Supplementary Material

Supplementary material and figures that support the ideas presented in the main text can be found in this document. Section 1 presents the theory behind the likelihood and travel time derivatives calculations, while Section 2 gives the theory behind the Bayesian filter computations. In Section 3, the supplementary figures are presented.

1 LIKELIHOOD AND TRAVEL TIME DERIVATIVE CALCULATIONS

The log-likelihood function at one vocalization time is defined by

$$l(x,y) = -\frac{\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}(x,y)\right)^2}{2\tau^2}, \tag{S1}$$

where the modeled travel time $T_{i,j} = T_{i,j}(x,y)$ is defined in equation (1). The derivatives of this log-likelihood with respect to lateral position (x,y) and the time shift are

$$\begin{aligned}
\frac{dl}{dx} &= \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{dT_{i,j}}{dx} = \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{(x-x_{i,j})}{c^2(T_{i,j}-\eta)},\\
\frac{dl}{dy} &= \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{dT_{i,j}}{dy} = \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{(y-y_{i,j})}{c^2(T_{i,j}-\eta)},\\
\frac{dl}{d\eta} &= \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{dT_{i,j}}{d\eta} = \frac{-1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right).
\end{aligned}$$

Considering a fixed (x,y), the latter equation defines a linear solution for η when setting $\frac{dl}{d\eta} = 0$. For the position (x,y), the second derivatives can be computed similarly

$$\begin{aligned}
\frac{d^2l}{dx^2} &= \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\left[\frac{dT_{i,j}}{dx}\right]^2 - \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{d^2T_{i,j}}{dx^2},\\
\frac{d^2l}{dxdy} &= \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\left[\frac{dT_{i,j}}{dx}\right]\left[\frac{dT_{i,j}}{dy}\right] - \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{d^2T_{i,j}}{dxdy},\\
\frac{d^2l}{dy^2} &= \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\left[\frac{dT_{i,j}}{dy}\right]^2 - \frac{1}{\tau^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}\right)\frac{d^2T_{i,j}}{dy^2}.
\end{aligned}$$

The travel time derivatives are

$$
\begin{aligned}
\frac{dT_{i,j}}{dx} &= \frac{(x-x_{i,j})}{c\sqrt{(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2}},\\
\frac{dT_{i,j}}{dy} &= \frac{(y-y_{i,j})}{c\sqrt{(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2}},\\
\frac{d^2T_{i,j}}{dx^2} &= \frac{1}{c\sqrt{(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2}}\\
&- \frac{(x-x_{i,j})^2}{c\left[(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2\right]^{3/2}},\\
\frac{d^2T_{i,j}}{dxdy} &= -\frac{(x-x_{i,j})(y-y_{i,j})}{c\left[(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2\right]^{3/2}},\\
\frac{d^2T_{i,j}}{dy^2} &= \frac{1}{c\sqrt{(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2}}\\
&- \frac{(y-y_{i,j})^2}{c\left[(x-x_{i,j})^2+(y-y_{i,j})^2+(z-z_{i,j})^2\right]^{3/2}}.
\end{aligned}
$$

Given a known or estimated position $(\hat{x}, \hat{y})$, the measurement noise variance τ^2 is estimated by

$$\hat{\tau}^2 = \frac{1}{n_1+n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(T_{i,j}^{(obs)} - T_{i,j}(\hat{x},\hat{y})\right)^2. \tag{S2}$$

2 BAYESIAN FILTERING CALCULATIONS

For completeness, we give a brief account of the recursive methods for computing the filtering and prediction steps for state space models. A recent description of these approaches is provided by Särkkä (2013).

We denote the travel time data at time s_k by $\boldsymbol{d}_k = \{T_{i,j,k}^{(obs)}; i = 1,\ldots,n_j, j = 1,2\}$. All the travel time data available up to this time are given by $\boldsymbol{D}_k = \{\boldsymbol{d}_k,\ldots,\boldsymbol{d}_1\}$.

We have the filtering state probability density function (PDF) at time s_{k-1} denoted $p(\boldsymbol{m}_{k-1}|\boldsymbol{D}_{k-1})$. Assume this is Gaussian with mean $\hat{\boldsymbol{\mu}}_{k-1}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_{k-1}$. The prediction PDF at time s_k is achieved by marginalizing over the previous state variable:

$$p(\boldsymbol{m}_k|\boldsymbol{D}_{k-1}) = \int p(\boldsymbol{m}_k,\boldsymbol{m}_{k-1}|\boldsymbol{D}_{k-1})d\boldsymbol{m}_{k-1}. \tag{S3}$$

Assuming a Gaussian approximation for the filtering distribution at time s_{k-1} and relying on the linear dynamical model in equation (6), we can solve equation (S3) to see that the predictive PDF is also Gaussian with mean $\bar{\boldsymbol{\mu}}_k = \boldsymbol{A}_{k-1,k}\hat{\boldsymbol{\mu}}_{k-1}$ and covariance matrix $\bar{\boldsymbol{\Sigma}}_k = \boldsymbol{A}_{k-1,k}\hat{\boldsymbol{\Sigma}}_{k-1}\boldsymbol{A}_{k-1,k}^t + \boldsymbol{S}_{k-1,k}$.

We denote this predicted state PDF at time s_k by $p(\boldsymbol{m}_k|\boldsymbol{D}_{k-1})$ (at the first time step, this is only the initial model $p(\boldsymbol{m}_1)$). The model is updated with data $\boldsymbol{d}_k$ to form the filtering PDF:

$$\begin{aligned} p(\boldsymbol{m}_k|\boldsymbol{D}_k) &= \frac{p(\boldsymbol{m}_k|\boldsymbol{D}_{k-1})p(\boldsymbol{d}_k|\boldsymbol{m}_k)}{p(\boldsymbol{d}_k|\boldsymbol{D}_{k-1})} \\ &\propto p(\boldsymbol{m}_k|\boldsymbol{D}_{k-1})p(\boldsymbol{d}_k|\boldsymbol{m}_k), \end{aligned} \tag{S4}$$

where we assume conditionally-independent data; $p(\boldsymbol{d}_k|\boldsymbol{m}_k, \boldsymbol{D}_{k-1}) = p(\boldsymbol{d}_k|\boldsymbol{m}_k)$. We approximate the PDF in equation (S4) by a Gaussian model with mean at the mode of equation (S4) and covariance defined from the curvature at this mode. The optimization is done similarly to the single-time approach described in equation (4). We now start with the position prediction $\boldsymbol{m}_k^0 = \bar{\boldsymbol{\mu}}_k$ and the iteration proceeds by $r = 1, \ldots$:

$$\boldsymbol{m}_k^r = \boldsymbol{m}_k^{r-1} - \left[\frac{d^2 l_k(\boldsymbol{m}_k^{r-1})}{d\boldsymbol{m}_k^2}\right]^{-1} \frac{dl_k(\boldsymbol{m}_k^{r-1})}{d\boldsymbol{m}_k}, \tag{S5}$$

where the log filtering PDF in equation (S4) at this time s_k is defined by

$$l_k(\boldsymbol{m}_k) = -\frac{1}{2}(\boldsymbol{m}_k - \bar{\boldsymbol{\mu}}_k)^t \bar{\boldsymbol{\Sigma}}_k^{-1}(\boldsymbol{m}_k - \bar{\boldsymbol{\mu}}_k) - \frac{1}{2\tau^2}[\boldsymbol{d}_k - \boldsymbol{T}_k(\boldsymbol{m}_k)]^t[\boldsymbol{d}_k - \boldsymbol{T}_k(\boldsymbol{m}_k)], \tag{S6}$$

using $\boldsymbol{T}_k$ to denote the vector of modeled travel time data at all fiber pick locations. The required derivatives are as calculated in Supplemental Material Section 1. The fitted mean is the mode $\hat{\boldsymbol{m}}_k$ obtained by the iterative scheme in equation (S5) and the fitted covariance of the Gaussian approximation is

$$\hat{\boldsymbol{\Sigma}}_k^{-1} = -\frac{d^2 l_k(\hat{\boldsymbol{m}}_k)}{d\boldsymbol{m}_k^2}. \tag{S7}$$

The prediction and filtering equations in equations (S3) and (S4) are common in tracking. They represent online calculations. Smoothing or posterior expressions can be used similarly for offline inspection. An effective backward calculation from an end time point s_K, moving in a stepwise manner to the first time s_1, defines the expression for the posterior PDF:

$$p(\boldsymbol{m}_k|\boldsymbol{D}_K) = \int \frac{p(\boldsymbol{m}_{k+1}|\boldsymbol{m}_k)p(\boldsymbol{m}_k|\boldsymbol{D}_k)}{p(\boldsymbol{m}_{k+1}|\boldsymbol{D}_k)} p(\boldsymbol{m}_{k+1}|\boldsymbol{D}_K) d\boldsymbol{m}_{k+1}. \tag{S8}$$

This relation builds on the Markovian state space modeling assumptions in equation (6) and the conditionally-independent assumptions of the data. Relying again on Gaussian approximations for the PDFs, all expressions in equation (S8) are available from the Gaussian approximations to the filtering and predictive PDF, as well as the smoothing PDF at the latter time t_{k+1} in the backward recursion. We can hence complete the integral to get an associated Gaussian approximation to the smoothing PDF at time t_k. Denoting the mean and covariance matrix of this smoothing PDF by $\tilde{\boldsymbol{\mu}}_k$ and $\tilde{\boldsymbol{\Sigma}}_k$, we have

$$\begin{aligned} \boldsymbol{J}_k &= \hat{\boldsymbol{\Sigma}}_k \boldsymbol{A}_{k,k+1}^t \bar{\boldsymbol{\Sigma}}_{k+1}^{-1}, \\ \tilde{\boldsymbol{\mu}}_k &= \hat{\boldsymbol{\mu}}_k + \boldsymbol{J}_k(\tilde{\boldsymbol{\mu}}_{k+1} - \bar{\boldsymbol{\mu}}_{k+1}) \\ \tilde{\boldsymbol{\Sigma}}_k &= \hat{\boldsymbol{\Sigma}}_k + \boldsymbol{J}_k(\tilde{\boldsymbol{\Sigma}}_{k+1} - \bar{\boldsymbol{\Sigma}}_{k+1})\boldsymbol{J}_k^t. \end{aligned} \tag{S9}$$

For details, see (Särkkä, 2013, Chapter 8).

3 SUPPLEMENTARY FIGURES

Two supplementary figures are given. Figure S1(A, B) shows an example of recorded air gun signals received on the inner (A) and outer cable (B). Furthermore, it also shows the comparison between the velocity of the ship derived from the GPS log (from a GNSS receiver) and the velocities in North-South and East-West directions from the estimators.

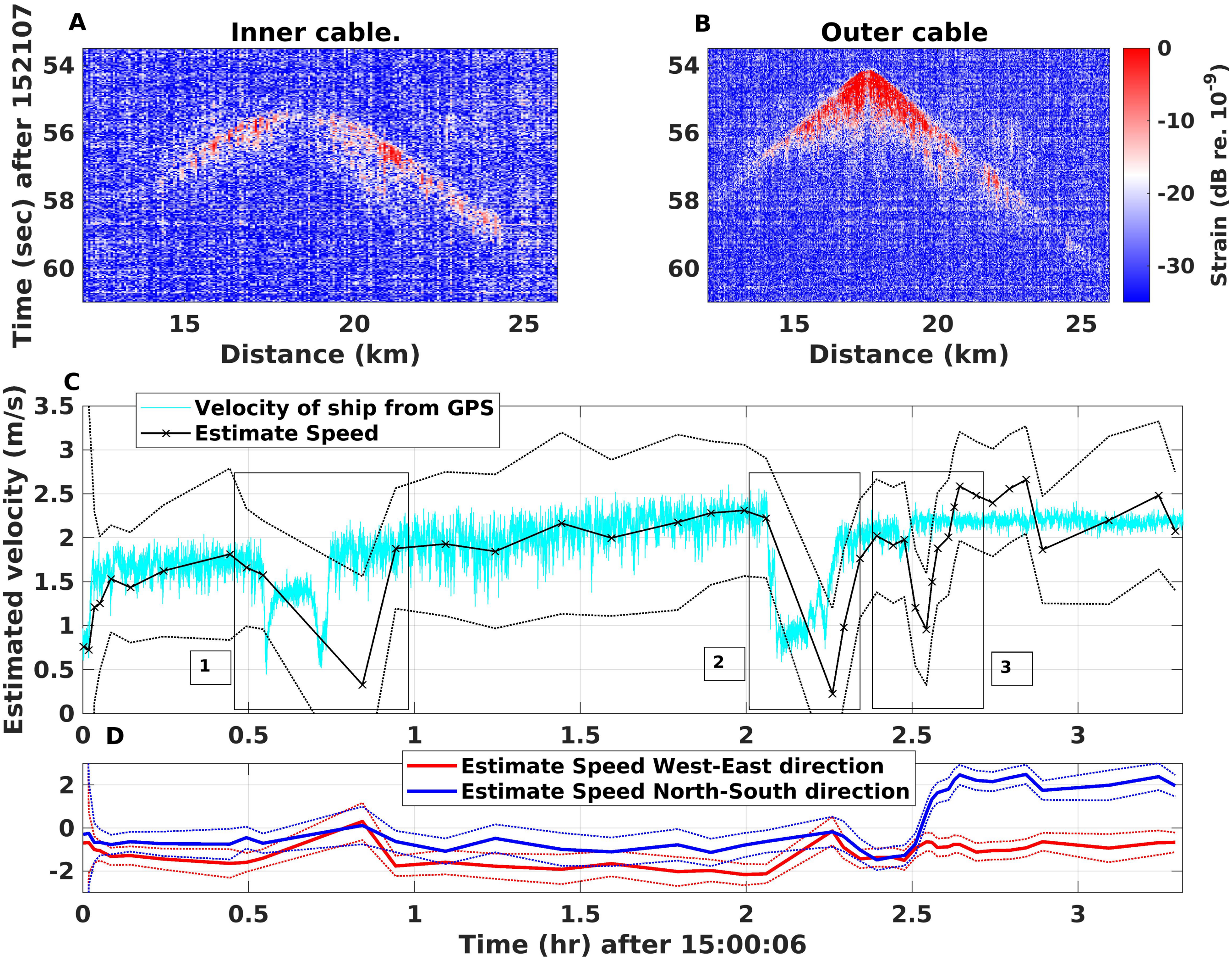

Figure S1: Supplemental information from the air gun processing. (A) An air gun shot received on the inner cable. (B) The same airgun shot on the outer cable. In this example, the air gun was fired directly above the outer cable. (C) A comparison between the velocity computed using GPS position recorded by a GNSS receiver and the Bayesian filter. (D) The decomposed velocities in West-East and North-South directions. The dashed lines in (C) and (D) are uncertainties related to the velocities. The numbers in (C) indicate where the ships made sharp turns, (1) and (3), and a loop, (2).

Figure S2 shows RMS levels for whale tracks (A), (G), and (E) computed in the same way as Figure 6. The first column of Figure S2 shows the RMS levels as a function of time. The second column of Figure S2 shows a zoomed-in version highlighting the inter-call and inter-series intervals for the various whale tracks.

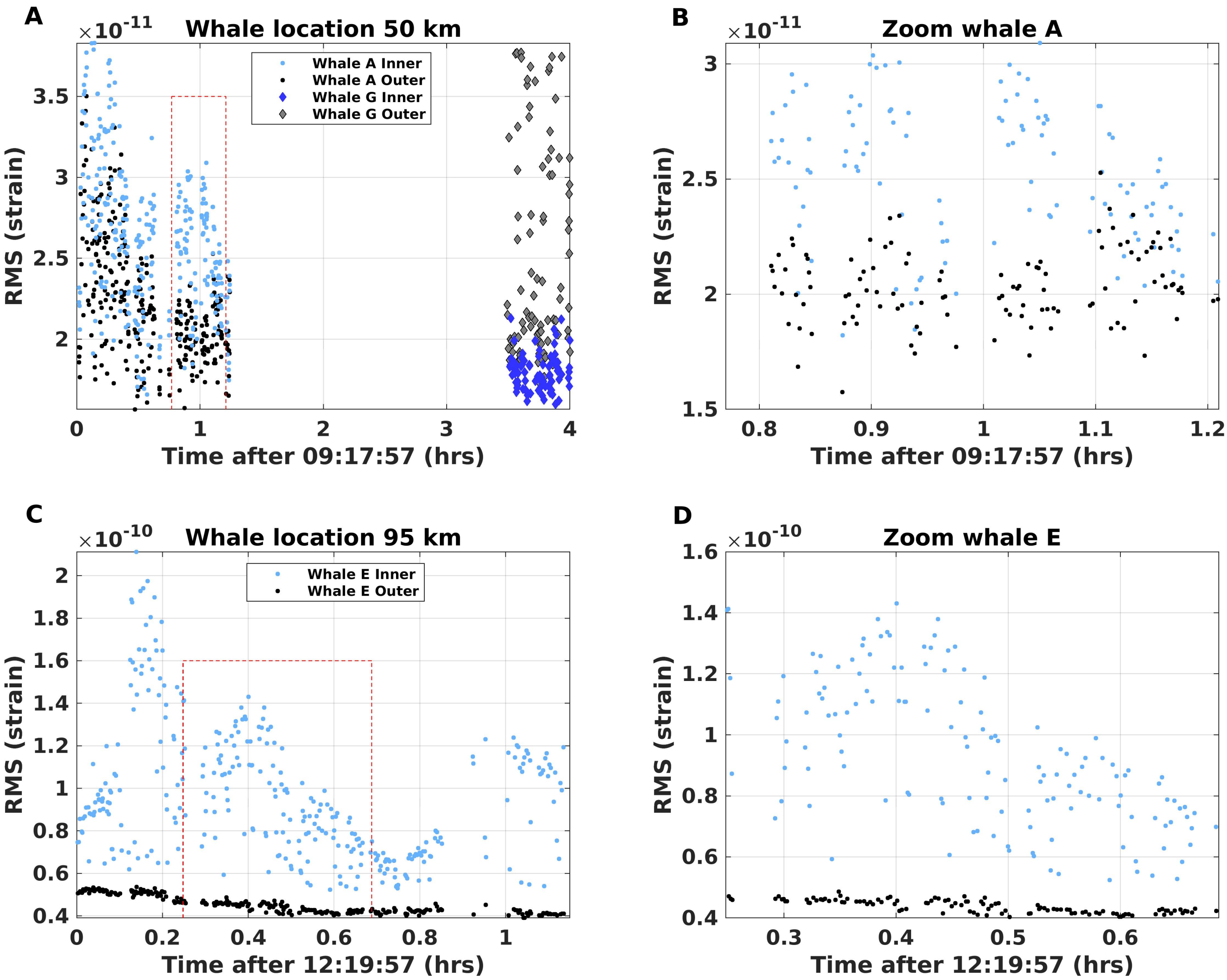

Figure S2: Observed RMS amplitudes of all Fin whale vocalizations computed by a rectangular window around the calls. (A, C) The vocalization for the whales in the two whale locations (45 and 95 km) as given in Figure 5. (B, D) Zoomed-in versions are meant to show the periods with whale calls and inter-call breaks.

REFERENCES

Särkkä, S. (2013). *Bayesian filtering and smoothing* (Cambridge university press)