

TEXT S1: Tag Data Filtering

Filtering included the removal of detections collected within the first 24 hrs of deployment to avoid confounding effects associated with tagging. Near-simultaneous detections at multiple receivers (i.e., < 5 s apart) were then aggregated into a single record. Consecutive detections from the same transmitter that occurred more frequently than the minimum transmission interval (i.e., < 60 sec) were also removed. Filtered acceleration and depth data were then paired based on the understanding that the tags transmit depth and acceleration at alternating intervals. To accomplish this, the two data sets were merged and ordered by time of detection. An R script was then used to recursively inspect the detections one-by-one to confirm that the type (i.e., acceleration or depth) alternated, the tag serial numbers matched, and detections were recorded within the maximum transmission interval (i.e., < 180 s). If consecutive detections met those rules, they were considered pairs and appended to a new dataset with temporally matched environmental data (i.e., AWAC, satellite, buoy data, and UNSO). Those detections that did not match were excluded from analysis.

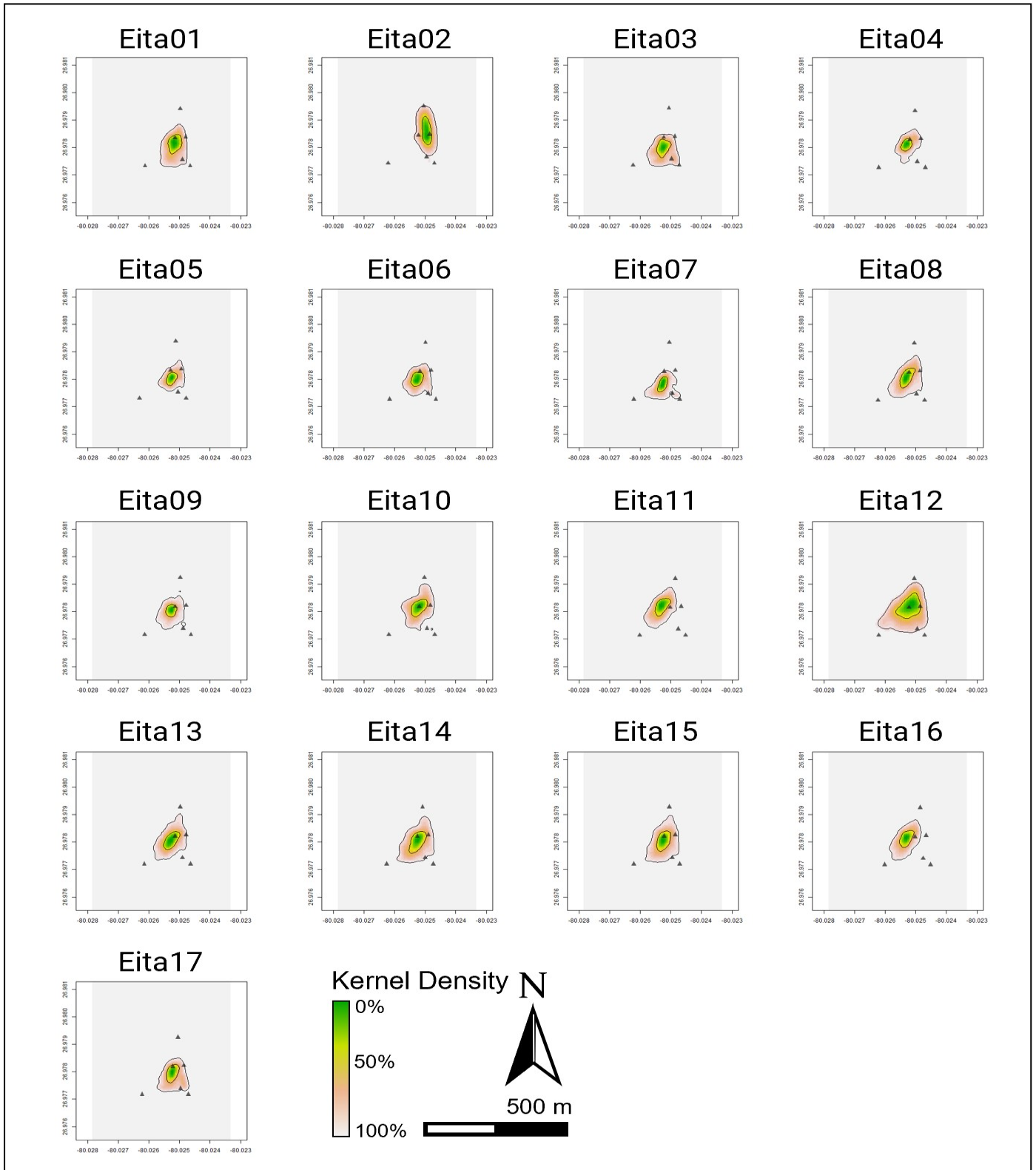


Fig. S1 Space use of individual goliath grouper. Black triangles are acoustic receivers, while space use is represented by utilization distributions.

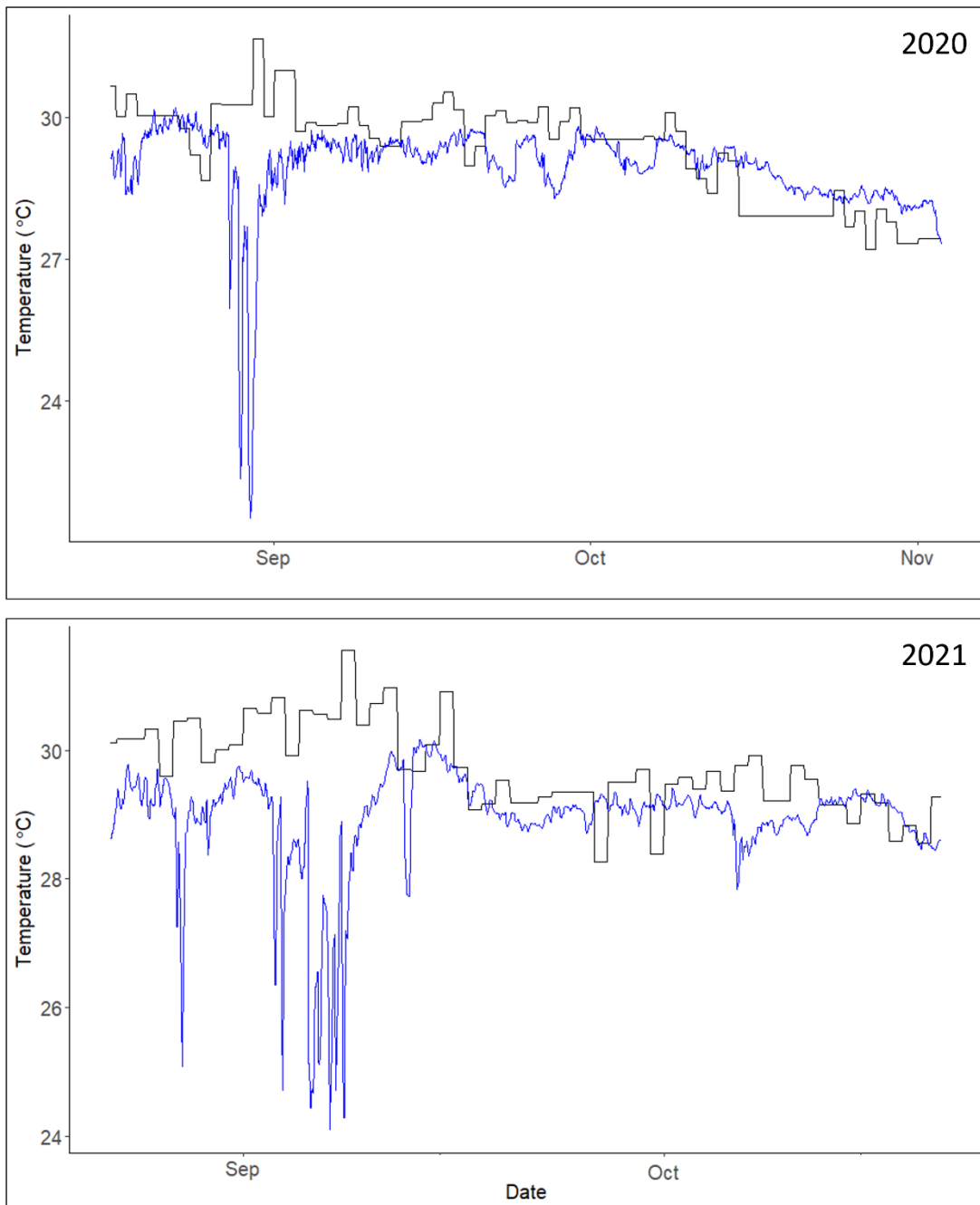


Fig. S2 Time series of bottom water temperature (blue line, measured by ADCP) and estimated sea surface temperature (black line, VIIRS) for 2020 and 2021 study periods.

TEXT S2: Hidden Markov models

An N -state hidden Markov model (HMM) is a doubly stochastic process composed of: (1) an observable state-dependent process, denoted by $\{X_t\}_{t=1}^T$, and (2) a (hidden) state process denoted by $\{S_t\}_{t=1}^T$ (Zucchini et al. 2017). The observations are taken to be conditionally independent and generated according to a set of state-dependent distributions, $\{f(X_t|S_t = n)\}_{n=1}^N$, and the states are taken to evolve over time according to a first-order N -state Markov chain with transition probability matrix (t.p.m.) Γ , with entries $\gamma_{ij} = \Pr(S_t = j | S_{t-1} = i)$ for $i, j \in \{1, \dots, N\}$. Lastly, the state at time $t = 1$ is taken to be generated according to the initial state distribution δ , with entries $\delta_n = \Pr(S_1 = n)$, for $n \in \{1, \dots, N\}$. In ecology, it is often of interest to understand how the probability of switching across states varies according to a set of covariates, e.g. time of day, temperature. We can reformulate the entries of the t.p.m. and write them in the following manner:

$$\gamma_{ij} = \frac{\exp(\alpha_{ij})}{1 + \sum_{l \neq i} \exp(\alpha_{il})} \text{ where } i, j = 1, \dots, N \quad (1)$$

where α_{ij} is assumed to be a constant in our baseline model, and $\alpha_{ii} = 0$, but later we will reformulate the transition probability to be a function of the environmental covariates of interest.

The diagonal entries of the transition probability matrix, suggest a tendency of the goliath groupers to continue in its current state, while the off-diagonal entries are equally informative, representing the probability of state change.

Our HMM analysis starts with a baseline HMM that assumes constant transition probabilities between states (Eq. (1)). Extending upon the baseline model, the full model integrates environmental covariates by allowing the transition probabilities to be time-varying (Leos-Barajas et al. 2016). Specifically, the transition probability from state i to state j at time t is reformulated to be a function of the environment covariates,

$$\gamma_{ij}(T_i, D_t, T_t, C_t, L_t) = \frac{\exp(\alpha_{ij} + \beta_{ij}^T T_i + \beta_{ij}^D D_t + \beta_{ij}^T T_t + \beta_{ij}^C C_t + \beta_{ij}^L L_t)}{1 + \sum_{k \neq i} \exp(\alpha_{ik} + \beta_{ik}^T T_i + \beta_{ik}^D D_t + \beta_{ik}^T T_t + \beta_{ik}^C C_t + \beta_{ik}^L L_t)} \quad (2)$$

On the right hand of the equation, the parameters α_{ij} represent the baseline log-odds of the transition, β_{ij}^T , β_{ij}^D , β_{ij}^T , β_{ij}^C , and β_{ij}^L are the coefficients quantifying the influence of each environmental covariates on the transition probability matrix entries at time t . By applying the traditional model selection technique, Akaike information criterion (AIC) to these different models, we will select the one with the smallest AIC.

Analysis

We used R (R Core Team 2022) to conduct the analysis and package `momentuHMM` (McClintock and Michelot 2018) to fit the hidden Markov model (HMM). To prepare our data for HMM, we conducted several essential adjustments to our data. This included the removal of duplicated sensor values and addressing missing data points. Notably, anomalies were observed in the data streams for Grouper 10410, leading us to conclude that its transmitter was lost. Consequently, we excluded this grouper from our analysis to maintain data integrity. Besides, we standardized our acoustic transmitter data stream to adhere to the regular time interval required by hidden Markov models. Given the irregular nature of the original data, we employed smoothing techniques, averaging data points within 10-minute intervals. This approach was selected after experimenting with various intervals ranging from 5 to 20 minutes. The 10-minute interval strikes a balance between providing an appropriate time range and generating sufficient data points to capture the essence of the data features, including data frequency and time interval characteristics. Lastly, we ordered the dataset by grouper ID and the collected dates and used the `prepData` function to prepare data for modelling.

With the prepared data, we used the `fitHMM` function to fit the baseline models without any covariates for year 2020 and 2021 separately, with the consideration that an eight-month time gap between the two-year data will lead to inadequate estimation. The `fitHMM` function uses the Viterbi algorithm to compute the most likely sequence of hidden states, which requires the user to provide number of states, probability distributions of the data streams, and the corresponding starting values for the distribution parameters.

For the choice of number of states, we compared the performance of a 2-state and 3-state HMMs (the most common choice in movement ecology literature, Bacheler et al., 2019) by a priori experiment, and the results suggested that a 2-state model adequately represented the behavior patterns. Therefore, we constructed a 2-state HMM to capture the predominant behaviors of the goliath grouper using acceleration rates as the primary observable data stream for both year data. The choice of probability distributions of the acceleration and the corresponding initial parameters were completely driven from exploratory data analysis (EDA) in Panel a) from Fig. S3. Specifically, we specified a log-normal distribution for all 2 states, as assuming a common distribution for all latent states is a classical way to fit HMMs (Leos-Barajas et al. 2016). The orange dashed lines in plots from Panel a), Fig. S3, indicated the initial location parameters (artificially suspect) for the log-normal distributions: for 2020, we specify $\mu_1 = 0$ for state 1 and $\mu_2 = -1.2$ for state 2; for 2021, we specify $\mu_1 = 0$ for state 1 and $\mu_2 = -1.5$ for state 2. Besides, we assumed the initial scales for both states to be $\sigma_1 = \sigma_2 = 0.2$ for both year model.

When fitting the full model, we standardized all covariates to mitigate the potentials for the impact of the covariates with small values (e.g., Water Current and Lunar Illumination) on the transition probabilities being obscured by the covariates with large values (e.g., Depth or Temperature), to improve model performance and interpretation. Moreover, the `fitHMM` function cannot handle missing covariates values, so we removed all the missing values from the covariates, which removed all the observations for one grouper in 2020 (ID: 10420) and another one in 2021 (ID: 10428). At the end of this stage, there were only 7 groupers remaining in 2020 and 8 groupers remaining in 2021 in our study data.

Results

We fitted a $N = 2$ HMM for all the groupers for each year separately, including the acceleration rate (m/s^2) series as the only observed state-dependent process. The estimated t.p.m. for both year 2020 and 2021 was

$$\hat{\Gamma}_{2020} = \begin{bmatrix} 0.938 & 0.062 \\ 0.038 & 0.962 \end{bmatrix}, \text{ and } \hat{\Gamma}_{2021} = \begin{bmatrix} 0.930 & 0.070 \\ 0.063 & 0.937 \end{bmatrix}$$

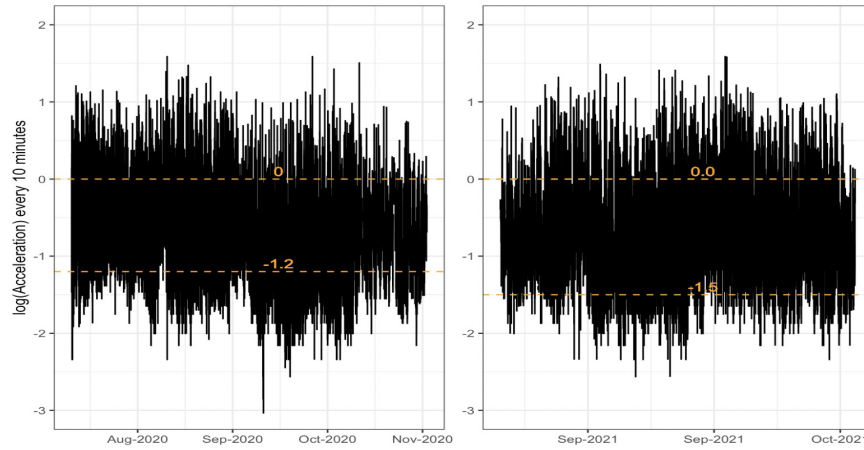
and the implied stationary distribution was (0.379 0.621) in 2020 and (0.471, 0.529) in 2021.

As a result, the baseline model successfully identified two biologically meaningful activity states (Fig. S3 b)). State 1 was characterized by a larger mean and variance in the acceleration rate, with the groupers moving at a mean rate of 0.73 m/s^2 every 10 minutes for 2020, and 0.67 m/s^2 every 10 minutes for 2021. This suggested that State 1 was a high activity state that might correspond to times when the groupers were actively travelling or foraging. Conversely, State 2 with a lower mean and variance in the acceleration rate, inferred that the groupers moved at a much lower mean rate of 0.34 m/s^2 every 10 minutes for 2020, 0.33 m/s^2 every 10 minutes for 2021. This suggested that the State 2 was a low activity state that might be relevant to the times when the groupers were resting. The distinction between the two stationary-state distribution was beneficial to decode the most likely state sequence behind the acceleration process (Fig. S3 c)) – we can see a relatively clear boundary between the colored points. This inferred that the baseline model, despite its simplicity, captured the main activity features performed by the goliath groupers in the dataset.

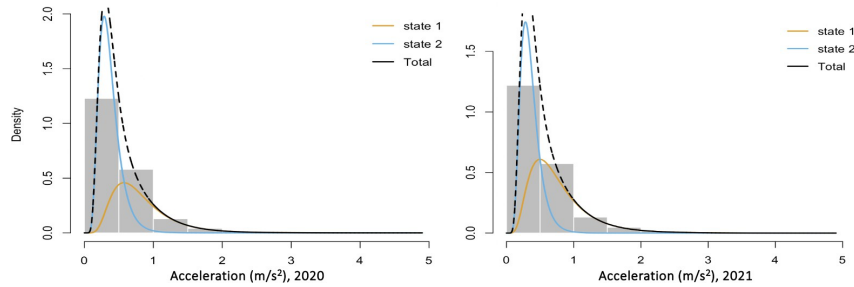
Finally, we experimented a 3-States HMMs without covariates, but the prediction performance did not provide significant improvement and the biological interpretability of the states worse off. Therefore, we considered only formulation with two states in the next section for model development.

Full Model

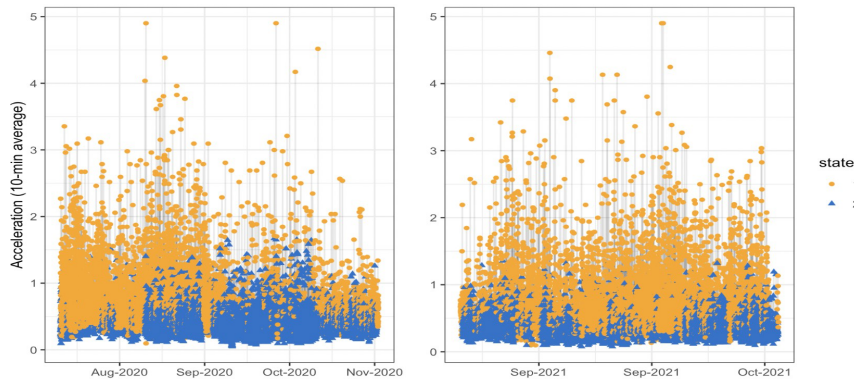
In this section, we extended the baseline model to allow the transition probabilities being a function of environmental covariates, including water current velocity (m/s), lunar illumination percentage, temperature ($^{\circ}\text{C}$), and Time of day (0-24 scale, with 12 represents midday). Moreover, the observed depth data stream helped to decode states in our prior experiment with 2-data-stream HMMs, so fish depth was included as covariate as well. We compared the models using AIC and selected the one with the smallest AIC value. In our analysis, the full model exhibited a smaller AIC compared to the base model.



a) Logarithm-transformed acceleration process smoothed by 10-minute interval



b) The Stationary State Distributions



c) Decoded Acceleration Time-Series

Fig. S3: HMM result for base model with left plots being 2020 and right plots being 2021. Panel (a) shows the time series plot for acceleration. The orange dashed lines indicated the choice for initial location parameters when fitting the HMM: the of the log-normal distributions, which were the assumed latent distribution of acceleration process. Panel (b) shows the resulted stationary state distributions, implying from the baseline 2-state HMMs had distinguishable shapes between the density lines, and Panel (c) shows the decoded state series, showing clear boundary between the two colours, which indicated the baseline model successfully identified the two hidden states.

LITERATURE CITED

- Bachelor, NM, Théo M, Robin TC, Shertzer KW. 2019. Fine-Scale Movement Patterns and Behavioral States of Gray Triggerfish *Balistes Capriscus* Determined from Acoustic Telemetry and Hidden Markov Models. *Fish. Res.* 215:76–89. <https://doi.org/10.1016/j.fishres.2019.02.014>
- Leos-Barajas V., Photopoulou T, Langrock R, Patterson TA, Watanabe YY, Murgatroyd M, Papastamatiou YP. (2016) Analysis of Animal Accelerometer Data Using Hidden Markov Models. *Meth. Ecol. Evol.* 8:161–73. <https://doi.org/10.1111/2041-210X.12657>
- McClintock BT, Theo M. (2018) MomentuHMM: R Package for Generalized Hidden Markov Models of Animal Movement. *Meth Ecol. Evol.* 9:1518–30. <https://doi.org/10.1111/2041-210X.12995>
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/>
- Zucchini W, MacDonald I, Langrock R. (2017). *Hidden Markov Models for Time Series: An Introduction Using R*, Second Edition. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press