sigma_D^2).$$

This fact is exploited to completely avoid the matrix inverse calculation when evaluating (12), which further reduces the computational burden. The derivations are included in Appendix B.3.

However, when evaluating $\langle \boldsymbol{\beta}, \boldsymbol{\eta}_G | \mathbf{X}, \mathbf{Y} \rangle$ in (11), we still need to deal with the long sequence $\mathbf{X}$. However, the $\mathbf{Y}$ is an unbiased observation of $\boldsymbol{\eta}_G$ and therefore $\langle \boldsymbol{\eta}_G | \mathbf{X}, \mathbf{Y} \rangle$ can be well approximated by $\langle \boldsymbol{\eta}_G | \mathbf{Y} \rangle$. This approximation is exceptionally good when $\sigma_D^2 > \sigma_G^2$. For $\boldsymbol{\beta}$, it can be well inferred from the difference between $\mathbf{X}_G \triangleq \mathbf{X}(t_{1:K})$ and $\mathbf{Y}$. In this way, we introduce the following approximation:

$$\langle \boldsymbol{\beta}, \boldsymbol{\eta}_G | \mathbf{X}, \mathbf{Y} \rangle \approx \langle \boldsymbol{\beta}, \boldsymbol{\eta}_G | \mathbf{X}_G, \mathbf{Y} \rangle. \tag{13}$$

With similar arguments, we also approximate the posterior of $\boldsymbol{\phi}$ by,

$$[\boldsymbol{\phi}|\mathbf{X},\mathbf{Y}] \approx [\boldsymbol{\phi}|\mathbf{X}_G,\mathbf{Y}] \tag{14}$$

The explicit expressions for (13) and (14) are included in Appendix B.4. Our simulations which are designed to mimic the real data sets have shown that the impact of the two approximation errors in (13) and (14) is negligible.

In summary, the posterior of $\boldsymbol{\eta}$ is approximated as follows:

$$\begin{aligned}
&[\boldsymbol{\eta},\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}] = \int [\boldsymbol{\eta},\boldsymbol{\beta}|\mathbf{X},\mathbf{Y},\boldsymbol{\phi}][\boldsymbol{\phi}|\mathbf{X},\mathbf{Y}]d\boldsymbol{\phi} \\
&= \int \langle \boldsymbol{\eta}(1:T\setminus t_{1:K})|\boldsymbol{\beta},\boldsymbol{\eta}_G,\mathbf{X},\mathbf{Y}\rangle \langle \boldsymbol{\eta}_G,\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}\rangle [\boldsymbol{\phi}|\mathbf{X},\mathbf{Y}]d\boldsymbol{\phi} \\
&= \int \left\{ \prod_{k=1}^{K-1} \langle \boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\boldsymbol{\beta},\eta(t_k),\eta(t_{k+1}),\mathbf{X}\rangle \right\} \langle \boldsymbol{\eta}_G,\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}\rangle [\boldsymbol{\phi}|\mathbf{X},\mathbf{Y}]d\boldsymbol{\phi} \\
&= \int \left\{ \prod_{k=1}^{K-1} \langle \boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\boldsymbol{\beta},\eta(t_k),\eta(t_{k+1}),\mathbf{X}(t_k:t_{k+1})\rangle \right\} \langle \boldsymbol{\eta}_G,\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}\rangle [\boldsymbol{\phi}|\mathbf{X},\mathbf{Y}]d\boldsymbol{\phi} \\
&\approx \int \left\{ \prod_{k=1}^{K-1} \langle \boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\boldsymbol{\beta},\eta(t_k),\eta(t_{k+1}),\mathbf{X}(t_k:t_{k+1})\rangle \right\} \langle \boldsymbol{\eta}_G,\boldsymbol{\beta}|\mathbf{X}_G,\mathbf{Y}\rangle [\boldsymbol{\phi}|\mathbf{X}_G,\mathbf{Y}]d\boldsymbol{\phi}.
\end{aligned} \tag{15}$$

Next, we will carry out the integration in Equation (15).

3.2 Integration over the variance parameters $\boldsymbol{\phi}$

According to the BM literature (Liu et al., 2011), the integration in (15) can be carried out by MCMC. However, we need to avoid the heavy computational burden of the MCMC

techniques in our application and it may not be practical to store the MCMC samples of the high dimensional parameter $\boldsymbol{\eta}$. The first alternative is to avoid the integration via the empirical Bayesian approach as in Casella (1985):

$$[\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{X}, \mathbf{Y}] \approx [\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{X}, \mathbf{Y}, \hat{\boldsymbol{\phi}}],$$

where $\hat{\boldsymbol{\phi}}$ is the maximum likelihood estimate of $\boldsymbol{\phi}$ after $\boldsymbol{\eta}, \boldsymbol{\beta}$ are marginalized out,

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \log([\boldsymbol{\phi}|\mathbf{X}_G, \mathbf{Y}]). \tag{16}$$

The empirical Bayesian approach is computationally simple, especially when we can explicitly evaluate the marginal likelihood. However, it fails to reflect the uncertainty in $\boldsymbol{\phi}$ and thus it underestimates the uncertainty in the posterior of $\boldsymbol{\eta}$. To overcome this issue, we use a numerical integration method like that in INLA (integrated nested Laplace approximation, Rue et al. (2009)), which approximates the integration on a grid decided by the likelihood surface.

Let $\mathbf{H}$ denote the 2×2 Hessian matrix of $\hat{\boldsymbol{\phi}} = \{\hat{\sigma}_H^2, \hat{\sigma}_D^2\}^T$ in (16) and $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$. With the eigenvalue decomposition $\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T$, the space of $\boldsymbol{\phi}$ can be explored by $\boldsymbol{\phi}(\mathbf{z}) = \hat{\boldsymbol{\phi}} + \mathbf{A}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$, where $\mathbf{z}$ is a 2×1 vector. To find the grid for numerical integration, we start from $\mathbf{z} = \mathbf{0}$ and search in the positive direction of z_1, that is, we increase $j \in \mathbb{N}^+$ and $\mathbf{z} = (j\delta_z, 0)$ as long as

$$\log([\boldsymbol{\phi}(\mathbf{0})|\mathbf{X}_G, \mathbf{Y}]) - \log([\boldsymbol{\phi}(\mathbf{z})|\mathbf{X}_G, \mathbf{Y}]) < \delta_\pi,$$

where δ_z is the step size and δ_π controls the magnitude of probability mass that will be included in the numerical integration. After searching on the positive side, we switch direction and search on the negative side of z_1. This procedure is repeated for both dimensions of $\mathbf{z}$. For our BM model above, if the search stops at J_1^+, J_2^+ steps in the positive directions of z_1 and z_2 respectively and J_1^-, J_2^- in their negative directions, a grid of size $(J_1^+ + J_1^- + 1) \times (J_2^+ + J_2^- + 1)$ is used in the numerical integration and the points on this grid are $\mathbf{z}_{j_1,j_2} = \delta_z(j_1, j_2)$, with $j_1 \in (-J_1^-, -J_1^- + 1, \ldots, 0, \ldots, J_1^+)$ and $j_2 \in (-J_2^-, -J_2^- + 1, \ldots, 0, \ldots, J_2^+)$. The integral in (15) is approximated by

$$[\boldsymbol{\eta}, \boldsymbol{\beta}|\mathbf{X}, \mathbf{Y}] \approx \sum_{j_1=-J_1^-}^{J_1^+} \sum_{j_2=-J_2^-}^{J_2^+} w_{j_1,j_2} \times [\boldsymbol{\eta}, \boldsymbol{\beta}|\boldsymbol{\phi}(\mathbf{z}_{j_1,j_2}), \mathbf{X}, \mathbf{Y}], \tag{17}$$

where

$$w_{j_1,j_2} = \frac{[\boldsymbol{\phi}(\mathbf{z}(j_1, j_2))|\mathbf{X}_G, \mathbf{Y}]}{\sum_{j_1} \sum_{j_2} [\boldsymbol{\phi}(\mathbf{z}(j_1, j_2))|\mathbf{X}_G, \mathbf{Y}]}.$$

Here (17) resembles Equation (5) of Rue et al. (2009). As the $\boldsymbol{\eta}, \boldsymbol{\beta}$ conditioning on $\boldsymbol{\phi}$ follows a multivariate Gaussian distribution, the posterior $[\boldsymbol{\eta}, \boldsymbol{\beta}|\mathbf{X}, \mathbf{Y}]$ can be approximated by a mixture of multivariate Gaussian densities, whose mean and variance can be easily calculated. The detailed expressions are in Appendix B.6.

3.3 Comparison to the conventional bias correction method

When compared to the conventional bias correction method in (1), our BM approach can account for both data and model uncertainty and provide a CI for the estimates of the animal's path. Moreover, there is an interesting connection between the BM posterior mean $\tilde{\eta}$ and the conventional corrected path $\hat{\eta}$ in Equation (1). Without loss of generality, let $K = 2$ (no GPS observations except the known start and end points) and $Y(1) = X(1) = 0$. So (1) can be written as

$$\hat{\eta}(t) = Y(T)\frac{t-1}{T-1} + \left[X(t) - X(T)\frac{t-1}{T-1}\right].$$

For BM, if $h(t) = 0$ for all t and $\boldsymbol{\phi}$ is known, the posterior mean $\tilde{\boldsymbol{\eta}}$ under the above assumptions can be simplified into

$$\tilde{\eta}(t) = Y(T)\frac{t-1}{T-1} + \frac{\sigma_H^2}{\sigma_H^2 + \sigma_D^2}\left[X(t) - X(T)\frac{t-1}{T-1}\right].$$

The first parts of $\hat{\eta}(t)$ and $\tilde{\eta}(t)$ represent linear interpolations between $Y(1) = 0$ and $Y(T)$, which determines the basic trend of the animal's path between two known points. The second part is a "bridge" constructed by the $X(t)$, which starts at $X(1) = 0$ and ends at $0 = X(T) - X(T)\frac{T-1}{T-1}$. This bridge can be treated as the "detail" for the animal's path, which is then added to the basic trend of the first part.

The difference between the conventional method and the simplified BM approach is the weight on the "detail". In the conventional approach, the "detail" is directly added to the basic trend while BM shrinks the detail by a factor of $\rho = \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2}$. According to our model, we cannot distinguish between $\eta(t)$ and $\xi(t)$ in $X(t)$ at those non–GPS points, as we only observe the sum of them, but we know that $\eta(t)$ accounts for the σ_H^2 part of the total $\sigma_H^2+\sigma_D^2$ variance (they are both of mean zero after $Y(T)\frac{t-1}{T-1}$ is removed). In this way, a fraction $\rho = \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2}$ of the detail is treated as signal in $\eta(t)$ and added to the basic linear trend.

Notice here we only compare the most simplified BM approach to the conventional approach. In practice, the BM $\tilde{\eta}$ is far more complicated than the form shown above with the parametric part from $\boldsymbol{\beta}$ and the integration over $\boldsymbol{\phi}$. According to our simulations and cross–validation of real data in Section 4.2, our $\tilde{\eta}$ is a substantial improvement on the conventional estimate $\hat{\eta}$.

4 Case studies

The proposed BM approach was applied to two data sets collected in the 2009 Bogoslof Island northern fur seal study (Benoit-Bird et al., 2013b). These two data sets, denoted by "Trip 1" and "Trip 2" came from two trips of different female seals. The first trip lasted about 6 days and second trip lasted about 7 days. The GPS device in this study was programmed to make one observation attempt every 15 minutes but a large fraction of these attempts failed to obtain a valid observation of the coordinates. For Trip 1, 274 GPS observations were available (including the start and end points), with an average gap of 36 minutes between two consecutive GPS observations. 130 GPS observations were available for Trip 2, with an average gap of 80 minutes between two consecutive GPS observations. It is noteworthy that

these GPS observations are not regularly spaced in time and the sample quantiles of the time gaps between two consecutive observations are summarized in Table 1. For the DR path, the DR tag was originally programmed to sample at 16Hz, but the DRA was only performed for the thinned data at 1Hz, which was believed to be sufficiently fine for the purpose of tracking the animals (Benoit-Bird et al., 2013b).

Table 1: The sample quantiles in minutes of the time gaps between two consecutive GPS observations in our case study.

	Min	10%	25%	50%	75%	90%	Max
Trip 1	14.75	15.00	15.45	18.40	31.68	82.79	953.65
Trip 2	14.75	15.00	15.05	30.00	113.05	130.77	698.47

4.1 BM analysis with a constant bias term

As one illustrative example of our approach, we let $\boldsymbol{\beta} = \beta_0 = h(t)$, for the parametric bias part in (5). The GPS observations of longitude and latitude are projected onto a plane in a point–wise fashion as in Wilson et al. (2007), such that the distance between any two consecutive GPS observations on this plane is their great circle distance and the angle between the line connecting the two points and the y–axis (latitude direction) equals the initial Bearing[3] between them. The projected x–direction and y–direction in kilometres are called "Easting" and "Northing" to distinguish them from the longitude and latitude in degrees. As the two dimensions are analyzed separately, we have four data sets in total, which are referred to as "Trip 1 Northing", "Trip 1 Easting", "Trip 2 Northing", and "Trip 2 Easting".

The GPS measurement variance in all of the analysis that follows was fixed at $\sigma_G^2 = 0.0625$, which was chosen based on Bryant (2007) and the average observed number of satellites in those two trips. For the numerical integration part, $\delta_z = 1$ and $\delta_\pi = 3$ were set according to suggestions in Rue et al. (2009), which often led to around 33 grid points in our data sets. All the computations required to obtain the posterior mean and variance for one data set can be done under 5 minutes of wall clock time on a regular laptop. The empirical Bayesian estimates of $\boldsymbol{\phi}$ are summarized in Table 2.

In Figure 2, we plot the GPS observations and DR path from Trip 1 latitude (Northing) together with our BM posterior mean and 95% credible interval. The bias of the DR path dramatically increases with time and it is around 100KM at the end of this trip. Nonetheless, the DR path incorporates some fluctuations that match those seen in the fluctuations of the GPS observations. Our BM approach successfully moves the DR path results to the correct position taken to be the one indicated by the GPS points, while keeping the fluctuations of the DR path to reflect the animal's fine scale movements.

The scale of Figure 2 is a magnified plot that more fully displays how our approach works in the fine scale. Therefore, Figure 3 zooms into the 2000–2400 minute portion of Figure 2. The conventional bias correction from Wilson et al. (2007) and linear interpolation between GPS observations are also included. The posterior mean from BM appears to be a shrunken version of the conventional bias correction, where the bumps in the conventional method are

[3]Initial Bearing, also known as the forward azimuth, denotes the angle between the great circle connecting the two points on the earth and the north direction at the former point in time.

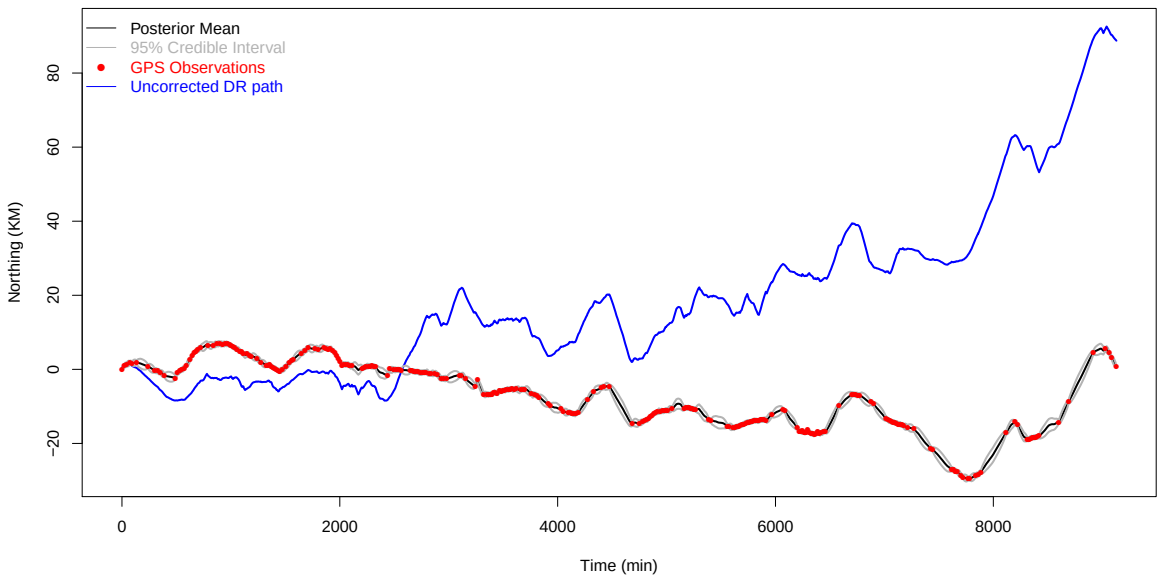

Figure 2: The GPS observations and DRA outputs of the Trip 1 Northing data set together with the BM results for it. The red points are the GPS observations. The blue curve is the uncorrected DR path. The black curve is the posterior mean of $\boldsymbol{\eta}$ from our BM model. The grey curves connect the 95% point–wise credible intervals at all the time points.

damped to the straight line connecting the two GPS points. This verifies our findings in Subsection 3.3.

Moreover, the CI for η in Figure 3 clearly displays a "bridge" structure. Namely, the CI is narrow when $\eta(t)$ is close the GPS observations and becomes wider in between two GPS points. This is plausible as we have direct observations of the path at the GPS points but less accurate information from the DRA for points in between GPS points. Plots obtained by analyzing the other three data sets are quite similar to those seen here in substance and hence are omitted for brevity.

In Table 2, we also summarize the averaged posterior standard errors (APSEs) of the $\eta(t), t = 1, 2, \ldots, T$ found in the four data sets. For the first trip, the APSE is around 500 meters in both dimensions, while it is around 1.3KM for the second trip. The differences in the averaged SEs are mainly decided by the gap between GPS observations. The longer the gap, the less accurate the corrected path is.

4.2 Cross validation comparisons

We also did a leave-one-out cross validation (LOOCV) analysis to further assess our BM approach. Here one GPS observation was deleted at a time when the model was trained. The difference between true observation and the prediction from our model was calculated and summarized as the cross validation root mean squared error (CV-RMSE). We also checked whether this observation is covered by the 95% posterior CI of $\eta(t)$, which is summarized as the coverage percentage. The LOOCV was also performed for the conventional method

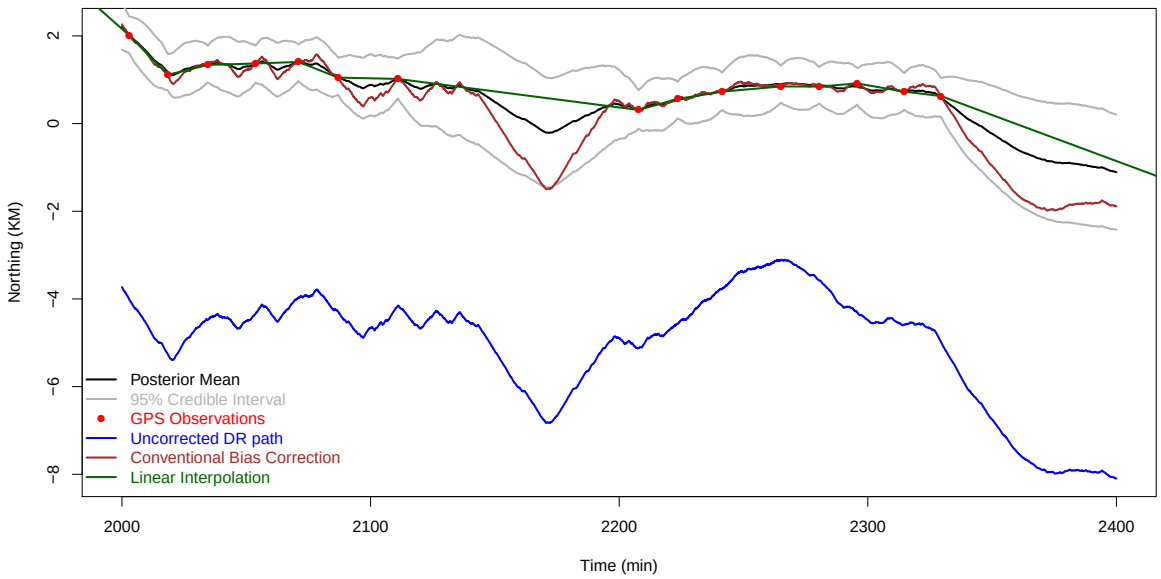

Figure 3: Zoom into the 2000-2400 mins part of Figure 2 (Trip 1 Northing data and BM results). The blue curve is the uncorrected DR path. The black curve is the posterior mean of $\boldsymbol{\eta}$ from our BM model. The grey curves connect the 95% point-wise credible intervals at all the time points. The conventional bias correction in Equation (1) is also included as the brown curve.

and linear interpolation (connecting two consecutive GPS by a straight line), where only the CV-RMSE was obtained. The results are presented in Table 3.

From Table 3, we see that the actual coverage of our 95% CIs is higher than the nominal level, which are further discussed in 4.4. On the other hand, it is easy to see that our BM approach has a smaller CV-RMSE than the linear interpolation method in all the four data sets while the conventional method has a larger CV-RMSE than the linear interpolation in two out of the four data sets, which indicates that the conventional method fails to use the information from the DR path appropriately. In this comparison of our BM to the conventional method, the CV-RMSE of our BM approach is smaller than those from the conventional method by around 1/3 in three of the four data sets. For Trip 2 Easting, the CV-RMSE of our method is slightly larger than that of the conventional method. This might be caused by the fact that $h(t) = \beta_0$ failed to model the bias of DRA well and inspired us to consider different parametric models for $h(t)$ of (5).

4.3 Comparison of different models of $h(t)$ for the Trip 2 longitude data set

Here polynomials $h(t) = \sum_{i=1}^{Q} \beta_i t^{i-1}$ up to order 6, $(Q = 0, 1, 2, \ldots, 6)$ were considered for Trip 2 Easting. The empirical Bayesian parameter estimates $\hat{\sigma}_H^2$ and $\hat{\sigma}_D^2$ under these models are in Table 4. It is interesting to find that estimates of σ_H^2 almost stay the same when Q increases. This is plausible as it is the parameter for the animal's Brownian Bridge movement, which is mainly decided by the GPS observations. On the other hand, the estimates of σ_D^2

Table 2: Empirical Bayesian estimates of $\boldsymbol{\phi}$ $(\hat{\sigma}_H^2, \hat{\sigma}_D^2)$ and the averaged posterior standard error (APSE) in kilometre (KM) of $\eta(t)$ in the BM analyses of the four data sets from our case study.

Data	$\hat{\sigma}_H^2$	$\hat{\sigma}_D^2$	APSE (KM)
Trip 1 Easting	0.0801	0.0353	0.5243
Trip 1 Northing	0.0267	0.0428	0.4438
Trip 2 Easting	0.1031	0.0767	1.2880
Trip 2 Northing	0.1029	0.1233	1.4510

Table 3: Results from the leave-one-out cross validation studies of the four data sets in our case study. The first column is the actual coverage percentage for the BM posterior 95% CI. The cross validation root mean squared errors (CV-RMSE) in kilometres (KM) from our BM approach, conventional method from Wilson et al. (2007) (Equation (1)), and linear interpolation are in the last three columns.

	BM Approach		Conventional Method	Linear Interpolation
Data Set	Coverage	CV-RMSE	CV-RMSE	CV-RMSE
Trip 1 Easting	99.3%	0.374	0.485	0.433
Trip 1 Northing	98.2%	0.386	0.496	0.509
Trip 2 Easting	99.2%	1.027	1.022	1.467
Trip 2 Northing	100%	0.849	1.148	1.061

decrease with increases in Q, as less randomness remains to be explained by the Brownian motion process when more flexible parametric models are characterized by $h(t)$. The smaller σ_D^2 also yields a larger $\rho = \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2}$, which indicates that more confidence can be placed on the DR path $X(t)$ and we thus get a more accurate $\eta(t)$. This explains why the posterior averaged SE decrease with increases in Q.

We also compared the above models by means of the LOOCV and summarize the results in Table 4. Clearly, the prediction accuracy of our model is improved by certain parametric models for $h(t)$. Models with $Q > 1$ have smaller CV-RMSE than the conventional method (1.022 as from Table 3). When $Q = 3$, the CV-RMSE is minimized. The coverage of the posterior CI varies slightly among those models and they are higher than the nominal coverage level.

4.4 Are the CI's conservative? Leave-5-out Cross Validations

In the above LOOCV, the actual coverage percentage of the CI is higher than the nominal level, which brings the concern that the CI is conservative. However, LOOCV just evaluated the performance of our method in a short period without GPS observations,(e.g. mostly around 30 mins) while the time gaps between GPS observation in our case study can be much long than that, as shown in Table 1. In order to assess the performance of our method in a longer period without GPS observations, we performed leave-5-out cross validations (L5OCV) of the same data sets. The gaps created by L5OCV were around 90min (Trip 1) and 150min (Trip 2), which are larger than the original 90% sample quantiles of the gaps in Table 1.

Table 4: Comparisons of different models for $h(t)$ in (5) for Trip 2 Easting (longitude) data. The first three columns of this table summarize the empirical Bayesian estimates of σ_H^2, σ_D^2 and the averaged posterior SE (APSE) in kilometre (KM). The last two columns are the cross-validation root mean squared error (CV-RMSE) and actual coverage percentage of the BM posterior 95% CI in the leave-one-out cross validation.

Q	$\hat{\sigma}_H^2$	$\hat{\sigma}_D^2$	APSE (KM)	CV-RMSE	Coverage
0	0.1032	0.0827	1.314	1.035	0.992
1	0.1032	0.0767	1.288	1.027	0.992
2	0.1031	0.0693	1.252	1.018	0.984
3	0.1029	0.0458	1.101	0.969	0.984
4	0.1029	0.0462	1.104	0.969	0.984
5	0.1029	0.0447	1.093	0.981	0.992
6	0.1028	0.0379	1.035	1.005	0.977

Thus L5OCV offered a more realistic evaluation of our method. Similarly as in LOOCV, five observations were left out at a time when the model was trained the difference between the observations and our model predictions were calculated. The results are summarized in Table 5. From this table, we can easily verify that the coverage percentage are around the nominal level and thus the CI's from our BM are not conservative. The CV-RMSE's from Table 5 once again demonstrate the superiority of our approach over the conventional method and linear interpolation.

Table 5: Results from the leave-five-out cross validation studies of the four data sets. The organization of this table is the same as Table 3.

	BM Approach		Conventional Method	Linear Interpolation
Data Set	Coverage	CV-RMSE	CV-RMSE	CV-RMSE
Trip 1 Easting	97.8%	0.726	1.039	1.131
Trip 1 Northing	95.9%	0.803	1.248	1.156
Trip 2 Easting	93.7%	2.516	2.674	3.839
Trip 2 Northing	92.9%	3.061	3.326	4.436

5 Conclusions

In this paper, we have developed a Brownian Bridge based Bayesian Melding approach to combine the sparse but accurate GPS observations with the high-resolution but biased DR path for the tracking of marine mammals. The posterior mean from our BM approach offers an accurate and high-resolution path of the tracked animals and the posterior credible intervals provides a reasonable statement of the uncertainty in our inferences. Moreover, our approach exploits the conditional independence property of the Brownian Bridge and Brownian Motion to dramatically reduce the heavy computational burden involved in dealing with large data sets. This work enables us to obtain a high–resolution in–situ record of the hydrographic data collected by the marine mammals, which may help to broaden our knowledge about parts of the ocean that are originally hard to observe and better address

the recent changes in the global climate. Besides the contribution to environmental studies, our BM also can serve as a foundation for many biological and ecological questions such as the animal's habitat preference and resource selection (Hooten et al., 2013). Many aspects of the BM approach can be also further improved, such as analyzing the two dimensions of the path simultaneously and developing simultaneous credible intervals.

Appendix A Details of the Dead-Reckoning Algorithm

There are different versions of the DRA depending on the device and the subjects been tracked, but the main steps (Elkaim and Decker, 2006; Wilson et al., 2007), can be summarized as follows:

1. Obtain the earth's gravity vector $\mathbf{g}$ and magnetic field vector $\mathbf{m}$ at the locations being studied.

2. Correct the systematic bias in the accelerometer and magnetometer (Grewal et al., 2007).

3. Smooth the accelerometer and magnetometer readings by a running mean or a low-pass filter.

4. Find the animal's orientation (often taken to be the tag's orientation, namely the three-dimensional direction of velocity) by solving Wahba's problem (Wahba, 1965), which is to solve for the rotation matrix $\mathbf{O}$ that minimizes the distance between the rotated gravity and magnetic force ($\mathbf{Og}$ and $\mathbf{Om}$) and the smoothed accelerometer reading $\tilde{\mathbf{a}}$ and magnetometer readings $\tilde{\mathbf{m}}$, namely,

$$\min_{\mathbf{O}} \left\{ ||\tilde{\mathbf{a}} - \mathbf{Og}||^2 + ||\tilde{\mathbf{m}} - \mathbf{Om}||^2 \right\}.$$

5. Obtain the animal's speed (norm of the velocity vector) by one of the following approaches:

   (a) Assume it is a known constant.

   (b) Assume it is measured by a speed meter (a wheel or paddle), which could only measure the speed of the animal with respect to the water, but not to the earth.

   (c) From the data recorded by the TDR, calculate the velocity in the depth direction v_z. Given the animal's orientation $\mathbf{o} = \{o_x, o_y, o_z\}$, the three dimensional velocity is then $\mathbf{v} = \mathbf{o}v_z/o_z$.

6. Starting from a known point, integrate the velocity to obtain the animal's trajectory.

Although the DRA has been used for years and the data collecting devices are gradually being improved by electrical engineers, there is still huge bias in the DRA results. As shown later in this section, the bias of the DRA in our case study can be as large as 100km at the end of a seven-day trip. According to our analysis and some simulation studies, the bias of the DRA mainly owes to the following things:

1. Orientation bias: Although the gravity $\mathbf{g}$ can be precisely obtained, we can only obtain $\mathbf{m}$ on a relatively sparse grid on the earth at a certain time point (Wilson et al., 2007) rather than at detailed location and time points during the animal's trip. The imprecise magnetic vector introduces bias into the solution to Wahba's problem.

   Moreover, an unbiased solution to the Wahba's problem (Step 4 above) requires the smoothed accelerometer reading $\tilde{\mathbf{a}}$ to be the projection of the gravity only. This is

unlikely to be done by the smoother because the accelerometer reading is the projection of the sum of all the external forces, including the gravity, acceleration from the animal and forces from wave, etc. Similarly, the bias correction methods may fail to remove the systematic bias in the accelerometer and magnetometer readings (Grewal et al., 2007).

2. Speed bias: The first two approaches above clearly involve some unrealistic assumptions. As for the third approach, it fails to work when the animal is floating on the surface (o_z is zero).

3. Discretization bias: Even if both biases above could be removed, we must still discretely approximate an integral in the continuous time domain. Although this problem can be alleviated by increasing the sampling frequency, that could reduce the tag's battery life and limit the scale of the study.

Appendix B Details of the Bayesian Melding Model

This section include the detailed derivations used in the inference of our BM model.

B.1 Explicit expression of $\langle \boldsymbol{\beta}, \boldsymbol{\eta} | \mathbf{X}, \mathbf{Y} \rangle$

Following the model and notations in (2) to (7), it is easy to find

$$\begin{aligned}\langle \boldsymbol{\beta}, \boldsymbol{\eta} | \mathbf{X}, \mathbf{Y} \rangle \propto & \langle \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{X}, \mathbf{Y} \rangle = \langle \boldsymbol{\eta} \rangle \langle \mathbf{Y} | \boldsymbol{\eta} \rangle \langle \mathbf{X} | \boldsymbol{\beta}, \boldsymbol{\eta} \rangle \\ \propto & |\mathbf{R}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\eta} - \mathbf{f})^T \mathbf{R}^{-1} (\boldsymbol{\eta} - \mathbf{f}) \right\} \times |\mathbf{D}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{G}\boldsymbol{\eta})^T \mathbf{D}^{-1} (\mathbf{Y} - \mathbf{G}\boldsymbol{\eta}) \right\} \\ & \times |\mathbf{C}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{X} - \mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\eta})^T \mathbf{C}^{-1} (\mathbf{X} - \mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\eta}) \right\}, \end{aligned} \tag{18}$$

where the following notations are used in addition to those from (2) to (7):

- $\mathbf{f} = \{f(2), f(3), \ldots, f(T-1)\}^T$, a $T-2$ vector as the prior mean of $\boldsymbol{\eta}$;
- $\mathbf{D} = \sigma_G^2 \mathbf{I}_{K-2}$ is the covariance matrix of $\mathbf{Y}$ conditioning on $\boldsymbol{\eta}$, where $\mathbf{I}_m$ stands for the identity matrix of dimension m;
- $\mathbf{Z}$ the design matrix for $\mathbf{h}$ as in (5)
- $\mathbf{G}$ is a $(K-2) \times (T-2)$ matrix, with

$$G_{k,j} = \begin{cases} 1, & j = t_k - 1 \\ 0, & \text{Otherwise} \end{cases}$$

  Notice here the first time point is removed;
- $\mathbf{R} = \mathbf{R}(2 : (T-1), 2 : (T-1))$ is the $(T-2) \times (T-2)$ covariance matrix for the Brownian Bridge;
- $\mathbf{C} = \mathbf{C}(2 : T, 2 : T)$ is the $(T-1) \times (T-1)$ covariance matrix for the Brownian Motion error term $\boldsymbol{\xi}$;

In order to further simplify (18), the following notations are introduced:

- $\boldsymbol{\zeta} = \{\boldsymbol{\beta}^T, \boldsymbol{\eta}^T\}^T$, which is the joint vector of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ of length $Q+T-2$;
- $\mathbf{U}$ is a $T-1 \times (Q+T-2)$ matrix with $\mathbf{U}(1:(T-1), 1:Q) = \mathbf{Z}$, $\mathbf{U}(1:(T-2), (Q+1):(Q+T-2)) = I_{T-2}$ and $\mathbf{U}(T-1, (Q+1):(Q+T-2)) = 0$, which maps $\mathbf{U}\boldsymbol{\zeta} = \mathbf{h} + \boldsymbol{\beta}$;
- $\mathbf{V}$ is a $(T-2) \times (Q+T-2)$ matrix with $\mathbf{V}(1:(T-2), 1:Q) = \mathbf{0}$ and $\mathbf{V}(1:(T-2), (Q+1):(Q+T-2)) = \mathbf{I}_{T-2}$, such that $\mathbf{V}\boldsymbol{\zeta} = \boldsymbol{\eta}$;
- $\mathbf{W}$ is a $(K-2) \times (Q+T-2)$ matrix with $\mathbf{W}(1:(K-2), 1:Q) = 0$ and $\mathbf{W}(1:(K-2), (Q+1):(Q+T-2)) = \mathbf{G}$, such that $\mathbf{W}\boldsymbol{\zeta} = \mathbf{G}\boldsymbol{\eta}$.

Some algebra could simplify (18) into

$$\begin{aligned}\langle \boldsymbol{\zeta} | \mathbf{X}, \mathbf{Y} \rangle \propto & |\mathbf{R}|^{-1/2} |\mathbf{D}|^{-1/2} |\mathbf{C}|^{-1/2} \times \\ & \exp\left\{ -\frac{1}{2} \left[(\boldsymbol{\zeta} - \mathbf{M}_1^{-1}\mathbf{M}_2)^T \mathbf{M}_1 (\boldsymbol{\zeta} - \mathbf{M}_1^{-1}\mathbf{M}_2) + M_3 - \mathbf{M}_2^T \mathbf{M}_1^{-1} \mathbf{M}_2 \right] \right\}, \\ \propto & \exp\left\{ -\frac{1}{2} \left[(\boldsymbol{\zeta} - \mathbf{M}_1^{-1}\mathbf{M}_2)^T \mathbf{M}_1 (\boldsymbol{\zeta} - \mathbf{M}_1^{-1}\mathbf{M}_2) \right] \right\} \end{aligned} \tag{19}$$

where

$$\begin{aligned} \mathbf{M}_1 =& \mathbf{V}^T \mathbf{R}^{-1} \mathbf{V} + \mathbf{W}^T \mathbf{D}^{-1} \mathbf{V}_G + \mathbf{U}^T \mathbf{C}^{-1} \mathbf{U} \\ \mathbf{M}_2 =& \mathbf{V}^T \mathbf{R}^{-1} \mathbf{f} + \mathbf{W}^T \mathbf{D}^{-1} \mathbf{Y} + \mathbf{U}^T \mathbf{C}^{-1} \mathbf{X} \\ M_3 =& \mathbf{f}^T \mathbf{R}^{-1} \mathbf{f} + \mathbf{Y}^T \mathbf{D}^{-1} \mathbf{Y} + \mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}. \end{aligned}$$

It is shown here that the posterior of $\boldsymbol{\beta}, \boldsymbol{\eta}$ conditioning on $\boldsymbol{\phi}$ is a multivariate Gaussian density. The main computation complexity of evaluating this density comes from calculation of $\mathbf{M}_1^{-1}\mathbf{M}_2$.

B.2 Derivation of (11) and (12)

It is well known that the Brownian Bridge and Brownian Motion process are Markovian (Stirzaker and Grimmett, 2001), such that:

$$\begin{aligned} [\eta(t) | \eta(t-1), \eta(t-2)] =& [\eta(t) | \eta(t-1)] \\ [\xi(t) | \xi(t-1), \xi(t-2)] =& [\xi(t) | \xi(t-1)], \end{aligned}$$

where $\eta(t)$ is a Brownian Motion process as in our model (2) and $\xi(t)$ is a Brownian Motion process as in (5). This Markovian property directly suggests the conditional independence property of these two process, such that:

$$\begin{aligned} [\eta(t-1, t+1) | \eta(t)] =& [\eta(t-1) | \eta(t)][\eta(t+1) | \eta(t)] \\ [\xi(t-1, t+1) | \xi(t)] =& [\xi(t-1) | \xi(t)][\xi(t+1) | \xi(t)]. \end{aligned}$$

These properties help us to derive (11) and (12). As an illustration, we assume $T = 5$ (recall that $\eta(1)$ and $\eta(5)$ are fixed. So only $\boldsymbol{\eta}(2,3,4)$ are random), and one GPS observation is available at $t = 3$. The first part of (10), $\langle \boldsymbol{\eta}(1:T \setminus t_{1:K}) | \boldsymbol{\beta}, \boldsymbol{\eta}_G, \mathbf{X}, \mathbf{Y} \rangle$ under this situation is

$$\langle \eta(2), \eta(4) | \eta(3), \mathbf{X}(2:4), Y(3), \boldsymbol{\beta} \rangle = \langle \eta(2), \eta(4) | \eta(3), \mathbf{X}(2:4), \boldsymbol{\beta} \rangle, \tag{20}$$

as $Y(3)$ only depends $\eta(3)$ and independent of all the other random variables. For brevity, we will hide $\boldsymbol{\beta}$ term in all the following derivation and let $X_i = X(i), \eta_i = \eta(i), \xi_i = \xi(i)$. Using the conditional independence property of η and ξ together with certain variable transformation, we can simplify (20) into

$$\begin{aligned}
\langle \eta_2, \eta_4 | \eta_3, X_{2:4} \rangle &= \frac{\langle \eta_2, \eta_3, \eta_4, X_2, X_3, X_4 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} = \frac{\langle \eta_2, \eta_3, \eta_4, \xi_2, \xi_3, \xi_4 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} \\
&= \frac{\langle \eta_2, \eta_3, \eta_4 \rangle \langle \xi_2, \xi_3, \xi_4 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} = \frac{\langle \eta_2 | \eta_3 \rangle \langle \eta_4 | \eta_3 \rangle \langle \eta_3 \rangle \langle \xi_2 | \xi_3 \rangle \langle \xi_4 | \xi_3 \rangle \langle \xi_3 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} \\
&= \frac{\langle \eta_2 | \eta_3 \rangle \langle \xi_2 | \xi_3 \rangle \langle \eta_4 | \eta_3 \rangle \langle \xi_4 | \xi_3 \rangle \langle \eta_3 \rangle \langle \xi_3 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} \\
&= \frac{\langle \eta_2, X_2 | \eta_3, X_3 \rangle \langle \eta_4, X_4 | \eta_3, X_3 \rangle \langle \eta_3, X_3 \rangle}{\langle \eta_3, X_2, X_3, X_4 \rangle} \\
&= \frac{\langle \eta_2, X_2 | \eta_3, X_3 \rangle \langle \eta_4, X_4 | \eta_3, X_3 \rangle}{\langle X_2, X_4 | \eta_3, X_3 \rangle} = \ldots = \frac{\langle \eta_2, X_2 | \eta_3, X_3 \rangle \langle \eta_4, X_4 | \eta_3, X_3 \rangle}{\langle X_2 | \eta_3, X_3 \rangle \langle X_4 | \eta_3, X_3 \rangle} \\
&= \langle \eta_2 | \eta_3, X_2, X_3 \rangle \langle \eta_4 | \eta_3, X_4, X_3 \rangle.
\end{aligned}$$

The above is an illustration on how we prove (11) and (12) and we omit the lengthy proof of the general case. On the other hand, these expressions can be easily proved via the conditional independence property of a graphical model as in Lauritzen (1996).

B.3 Explicit expression of (12)

In the above subsection, we have shown that:

$$\begin{aligned}
&\langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, \mathbf{X}, \mathbf{Y} \rangle = \\
&\qquad \langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, \mathbf{X}(t_k : t_{k+1}) \rangle.
\end{aligned}$$

Here we derive the explicit expression of the right-hand side above. First, it is easy to use the conditional independence property to find:

$$\begin{aligned}
&\langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, \mathbf{X}(t_k : t_{k+1}) \rangle \\
&\qquad = \frac{\langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, X(t_k), X(t_{k+1}) \rangle}{\langle \mathbf{X}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, X(t_k), X(t_{k+1}) \rangle} \\
&\qquad = \frac{\langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}) \rangle}{\langle \mathbf{X}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, X(t_k), X(t_{k+1}) \rangle}.
\end{aligned}$$

Define two new variables,

$$\begin{aligned}
\boldsymbol{\eta}^c(t_k + 1 : t_{k+1} - 1) =& \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}) \\
\mathbf{X}^c(t_k + 1 : t_{k+1} - 1) =& \mathbf{X}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, X(t_k), X(t_{k+1}),
\end{aligned}$$

such that

$$\begin{aligned}
&\langle \boldsymbol{\eta}(t_k + 1 : t_{k+1} - 1) | \eta(t_k), \eta(t_{k+1}), \boldsymbol{\beta}, \mathbf{X}(t_k : t_{k+1}) \rangle \\
&\qquad = \langle \boldsymbol{\eta}^c(t_k + 1 : t_{k+1} - 1) | \mathbf{X}^c(t_k + 1 : t_{k+1} - 1) \rangle
\end{aligned} \tag{21}$$

Accoridng to the basic properties of the Brownian Bridge, it is easy to verify that

$$\boldsymbol{\eta}^c(t_k+1:t_{k+1}-1) \sim \text{MVN}(\mathbf{f}_k, \sigma_H^2\mathbf{R}_k), \tag{22}$$

where

$$\begin{aligned}
f_k(t) =& f(t) + a_k'(t)(\eta(t_k) - f(t_k)) + a_k(t)(\eta(t_{k+1}) - f(t_{k+1})) \\
=& a_k'(t)\eta(t_k) + a_k(t)\eta(t_{k+1}) \\
a_k(t) =& \frac{t-t_k}{t_{k+1}-t_k} \\
a_k'(t) =& 1 - a_k(t) = \frac{t_{k+1}-t}{t_{k+1}-t_k} \\
R_k(s,t) =& \frac{(s-t_k)(t_{k+1}-t)}{t_{k+1}-t_k}, \qquad\qquad t_k < s \le t < t_{k+1},
\end{aligned}$$

and $\text{MVN}(\cdot,\cdot)$ denotes a multivariate normal distribution.

The definition of $\mathbf{X}^c$ and $\mathbf{X}$ in (5) implies that $\mathbf{X}^c$ is the sum of a deterministic term and two independent Gaussian processes:

$$\mathbf{X}^c(t_k+1:t_{k+1}-1) = \mathbf{h}(t_k+1:t_{k+1}-1) + \boldsymbol{\eta}^c(t_k+1:t_{k+1}-1) + \boldsymbol{\xi}^c(t_k+1:t_{k+1}-1),$$

where $\mathbf{h}$ is defined in (5) and $\boldsymbol{\xi}^c$ is defined similarly as $\boldsymbol{\eta}^c$:

$$\boldsymbol{\xi}^c(t_k+1:t_{k+1}-1) = \boldsymbol{\xi}(t_k+1:t_{k+1}-1)|\xi(t_k), \xi(t_{k+1}) \sim \text{MVN}(\mathbf{g}_k, \sigma_D^2\mathbf{R}_k)$$

with $g_k(t) = a_k'(t)\xi(t_k) + a_k(t)\xi(t_{k+1})$.

In this way, the marginal distribution of $\mathbf{X}^c$ is

$$\mathbf{X}^c(t_k+1:t_{k+1}-1) \sim \text{MVN}(\mathbf{u}_k, (\sigma_H^2+\sigma_D^2)\mathbf{R}_k), \tag{23}$$

where

$$\begin{aligned}
u_k(t) &= h(t) + f_k(t) + g_k(t) \\
&= h(t) + a_k'(t)(\eta(t_k) + \xi(t_k)) + a_k(t)(\eta(t_{k+1}) + \xi(t_{k+1})) \\
&= h(t) + a_k'(t)(X(t_k) - h(t_k)) + a_k(t)(X(t_{k+1}) - h(t_{k+1})).
\end{aligned}$$

Also, the covariance between $\boldsymbol{\eta}^c$ and $\mathbf{X}^c$ is

$$\text{Cov}\,(\boldsymbol{\eta}^c, \mathbf{X}^c) = \text{Cov}\,(\boldsymbol{\eta}^c, \boldsymbol{\eta}^c) = \sigma_H^2\mathbf{R}_k. \tag{24}$$

With (22), (23), and (24), we can easily find that there is no need to invert any matrices when computing (21). As when the conditional covariance is calculated, the inverse of $\mathbf{R}_k$ will be canceled by $\mathbf{R}_k$, such that,

$$\begin{aligned}
\text{Cov}\,(\boldsymbol{\eta}^c, \mathbf{X}^c)\,(\text{Cov}\,(\mathbf{X}^c))^{-1} =& \sigma_H^2\mathbf{R}_k\left((\sigma_H^2+\sigma_D^2)\mathbf{R}_k)\right)^{-1} = \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2}\mathbf{I}_{t_{k+1}-t_k-1} \\
\text{Cov}\,(\boldsymbol{\eta}^c, \mathbf{X}^c)\,(\text{Cov}\,(\mathbf{X}^c))^{-1}\,\text{Cov}\,(\mathbf{X}^c, \boldsymbol{\eta}^c) =& \sigma_H^2(1 - \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2})\mathbf{R}_k = \frac{\sigma_H^2\sigma_D^2}{\sigma_H^2+\sigma_D^2}\mathbf{R}_k.
\end{aligned}$$

We only need to calculate $\rho = \frac{\sigma_H^2}{\sigma_H^2+\sigma_D^2}$ for the conditional mean and covariance. In this way, the desired posterior in (21) is

$$\begin{aligned}
&\langle \boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\eta(t_k),\eta(t_{k+1}),\boldsymbol{\beta},\mathbf{X}(t_k:t_{k+1})\rangle \\
&\qquad =\langle \boldsymbol{\eta}^c(t_k+1:t_{k+1}-1)|\mathbf{X}^c(t_k+1:t_{k+1}-1)\rangle \\
&\qquad \sim \text{MVN}\left(\mathbf{f}_k+\rho\left(\mathbf{X}(t_k+1:t_{k+1}-1)-\mathbf{u}_k\right),\rho\sigma_D^2\mathbf{R}_k\right) \qquad (25)
\end{aligned}$$

B.4 Explicit expression of $[\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\eta}_G|\mathbf{X}_G,\mathbf{Y}]$

Following notations in Appendix B.1

$$\begin{aligned}
&[\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\eta}_G|\mathbf{X}_G,\mathbf{Y}] \propto [\boldsymbol{\phi}][\boldsymbol{\eta}(t_{1:K})|\boldsymbol{\phi}][\mathbf{Y}|\boldsymbol{\eta}(t_{1:K})][\mathbf{X}(t_{1:K})|\boldsymbol{\beta},\boldsymbol{\eta}(t_{1:K}),\boldsymbol{\phi}] \\
&\quad \propto [\boldsymbol{\phi}]|\mathbf{R}_G|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\eta}_G-\mathbf{f}_G)^T\mathbf{R}_G^{-1}(\boldsymbol{\eta}_G-\mathbf{f}_G)\right\}\times|\mathbf{D}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\eta}_G)^T\mathbf{D}^{-1}(\mathbf{Y}-\boldsymbol{\eta}_G)\right\} \\
&\qquad \times|\mathbf{C}_G|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{X}_G-\mathbf{Z}_G\boldsymbol{\beta}-\boldsymbol{\eta}_G)^T\mathbf{C}_G^{-1}(\mathbf{X}_G-\mathbf{Z}_G\boldsymbol{\beta}-\boldsymbol{\eta}_G)\right\}, \qquad (26)
\end{aligned}$$

where $\mathbf{X}_G=\mathbf{X}(t_{1:K})$ and $\mathbf{R}_G=\mathbf{R}(\mathbf{t_{1:K}},\mathbf{t_{1:K}})$. Similar notations apply to $\boldsymbol{\eta}_G$, $\mathbf{f}_G,\mathbf{C}_G,\mathbf{Z}_G$. Also, we will introduce the following notations similar to those in Appendix B.1, which is the sub-vector or sub-matrix of those in Appendix B.1 with respect to the GPS observations:

- $\boldsymbol{\zeta}_G=\{\boldsymbol{\beta}^T,\boldsymbol{\eta}_G^T\}^T$, which is the joint vector of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}_G$ of length $Q+K-2$;
- $\mathbf{U}_G$ is a $K-1\times(Q+K-2)$ matrix with $\mathbf{U}_G(1:(K-1),1:Q)=\mathbf{Z}_G$, $\mathbf{U}_G(1:(K-2),(Q+1):(Q+T-2))=I_{K-2}$ and $\mathbf{U}_G(K-1,(Q+1):(Q+T-2))=0$, which maps $\mathbf{U}_G\boldsymbol{\zeta}_G=\mathbf{h}_G+\boldsymbol{\beta}_G$;
- $\mathbf{V}_G$ is a $(K-2)\times(Q+K-2)$ matrix with $\mathbf{V}_G(1:(K-2),1:Q)=\mathbf{0}$ and $\mathbf{V}_G(1:(K-2),(Q+1):(Q+K-2))=\mathbf{I}_{K-2}$, such that $\mathbf{V}_G\boldsymbol{\zeta}_G=\boldsymbol{\eta}_G$;

Some algebra could simplify (26) into

$$\begin{aligned}
&[\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\eta}_G|\mathbf{X}_G,\mathbf{Y}] \propto \\
&\quad \frac{1}{\sigma_H^2}\frac{1}{\sigma_D^2}\times|\mathbf{R}_G|^{-1/2}|\mathbf{D}|^{-1/2}|\mathbf{C}_G|^{-1/2}\times \\
&\quad \exp\left\{-\frac{1}{2}\left[(\boldsymbol{\zeta}_G-\mathbf{M}_{G1}^{-1}\mathbf{M}_{G2})^T\mathbf{M}_{G1}(\boldsymbol{\zeta}_G-\mathbf{M}_{G1}^{-1}\mathbf{M}_{G2})+M_{G3}-\mathbf{M}_{G2}^T\mathbf{M}_{G1}^{-1}\mathbf{M}_{G2}\right]\right\}, \qquad (27)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{M}_{G1} =& \mathbf{V}_G^T\mathbf{R}_G^{-1}\mathbf{V}_G+\mathbf{V}_G^T\mathbf{D}^{-1}\mathbf{V}_G+\mathbf{U}_G^T\mathbf{C}^{-1}\mathbf{U}_G \\
\mathbf{M}_{G2} =& \mathbf{V}_G^T\mathbf{R}_G^{-1}\mathbf{f}_G+\mathbf{V}_G^T\mathbf{D}^{-1}\mathbf{Y}+\mathbf{U}_G^T\mathbf{C}_G^{-1}\mathbf{X}_G \\
M_{G3} =& \mathbf{f}_G^T\mathbf{R}_G^{-1}\mathbf{f}_G+\mathbf{Y}^T\mathbf{D}^{-1}\mathbf{Y}+\mathbf{X}_G^T\mathbf{C}_G^{-1}\mathbf{X}_G.
\end{aligned}$$

Following (27), it is easy to learn that

$$[\boldsymbol{\zeta}_G|\mathbf{X}_G,\mathbf{Y},\boldsymbol{\phi}]\sim MVN\left(\mathbf{M}_{G1}^{-1}\mathbf{M}_{G2},\mathbf{M}_{G1}^{-1}\right). \qquad (28)$$

Integrate $\boldsymbol{\zeta}_G$ out, we have

$$[\boldsymbol{\phi}|\mathbf{X}_G, \mathbf{Y}] \propto \frac{1}{\sigma_H^2}\frac{1}{\sigma_D^2}|\mathbf{R}_G|^{-1/2}|\mathbf{D}|^{-1/2}|\mathbf{C}_G|^{-1/2}|\mathbf{M}_{G1}|^{-1/2}\exp\left\{-\frac{1}{2}\left[M_{G3} - \mathbf{M}_{G2}^T\mathbf{M}_{G1}^{-1}\mathbf{M}_{G2}\right]\right\} \tag{29}$$

B.5 Marginal distribution of $\boldsymbol{\eta}$ at the non-GPS points

With (28), we can marginalize $\boldsymbol{\eta}_G, \boldsymbol{\beta}$ out in (25) to obtain $\langle\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\mathbf{X},\mathbf{Y}\rangle$, which is later used in the numerical integration. Let $\boldsymbol{\zeta}_k = \{\boldsymbol{\beta}^T, \eta(t_k), \eta(t_{k+1})\}^T$ and its mean and covariance matrix obtained in (28) be denoted as $\tilde{\boldsymbol{\zeta}}_k, \tilde{\boldsymbol{\Sigma}}_k$ respectively.

The $\mathbf{u}_k$ of (23) can be written as as a linear transformation of $\boldsymbol{\zeta}_k$, such that $\mathbf{u}_k = \mathbf{B}_k\boldsymbol{\zeta}_k$, where $\mathbf{B}_k = \left[\rho\left(\mathbf{Z}_k - \mathbf{A}_k\mathbf{Z}_k^G\right), \mathbf{A}_k\right]$. $\mathbf{Z}_k$ is the rows of the design matrix, $\mathbf{Z}(t_k+1:t_{k+1}-1, 1:Q)$, which corresponds to this period of the non-GPS observation. $\mathbf{Z}_k^G = \mathbf{Z}(\mathbf{t}_{k,k+1}, 1:Q)$ corresponds to the rows of the design matrix of the two GPS observations. $\mathbf{A}_k$ is the matrix of the linear weights of $a_k'(t), a_k(t)$ as in Equation (22). Marginalize $\boldsymbol{\zeta}_k$ out in (25) results in

$$\langle\boldsymbol{\eta}(t_k+1:t_{k+1}-1)|\mathbf{X},\mathbf{Y}\rangle \sim \mathrm{MVN}\left(\mathbf{B}_k\tilde{\boldsymbol{\zeta}}_k, \mathbf{B}_k\tilde{\boldsymbol{\Sigma}}_k\mathbf{B}_k^T\right) \tag{30}$$

B.6 Integration part of (15) via a Gaussian mixture

As introduced in Subsection 3.2, we evaluate the integration part of (15) via the numerical integration in (17). The distribution of $\boldsymbol{\zeta}$ is multivariate normal conditioning on $\boldsymbol{\phi}$ as in (28) and (30) and therefore the posterior density of $\boldsymbol{\zeta}$ can be approximated by a mixture of multivariate normal densities. Let $\tilde{\boldsymbol{\zeta}}^{(i)}$ and $\tilde{\boldsymbol{\Sigma}}^{(i)}$ be the posterior mean and covariance of $\boldsymbol{\zeta}$ conditioning on the i-th grid point of $\boldsymbol{\phi}(\mathbf{z}(j_1, j_2))$ (Here j_1, j_2 are collapsed into a single index set and $L = (J_1^+ + J_1^- + 1) \times (J_2^+ + J_2^- + 1)$). We simplify (17) into

$$[\boldsymbol{\zeta}|\mathbf{X},\mathbf{Y}] \approx \sum_{i=1}^{L} w_i[\boldsymbol{\zeta}|\boldsymbol{\phi}^{(i)}, \mathbf{X}, \mathbf{Y}] = \sum_{i=1}^{L} w_i\Psi(\cdot; \tilde{\boldsymbol{\zeta}}^{(i)}, \tilde{\boldsymbol{\Sigma}}^{(i)}),$$

where $\boldsymbol{\Psi}$ stands for the probability density function of the multivariate normal distribution. For our application, we are only interested in finding out the posterior mean and variance of $\boldsymbol{\zeta}$. As in (Frühwirth-Schnatter, 2006), the posterior mean equals

$$\tilde{\boldsymbol{\zeta}} = \sum_{i=1}^{L} w_i\tilde{\boldsymbol{\zeta}}^{(i)},$$

and the posterior variance of the k-th element of $\boldsymbol{\zeta}$ is

$$\tilde{\sigma}_k^2 = \sum_{i=1}^{L} w_i\left[\tilde{\Sigma}^{(i)}(k,k) + \sum_{i=1}^{L}\left(\tilde{\zeta}^{(i)}(k) - \tilde{\zeta}(k)\right)^2\right].$$

References

Benoit-Bird, K., Battaile, B., Nordstrom, C., and Trites, A. (2013a). Foraging behavior of northern fur seals closely matches the hierarchical patch scales of prey. *Marine Ecology Progress Series*, 479:283–302.

Benoit-Bird, K. J., Battaile, B. C., Heppell, S. a., Hoover, B., and et al (2013b). Prey patch patterns predict habitat use by top marine predators with diverse foraging strategies. *PloS one*, 8(1):e53348.

Boehme, L., Kovacs, K., and Lydersen, C. (2010). Biologging in the global ocean observing system. *Proceedings of OceanObs 09: Sustained Ocean Observations and Information for Society*, 2(September):21–25.

Boehme, L., Meredith, M. P., Thorpe, S. E., Biuw, M., and Fedak, M. (2008). Antarctic circumpolar current frontal system in the South Atlantic: Monitoring using merged Argo and animal-borne sensor data. *Journal of Geophysical Research C: Oceans*, 113.

Bryant, E. (2007). 2D Location Accuracy Statistics for Fastloc . Technical Report September, Wildtrack Telemetry System Ltd.

Casella, G. (1985). An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87.

Dowd, M. and Joy, R. (2011). Estimating behavioral parameters in animal movement models using a state-augmented particle filter. *Ecology*, 92(3):568–575.

Elkaim, G. and Decker, E. (2006). Marine Mammal Marker (MAMMARK) dead reckoning sensor for In-Situ environmental monitoring. *Proc. of the ION/IEEE PLANS*.

Evans, W. (1970). Uses of advanced space technology and upgrading the future of oceanography. *AIAA Paper*, 3:7–01273.

Foley, K. M. and Fuentes, M. (2008). A statistical framework to combine multivariate spatial data and physical models for Hurricane surface wind prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(1):37–59.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*. Springer.

Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics*, (March):36–45.

Grewal, M. S., Weill, L. R., and Andrews, A. P. (2007). *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons.

Hazel, J. (2009). Evaluation of fast-acquisition GPS in stationary tests and fine-scale tracking of green turtles. *Journal of Experimental Marine Biology and Ecology*, 374(1):58–68.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (1993). Global positioning system. theory and practice. *Global Positioning System. Theory and practice., by Hofmann-Wellenhof, B.; Lichtenegger, H.; Collins, J.. Springer, Wien (Austria), 1993, 347 p., ISBN 3-211-82477-4, Price DM 79.00. ISBN 0-387-82477-4 (USA).*, 1.

Hooten, M. B., Hanks, E. M., Johnson, D. S., and Alldredge, M. W. (2013). Reconciling resource utilization and resource selection functions. *The Journal of animal ecology*, 82(6):1146–54.

Horne, J. S., Garton, E., Krone, S., and Lewis, J. (2007). Analyzing animal movements using Brownian bridges. *Ecology*, 88(9):2354–2363.

Humphries, N. E., Queiroz, N., Dyer, J. R. M., Pade, N. G., and et al (2010). Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature*, 465(7301):1066–9.

Isachsen, P. l. E., Sø rlie, S. R., Mauritzen, C., Lydersen, C., and et al (2014). Upper-ocean hydrography of the Nordic Seas during the International Polar Year (20072008) as observed by instrumented seals and Argo floats. *Deep Sea Research Part I: Oceanographic Research Papers*, 93:41–59.

Johnson, D., London, J., Lea, M.-a., and Durban, J. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.

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

Kranstauber, B., Kays, R., Lapoint, S. D., Wikelski, M., and Safi, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *The Journal of animal ecology*, 81(4):738–46.

Kranstauber, B., Safi, K., and Bartumeus, F. (2014). Bivariate Gaussian bridges: directional factorization of diffusion in Brownian bridge models. *Movement Ecology*, 2(1):5.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford, UK.

Liu, Y., Battaile, B. C., Zidek, J. V., and Trites, A. (2014). Bayesian melding of the dead-reckoned path and gps measurements for an accurate and high-resolution path of marine mammals. *arXiv preprint arXiv: 1411.6683*.

Liu, Z., Le, N. D., and Zidek, J. V. (2011). An empirical assessment of Bayesian melding for mapping ozone pollution. *Environmetrics*, 22(3):340–353.

Lydersen, C. (2002). Salinity and temperature structure of a freezing Arctic fjordmonitored by white whales (Delphinapterus leucas). *Geophysical Research Letters*, 29(23):2119.

Meredith, M. P., Nicholls, K. W., Renfrew, I. a., Boehme, L., Biuw, M., and Fedak, M. (2011). Seasonal evolution of the upper-ocean adjacent to the South Orkney Islands, Southern Ocean: Results from a lazy biological mooring. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(13-16):1569–1579.

Nordstrom, C. a., Benoit-Bird, K. J., Battaile, B. C., and Trites, A. W. (2013). Northern fur seals augment ship-derived ocean temperatures with higher temporal and spatial resolution data in the eastern Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 94:257–273.

Padman, L., Costa, D. P., Bolmer, S. T., Goebel, M. E., Huckstadt, L. a., Jenkins, A., McDonald, B. I., and Shoosmith, D. R. (2010). Seals map bathymetry of the Antarctic continental shelf. *Geophysical Research Letters*, 37(21):n/a–n/a.

Rue, H. v., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Rutz, C. and Hays, G. C. (2009). New frontiers in biologging science.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical science*, 4(4):409–423.

Sahu, S. (2010). Fusing point and areal level spacetime data with application to wet deposition. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, pages 77–103.

Sawyer, H., Kauffman, M. J., Nielson, R. M., and Horne, J. S. (2009). Identifying and prioritizing ungulate migration routes for landscape-level conservation. *Ecological applications : a publication of the Ecological Society of America*, 19(8):2016–25.

Stirzaker, G. G. D. and Grimmett, D. (2001). *Probability and random processes*. Oxford Science Publications.

Wahba, G. (1965). A least squares estimate of satellite attitude. *SIAM review*, 7(3):409–409.

Wilson, R., Shepard, E., and Liebsch, N. (2008). Prying into the intimate details of animal lives: use of a daily diary on animals. *Endangered Species Research*, 4(January):123–137.

Wilson, R. P., Liebsch, N., Davies, I. M., Quintana, F., Weimerskirch, H., Storch, S., and et al (2007). All at sea with animal tracks; methodological and analytical solutions for the resolution of movement. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(3-4):193–210.