

# Supplementary Material

# **1 SUPPLEMENTARY METHODS**

## 1.1 Statistical analysis of trajectories

We used multivariate hidden Markov models (HMM) to analyse the time-regular trajectories of 33 great white sharks (*Carcharodon carcharias*). HMMs are models for analysing time series of observations recorded at regular time intervals. They are now commonly used in the animal movement context for inferring the underlying movement modes that give rise to observed movement metrics (e.g., Langrock et al. (2012)). The classical HMM formulation assumes a first-order dependence between the underlying states, and between the observations and the underlying states. HMMs for animal movement often involve modelling two derived variables (one scalar and one angular) from relocation data, e.g., the step length and turning angle between consecutive observed locations in the case of terrestrial movement.

The trajectory reconstruction provided a step length and a turning angle. We modelled step length using a gamma distribution, which is suitable for modelling continuous, positive-valued variables, as in step length. We modelled the angular state variable using a wrapped Cauchy distribution, which has two parameters, the mean and concentration, where the latter relates to how peaked the distribution is. We only estimated the concentration parameter while assuming the mean to be zero, since we had no reason to expect turning to be systematically biased in one direction.

We intentionally kept the structure of the HMM simple to facilitate interpretability, since little is known about white shark movement behaviour, or the relationship between movement and fisheries interactions and use of MPAs. We used the results of the HMM in subsequent spatial analyses to 1) quantify overlap between shark movement and exposure to risk of capture in various fisheries, and 2) quantify overlap between shark movement and MPAs. This two-stage analysis is appealing because it allows for a simple, interpretable HMM, and distinct, interpretable spatial analyses. The downside of this approach is the lack of propagation of uncertainty from the estimated locations and state assignments through to the spatial analyses, and the lack of an explicit link between the inferred states and the spatial locations at which they occur. Under our current modelling setup we cannot answer the question "why do certain states happen in particular locations?". To answer this question we would need a more complex model structure that would likely hinder interpretability. With this in mind, and in light of the paucity of the current knowledge base about how white sharks use their environment, particularly within the South African EEZ, we chose to use this two-stage approach in this study (see Glennie et al. (2021) for more details on the pitfalls of HMMs for ecology). A map of state probabilities in included in the main manuscript (Figure 4, main manuscript) to make the degree of uncertainty in state assignments more explicit. A visual assessment suggests that state assignments are generally clear-cut, with low uncertainty, since most colours are near either end of the colour spectrum.

## 1.1.1 HMM structure and implementation

Following a standard animal movement HMM structure, an unobserved Markov chain is assumed to determine the behavioural states and the parameters of the state-dependent distributions associated with the observed variables (Zucchini et al. 2017). In this HMM analysis, the movement states are treated as unknown, and our research question relates to inferring information about the different types of movement carried out by white sharks. We fit one model to all sharks jointly and therefore estimate one set of parameters for all individuals.

The HMM parameters are estimated using numerical maximisation of the likelihood, implemented in R R Core Team (2019), using the momentuHMM package McClintock and Michelot (2018). In this package, the computation of the covariate-dependent transition probability matrices and the forward algorithm are coded in C++. The forward algorithm is an efficient way of evaluating the likelihood and is one reason for the popularity of HMMs – it makes them fast to fit. The forward algorithm corresponds to a recursive calculation of the likelihood with computational costs that scale linearly with the number of observed time points and this renders numerical maximum likelihood estimation feasible Zucchini et al. (2017).

Based on a lack of prior knowledge about white shark movement behaviour from the literature, and the scale of the error associated with the locations in the current dataset, we only fitted HMMs with 2 movement states in the interest of biological interpretability, following advice by Pohle et al. (2017). We expected sharks to exhibit more rapid, directed movement towards when travelling to an area of interest, compared to more sinuous, slower movement while in an area of interest, although these motivational assumptions are largely speculative in the absence of information about white shark movement ecology and hunting behaviour.

#### 1.1.2 Likelihood of the HMMs

Following Supplementary Material S4 from Photopoulou et al. 2020, we can write the likelihood of an HMM with N states and observation vectors  $z_1, ..., z_T$  can be written as a matrix product:

$$L = \delta \mathbf{P}(\mathbf{z_1}) \prod_{t=2}^{T} \Gamma_t \mathbf{P}(\mathbf{z_t}) \mathbf{1}'$$
(S1)

where  $\delta$  is a row-vector containing the initial state distribution,  $\Gamma_t$  represents the  $N \times N$  transition probability matrix at time point t and 1' is a row-vector of ones. P denotes a  $N \times N$  diagonal matrix containing the values of the N joint state-dependent densities evaluated at the observation vector  $z_t$ . We assume the observed variables to be contemporaneously conditionally independent, given the current state. Thus, for each state, the joint state-dependent density is the product of the univariate state-dependent densities which are associated to the observed variables.

In this analysis the observation vector at a given time step,  $z_t$  is of length two, and includes step-length and turning angle at each time step t along the movement track. To investigate the effect of season and individual covariates on movement behaviour, we include the temporal covariate day of the year (doy), the individual's sex, and its total length (TL) as covariates on the transition probabilities using a mutinomial logit link, so that the probability of transitioning from state i to state j at time t is:

$$\ln\left(\frac{\gamma_{ij}(t)}{\gamma_{ii}(t)}\right) = \beta_{0ij} + \beta_{1ij}\cos(2\pi \mathrm{doy}_t/365) + \beta_{2ij}\sin(2\pi \mathrm{doy}_t/365) + \beta_{3ij}\mathrm{sex}_t + \beta_{4ij}\mathrm{TL}_t \quad (S2)$$

Given we are not fitting any random effects, the log-likelihood of interest is the sum of log-likelihoods corresponding to the different sharks.

### 1.1.3 Model assessment

We used the AIC weights to choose between models with different covariates. We used the Viterbi algorithm to obtain the most likely state sequence for the model with most support, and we also calculated the state probabilities, which give the probability of being in a given state at each point in the observed time series and conveys uncertainty in state assignments. We examined model fit for each experiment by calculating the pseudo-residuals for each of the state-dependent variables (step length, turning angle) and

checking their distributions and the residual autocorrelation, using the acf function in R R Core Team (2019).

## 2 SUPPLEMENTARY RESULTS

## 2.1 Candidate models

In the main text we presented only the top model. Here we describe all models that we tried, including their AIC scores and AIC weights (Table S1). The top two models are essentially the same, since total length is used to define lifestage or maturity, but we present both models here for completeness.

#### 2.2 Diagnostics on best models

The distribution of pseudo-residuals for both state-dependent variables was symmetrical, suggesting no systematic lack of model fit (Figure S1). There was some residual autocorrelation for step length up to a lag of about 60 time steps, suggesting a more persistent serial dependence in consecutive step lengths than was captured by this model. The model was fitted at a 12-hour time step, so 60 steps translates to about 30 days. This might have to do with the time spent in a given state, and would be interesting to investigate in a future study.



**Figure S1.** Pseudo-residuals for each state-dependent variable (step length and turning angle) generated from a HMM for white shark movement. The x-axis shows the theoretical quantiles for the residuals, and the y-axis shows the density of residuals so that the area under each histogram sums to 1. The spread of the distributions for the residuals is slightly wide but they are both nice and symmetrical.

Table S1. All models fitted to the white shark tracking data. Model names correspond to the object names in the R code found in the GitHub repository https://github.com/theoniphotopoulou/white\_shark\_hmm (a final snapshot of this repository will replace this link for publication). The covariates used in the following models are sex (female, male), total length (TL; in centimetres), lifestage (juvenile, sub-adult, adult; derived from total length) and day of the year (doy; 1-365).

Model name	Covariata(s)	Acting on	AIC		
Widder fiame	Covariate(s)	Acting on	score	weight	
wsm6a	Sex + TL + cos(doy) + sin(doy)	Transition probabilities	159878	0.704	
wsm1	Flower distance	Transition probabilities	159879	0.294	
wsm4	Lifestage + $cos(doy) + sin(doy)$	Transition probabilities	159890	0.002	
wsm5	Sex + lifestage	Transition probabilities	159892	0.000	
wsm3	Sex + cos(doy) + sin(doy)	Transition probabilities	159904	0.000	
wsm2	$\cos(doy) + \sin(doy)$	Transition probabilities	159951	0.000	
wsm0	-	-	159956	0.000	
wsm7	Sex + TL, Cos(doy) + sin(doy)	Transition probabilities, Step length mean	No conv	ergence	

### 2.3 Summary of shark locations and fishing effort

Here we show annual breakdown of the number of active shark tags, the total daily locations and the total fishing effort as the sum of hooks for the two fisheries; pelagic and demersal shark longline.

**Table S2.** Annual summary of the number of active shark tags, daily locations and the number of hooks set in the pelagic and demersal longline fisheries.

Vaar	Number of active	Total daily	Total fishing effort (hooks)			
Ieal	shark tags	locations	Pelagic longline	Demersal longline		
2012	31	2392	4242642	440200		
2013	14	872	4778861	326670		
2014	4	76	3031507	137720		

#### 2.4 Administration of drugs and their dosages used on white sharks

The tagging program was designed to release tagged animals in good condition with 'normal' physiological and behavioural functioning. In South Africa, legally, a veterinarian must be present during surgical or drug procedures. A registered veterinarian from The Department of Agriculture Veterinary Services was present to monitor and assist with capture, tagging, sampling, surgery and drug administration of white sharks (*Carcharodon carcharias*) satellite tagged in this study.

Therapeutic drugs were administered as a precaution to captured white sharks to reduce the effects of stress and capture myopathy, metabolic acidosis and post-release infection (Table S3 and S4).

**Table S3.** A list of therapeutic drugs used on white sharks in the study to reduce stress, capture myopathy, metabolic acidosis and post-release infection.

Drug	Reason for use
Biosolamine (ml)	To reduce capture myopathy
PREDEF2X (ml)	An anti-inflammatory
Amikacin (1000mg/ml)	As a general antibiotic
Amikacin (500mg/ml)	As a general antibiotic
Vitamin B complex (ml)	To reduce stress
Vitamin AD3E (ml)	To reduce stress

Table S4.	The dosage	regimen (ml)	) used on	white sha	arks in th	e study to	o reduce	stress, c	apture m	yopathy,
metabolic a	acidosis and	post-release i	infection.	Dosages	were det	ermined l	based on	the estin	nated wei	ight (kg)
of the shar	k.	-		•						

Total	Mass	PREDEF2X	Biosolamine	Amikacin	Amikacin	Vitamin B	Vitamin
length	class	(ml)	(ml)	(1000mg/ml)	(500mg/ml)	complex	AD3E
(m)	(kg)			_	_	(ml)	(ml)
2.10	100	3	5	1.5	3	5.0	1
2.80	200	6	10	3.0	6	7.5	2
3.25	300	9	15	4.5	9	10.0	3
3.55	400	12	20	6.0	12	12.5	4
3.75	500	15	25	7.5	15	15.0	5
3.85	600	18	30	9.0	18	17.5	6
4.15	700	21	35	10.5	21	20.0	7
4.20	800	24	40	12.0	24	22.5	8
4.35	900	27	45	13.5	27	25.0	9
4.40	1000	30	50	15.0	30	25.0	10
4.55	1100	30	50	16.5	33	25.0	11
4.60	1200	30	50	18.0	36	25.0	12
4.70	1300	30	50	19.5	39	25.0	13
4.75	1400	30	50	21.0	42	25.0	14
4.80	1500	30	50	22.5	45	25.0	15
4.90	1600	30	50	22.5	45	25.0	16

# REFERENCES

- Glennie, R., Adam, T., Leos Barajas, V., Michelot, T., Photopoulou, T., and McClintock, B. (2021). Hidden Markov Models: Pitfalls and Opportunities in Ecology. *Resubmitted after minor revisions for Methods in Ecology and Evolution*
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93, 2336–2342
- McClintock, B. T. and Michelot, T. (2018). momentuhmm: R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution* 9, 1518–1530. doi:https://doi.org/10. 1111/2041-210X.12995
- Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics* 22, 270–293
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R* (Chapman and Hall/CRC)