

SUPPLEMENTARY FILE TO "VARYING COEFFICIENTS CORRELATED VELOCITY MODELS IN COMPLEX LANDSCAPES WITH BOUNDARIES APPLIED TO NARWHAL RESPONSES TO NOISE EXPOSURE"

BY ALEXANDRE DELPORTE^{1,a}, SUSANNE DITLEVSEN^{2,c} AND ADELINE SAMSON^{1,b}

¹Laboratoire Jean Kuntzmann, Université Grenoble-Alpes, ^aalexandre.delporte@univ-grenoble-alpes.fr;

^badeline.leclercq-samson@univ-grenoble-alpes.fr

²Department of Mathematical Sciences, University of Copenhagen, ^csusanne@math.ku.dk

APPENDIX A: MORE DETAILS ABOUT THE DATA

This section is a complement to Section 2 of the paper. We provide more information about the narwhal movement data used for the analysis of behavioral disturbance.

The time step between consecutive GPS observations is not constant. Its median is 4.8 minutes and its mean is 9.3 minutes. We show the histogram of the time steps in Figure 1.

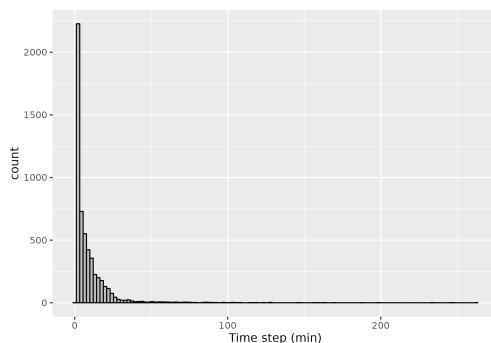


Fig 1: Histogram of time steps

The observations are divided into unexposed periods, for which the narwhals are not in line of sight with the ship; trial periods, when the narwhals are exposed to the ship and airguns are shot; and intertrial periods, when the narwhals are exposed to the ship but airguns are not shot. These periods are indicated by a categorical variable T_{ship} in the dataset. Figure 2 shows how the exposure periods are distributed among the 6 narwhals that were tracked. Our analysis in section 6 does not distinguish between intertrial and trial periods. They are both treated as exposure periods, though the nature and intensity of the behavioral response might differ for the two periods. We adopted this approach due to the lack of intertrial data as well as a potential persistence in time of the behavior shift due to airgun exposure during trial periods.

Table 1 shows how the data is distributed among the different narwhals. Figure 3 shows all the tracks before and after exposure with a base map of Scoresby Sound fjords system.

All the relevant covariates used for the analysis of narwhals movement are summarized in Table 2.

Keywords and phrases: behavioral response study, stochastic differential equations, mixed effect model.

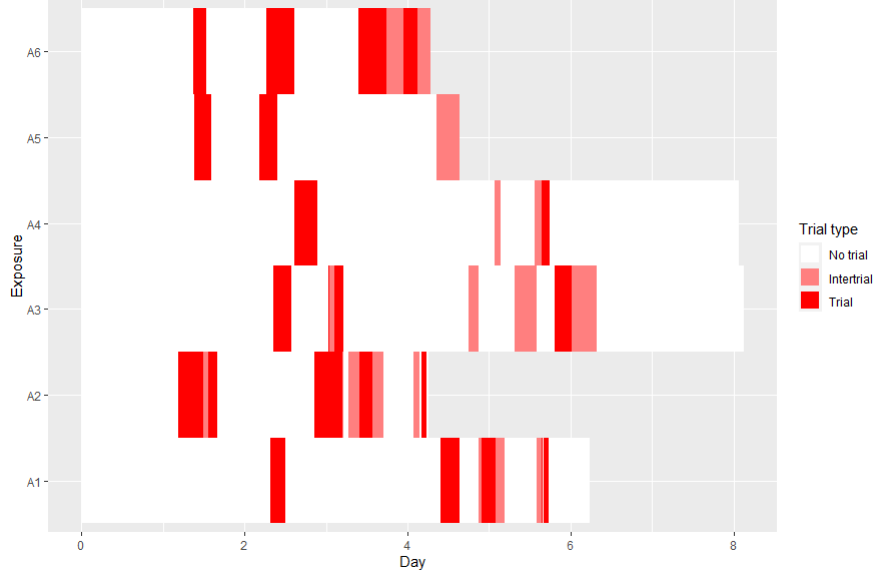


Fig 2: Trial and Intertrial periods for each narwhal

Narwhal ID	Number of measurement before exposure	Number of measurements during exposure
A1	354	576
A2	151	515
A3	397	680
A4	127	642
A5	207	419
A6	322	425
Total	1558	3257

TABLE 1

Distribution of the data among the 6 individuals

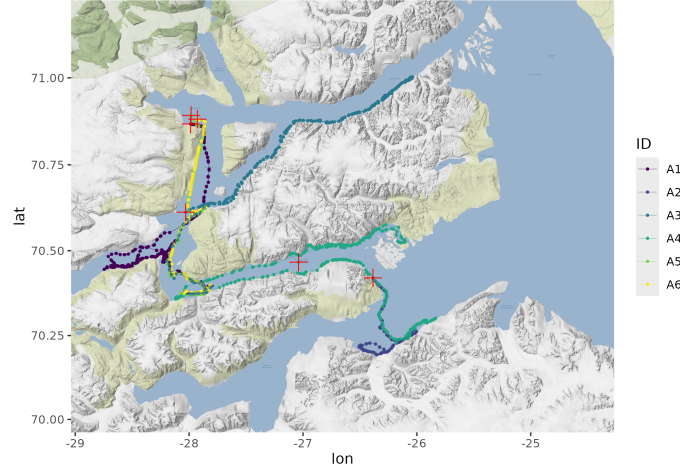
Covariate	Unit	Description	Domain
$D^{ship}(t)$	km	distance in kilometers between the narwhal and the ship at time t	\mathbb{R}^+
$E^{ship}(t) = \frac{1}{D^{ship}(t)}$	km^{-1}	global exposure level of the narwhal to the ship disturbance at time t	\mathbb{R}^+
$D^{shore}(t)$	km	distance between the narwhal and the nearest point on the shore at time t	\mathbb{R}^+
$I^{shore}(t) = \frac{1}{D^{shore}(t)}$	km^{-1}	global exposure level of the narwhal to the shore at time t	\mathbb{R}^+
$\Theta(t)$	rad	angle between the vector that goes from the nearest shore point to the narwhal's position and the empirical velocity vector at time t	$[-\pi, \pi]$

TABLE 2

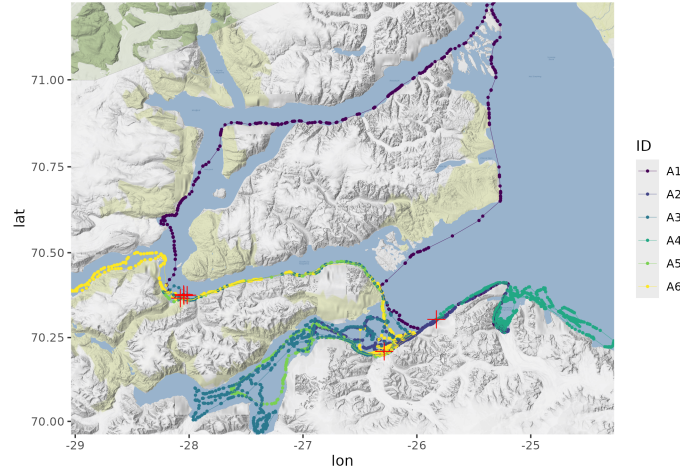
Summary of the covariates

APPENDIX B: PROOF OF PROPOSITION 3.1

Here, we prove Proposition 3.1 The proof is inspired by the results in (Gurarie et al., 2017; Johnson et al., 2008) and the proof of the transition density of the velocity process in (Albertsen, 2018).



(a) Tracks before exposure experiments



(b) Tracks during exposure experiments

Fig 3: Movement data of East Greenland narwhals. The red crosses indicate initial positions.

PROOF. The velocity process is an Ornstein-Uhlenbeck process. For $t \geq 0$ and $\Delta > 0$,

$$(1) \quad V(t+\Delta) = \exp(-A\Delta)V(t) + (I_2 - \exp(-A\Delta))\mu + \frac{2\nu}{\sqrt{\pi\tau}} \int_t^{t+\Delta} \exp(A(s-(t+\Delta)))dW(s)$$

It has Gaussian transition density with mean

$$(2) \quad \mathbb{E}(V(t+\Delta|V(t))) = \exp(-A\Delta)V(t) + (I_2 - \exp(-A\Delta))\mu$$

and covariance matrix

$$\begin{aligned} \text{Var}(V(t+\Delta)|V(t)) &= \frac{4\nu^2}{\pi\tau} \int_t^{t+\Delta} \exp(A(s-(t+\Delta))) \exp(A(s-(t+\Delta)))^\top ds \\ &= \frac{4\nu^2}{\pi\tau} \int_0^\Delta \exp(-Au) \exp(-Au)^\top du \end{aligned}$$

Since $\exp(-Au) = \exp(-\frac{u}{\tau})R_{-\omega u}$ where $R_{-\omega u}$ is the rotation matrix with angle $-\omega u$, the matrix product $\exp(-Au) \exp(-Au)^\top$ is simply $\exp(-\frac{2u}{\tau})I_2$. We deduce that the two com-

ponents of the velocity are independent and have the same variance, denoted $q_2(\Delta)$. The variance is

$$(3) \quad q_2(\Delta) = \frac{2\nu^2}{\pi} \left(1 - \exp\left(-\frac{2\Delta}{\tau}\right) \right).$$

These results are found in [Gurarie et al. \(2017\)](#). In the sequel, we use the notation

$$\zeta(t, s) = \frac{2\nu}{\sqrt{\pi\tau}} \int_t^s \exp(A(u-s)) dW(u).$$

Using that $V(s) = \mu + \exp(-A(s-t))(V(t) - \mu) + \zeta(t, s)$, we have

$$\begin{aligned} X(t+\Delta) &= X(t) + \int_t^{t+\Delta} V(s) ds \\ &= X(t) + \mu\Delta + \int_t^{t+\Delta} \exp(-A(s-t))(X(t) - \mu) ds \\ &\quad + \int_t^{t+\Delta} \zeta(t, s) ds \\ &= X(t) + \mu\Delta + (A^{-1}(V(t) - \mu) - A^{-1}\exp(-A\Delta))(X(t) - \mu) \\ &\quad + \int_t^{t+\Delta} \zeta(t, s) ds \end{aligned}$$

Thus,

$$(4) \quad X(t+\Delta) = X(t) + \mu\Delta + A^{-1}(I_2 - \exp(-A\Delta))(V(t) - \mu) + \xi(t, t+\Delta)$$

where $\xi(t, t+\Delta) = \int_t^{t+\Delta} \zeta(t, s) ds$. The location process is also Gaussian with mean

$$(5) \quad \mathbb{E}(X(t+\Delta)|V(t), X(t)) = X(t) + \mu\Delta + A^{-1}(I_2 - \exp(-A\Delta))(V(t) - \mu).$$

To get an expression of the covariance matrix, first rewrite

$$\begin{aligned} \xi(t, t+\Delta) &= \int_t^{t+\Delta} \frac{2\nu}{\sqrt{\pi\tau}} \left(\int_t^s \exp(-A((u-s))) dW(u) \right) ds \\ &= \frac{2\nu}{\sqrt{\pi\tau}} \int_t^{t+\Delta} (A^{-1} - A^{-1}\exp(A(u-t-\Delta))) dW(u) \\ &= \frac{2\nu}{\sqrt{\pi\tau}} \int_t^{t+\Delta} A^{-1}(I_2 - \exp(A(u-t-\Delta))) dW(u) \end{aligned}$$

Then use Ito's isometry

$$\begin{aligned} \text{Var}(X(t+\Delta)|X(t), V(t)) &= \frac{4\nu^2}{\pi\tau} \int_t^{t+\Delta} A^{-1}(I_2 - \exp(A(u-t-\Delta))) \\ &\quad \times (I_2 - \exp(A(u-t-\Delta)))^\top A^{-\top} du \\ &= \frac{4\nu^2}{\pi\tau} \int_0^\Delta A^{-1}(I_2 - \exp(-Ar)) \\ &\quad \times (I_2 - \exp(-Ar))^\top A^{-\top} dr. \end{aligned}$$

This integral can be computed explicitly since

$$A^{-1}(I_2 - \exp(-Ar))(I_2 - \exp(-Ar))^\top A^{-\top} = \frac{1}{C} f(r) I_2$$

where $f(r) = 1 - 2 \exp\left(-\frac{r}{\tau}\right) \cos(\omega r) + \exp\left(\frac{-2r}{\tau}\right)$ and $C = \frac{1}{\tau^2} + \omega^2$. We obtain that $X_1(t + \Delta)$ and $X_2(t + \Delta)$ are independent and have the same variance, denoted $q_1(\Delta)$. Writing $\sigma = \frac{2\nu}{\sqrt{\pi\tau}}$, the variance is

$$q_1(\Delta) = \frac{\sigma^2}{C} \left(\Delta - 2 \frac{\omega \sin(\omega\Delta) - \frac{1}{\tau} \cos(\omega\Delta)}{\frac{1}{\tau^2} + \omega^2} \exp\left(-\frac{\Delta}{\tau}\right) + \frac{\tau}{2} \left(\frac{\omega^2 - \frac{3}{\tau^2}}{\frac{1}{\tau^2} + \omega^2} - \exp\left(-\frac{2\Delta}{\tau}\right) \right) \right)$$

Now we compute the covariance between X and V to get the full covariance matrix of U :

$$\begin{aligned} \Gamma(\Delta) &= \frac{4\nu^2}{\pi\tau} \mathbb{E} \left(\left(\int_t^{t+\Delta} A^{-1}(I_2 - \exp(A(u-t-\Delta))) dW(u) \right) \right. \\ &\quad \times \left. \left(\int_t^{t+\Delta} \exp(A(s-(t+\Delta))) dW(s) \right)^\top \right) \\ &= \int_t^{t+\Delta} A^{-1}(I_2 - \exp(A(u-(t+\Delta)))) \exp(A(u-(t+\Delta)))^\top du \\ &= \frac{4\nu^2}{\pi\tau} \int_0^\Delta A^{-1}(I_2 - \exp(-Ar)) \exp(-Ar)^\top dr. \end{aligned}$$

Then,

$$A^{-1}(I_2 - \exp(-Ar)) \exp(-Ar)^\top = \frac{1}{C} \exp\left(-\frac{r}{\tau}\right) \begin{pmatrix} g(r) & h(r) \\ -h(r) & g(r) \end{pmatrix}$$

where

$$\begin{aligned} g(r) &= \frac{1}{\tau} \left(\cos(\omega r) + \exp\left(-\frac{r}{\tau}\right) \right) - \omega \sin(\omega r), \\ h(r) &= -\frac{1}{\tau} \sin(\omega r) + \omega \left(\cos(\omega r) - \exp\left(-\frac{r}{\tau}\right) \right). \end{aligned}$$

Finally we get

$$\gamma_1(\Delta) = \frac{\sigma^2}{2C} \left(1 + \exp\left(-\frac{2\Delta}{\tau}\right) - 2 \exp\left(-\frac{\Delta}{\tau}\right) - 2 \exp\left(-\frac{\Delta}{\tau}\right) \cos(\omega\Delta) \right),$$

$$\gamma_2(\Delta) = \frac{\sigma^2}{C} \left(\exp\left(-\frac{\Delta}{\tau}\right) \sin(\omega\Delta) - \frac{\omega\tau}{2} \left(1 - \exp\left(-2\frac{\Delta}{\tau}\right) \right) \right).$$

In the specific case $\omega = 0$, we obtain $C = \frac{1}{\tau^2}$ and the variance of X becomes

$$q_1(\Delta) = \sigma^2 \tau^2 \left(\Delta + 2\tau \exp\left(-\frac{\Delta}{\tau}\right) + \frac{\tau}{2} \left(-3 - \exp\left(-\frac{2\Delta}{\tau}\right) \right) \right).$$

Writing $\beta = \frac{1}{\tau}$ and reorganizing the terms, we obtain

$$(6) \quad q_1(\Delta) = \frac{\sigma^2}{\beta^2} \left(\Delta - 2 \frac{1 - \exp(-\beta\Delta)}{\beta} + \frac{1 - \exp(-2\beta\Delta)}{2\beta} \right).$$

This result match equation (6) in (Johnson et al., 2008). Similarly, in the case $\omega = 0$, we get $\gamma_2 = 0$ and the expression for γ_1 match equation (7) in (Johnson et al., 2008). \square

APPENDIX C: MEASUREMENT ERROR

In the application in Section 6, different values of the measurement error were estimated for the data before and after exposure: 35 m before and 48 m after. Both these values are consistent with the results in (Wensveen, Thomas and Miller, 2015). However, the post exposure estimation gave non-positive definite Hessian matrix for the negative log-likelihood, which prevents from using the information matrix equality to get confidence intervals of the estimates. Fixing a 35 m measurement error value when fitting the response model led to the same issue. We therefore tried different values of the measurement error, and kept the one that gave a positive definite hessian matrix and had the highest log-likelihood value. It turned out to be 50 m, very close to the initially estimated 48 m. Table 3 shows these results. In comparison, the final values of the log-likelihood when σ_{obs} is estimated from the data are respectively 4273 and 8043 before and after exposure, while the estimate of τ_0 is 1.10, and the estimates of α_τ and α_ν are respectively -4.19 and 0.66 , which is in the confidence interval of the final estimations we kept (those obtained for $\sigma_{obs} = 50$ m).

σ_{obs} (m)	$\hat{\tau}_0$	Baseline llk	Response llk	P.d hessian	α_τ	α_ν
30	0.96 ± 0.15	4261	8038	No	0.29	2.17
40	1.18 ± 0.16	4266	8014	No	-2.06	0.60
45	1.29 ± 0.17	4243	7965	No	-3.64	0.49
50	1.35 ± 0.16	4208	7861	Yes	-3.43 ± 0.70	0.74 ± 0.27
75	1.63 ± 0.19	3944	7225	Yes	-3.88 ± 0.73	0.76 ± 0.30
100	1.85 ± 0.22	3640	6590	Yes	-4.27 ± 0.74	0.63 ± 0.29

TABLE 3

Estimate for the baseline and response models for several fixed measurement errors.

APPENDIX D: CODE EXAMPLE

We illustrate briefly how to fit our baseline SDE model and obtain the results with smoothSDE R package Michelot et al. (2021). The version of the package including our new model is available [here](#). We suppose the package has been loaded. We consider a dataframe dataBE containing the preprocessed observations before exposure to the ship in the columns x and y, an animal identifier in a column ID and columns Ishore and AngleNormal for the covariates I^{shore} and Θ . The first step consists in choosing initial SDE parameters and model formulas. For the model we consider, there are five parameters μ_1 , μ_2 , τ , ν , and ω , and each of them needs a formula. Specification of the formulas is identical to the R package mgcv. Among the parameters, μ_1 , μ_2 will be set to 0, while τ , ν and ω are expressed as in section 4.1.

```

1 #number of observation
2 n_pre<-nrow(dataBE)
3
4 #initial parameters
5 par0 <- c(0,0,1,4,0)
6
7 #model formulas
8 formulas <- list(mu1 = ~1 ,mu2 =~1,tau =~s(ID,bs="re"),nu=~s(ID,bs="re"),
9 omega=~ti(AngleNormal,k=5,bs="cs")+ti(Ishore,k=5,bs="cs")+
10 ti(AngleNormal,Ishore,k=c(5,5),bs="cs"))

```

We then specify the measurement error for each observation in an array of covariance matrices. We suppose they are all diagonal with the same standard deviation `sigma_obs`. We will fix this measurement error.

```
1 # 50m measurement error
2 sigma_obs=0.05
3 H=array(rep(sigma_obs^2*diag(2), n_pre), dim=c(2,2,n_pre))
```

We can then create the SDE object as in [Michélot et al. \(2021\)](#). We choose the type of SDE in the argument `type`. Here, it is `RACVM` (see Section 3.1) since we want to include a non zero rotation parameter ω . The name of the columns where the observations are found is specified in the `response` argument. We specify the measurement error matrix `H` in the argument `other_data`. Fixed parameters are indicated in the argument `fixpar`.

```
1 #create SDE object
2 baseline_50m<- SDE$new(formulas = formulas, data = dataBE, type = "RACVM",
3 response = c("x", "y"), par0 = par0, other_data=list("H"=H),
4 fixpar=c("mu1", "mu2"))
```

To fix specific parameters in the statistical model, we need to use the `map` attribute. Here we use it to specify that the smoothing parameters should be fixed. Then we update the smoothing parameters to 1, and fit the SDE model.

```
1 #update map to fix smoothing parameters
2 baseline_50m$update_map(list("log_lambda"=factor(c(1,2,rep(NA,4)))))
3
4 #update smoothing parameters values
5 init_lambda=rep(1,6)
6 baseline_50m$update_lambda(init_lambda)
7
8 #fit the model
9 baseline_50m$fit()
```

The results of the optimization are stored in the attribute `tmb_rep`. We can extract the estimated parameters along with the standard errors.

```
1 #estimates
2 estimates_bas_50m=as.list(baseline_50m$tmb_rep(), what="Est")
3 #standard error
4 std_bas_50m=as.list(baseline_50m$tmb_rep(), what="Std")
```

Finally, we would like to plot all the smooth parameters as a function of the covariates. We can do it with the `get_all_plots` method. We only need to specify the range of each

covariate value we want to plot, a link function if we don't want to have directly the covariate on the x-axis but rather a function of the covariate, and the x-axis label of the plots. We put the option `show_CI="pointwise"` to show the pointwise confidence intervals on the plots.

```

1 #range of the covariates
2 D_low=0.073
3 D_up=3
4 xmin=list("Ishore"]=1/D_up)
5 xmax=list("Ishore"]=1/D_low)
6 #link function
7 link=list("Ishore")=(\ (x) 1/x))
8 #label
9 xlabel=list("Ishore"]="Distance to shore")
10
11 #draw plots
12 plots_bas_50m=baseline_50m$get_all_plots(model_name="baseline_50m",
13 xmin=xmin,xmax=xmax, link=link, xlabel=xlabel, show_CI="pointwise", save=TRUE)

```

REFERENCES

- ALBERTSEN, C. M. (2018). Generalizing the first-difference correlated random walk for marine animal movement data. arXiv:1806.08582 [q-bio].
- GURARIE, E., FLEMING, C. H., FAGAN, W. F., LAIDRE, K. L., HERNÁNDEZ-PLIEGO, J. and OVASKAINEN, O. (2017). Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement Ecology* **5** 13. <https://doi.org/10.1186/s40462-017-0103-3>
- JOHNSON, D. S., LONDON, J. M., LEA, M.-A. and DURBAN, J. W. (2008). Continuous time correlated random walk model for animal telemetry data. *Ecology* **89** 1208–1215. <https://doi.org/10.1890/07-1032.1>
- MICHELOT, T., GLENNIE, R., HARRIS, C. and THOMAS, L. (2021). Varying-Coefficient Stochastic Differential Equations with Applications in Ecology. *Journal of Agricultural, Biological and Environmental Statistics* **26** 446–463. <https://doi.org/10.1007/s13253-021-00450-6>
- WENSVEEN, P. J., THOMAS, L. and MILLER, P. J. O. (2015). A path reconstruction method integrating dead-reckoning and position fixes applied to humpback whales. *Movement Ecology* **3** 31. <https://doi.org/10.1186/s40462-015-0061-6>