Supplementary Materials and Methods

Bayesian modeling framework for the detection of heart beats from ECG Data

1 The model and the problem

We assume that the observed data are given by

$$y_t = \sum_k c_k s_{t-\tau_k} + w_t, \tag{1}$$

where c_k is an unknown constant, s_t is a known signal of finite duration L, τ_k is the delay of the kth replica of the signal, and the samples w_t are independent and zero mean Gaussian, i.e., $w_t \sim \mathcal{N}(0, \sigma^2)$, with the variance of the noise being unknown.

The problem is to estimate the time instants τ_k and treat the remaining unknowns as nuisance parameters. In solving this problem, we follow the Bayesian methodology.

2 Solution based on the raw data

Let us assume that the $(k-1)$st beat was detected at $\widehat{\tau}_{k-1}$. We assume that approximately the next beat is in the proximity of $\widehat{\tau}_{k-1} + \widehat{T}_k$, where $\widehat{T}_k = \widehat{\tau}_{k-1} - \widehat{\tau}_{k-2}$. The idea is to test several models of the form $p(\mathcal{M}_{\ell,k}|y_{t_1:t_2}, s, \tau_{\ell,k})$,

where ℓ is an index of the model ($\ell \in \{1, 2, \ldots, 2M+1\}$, with $2M+1$ being the number of models to choose from.[1]), $y_{t_1:t_2} \equiv \{y_{t_1}, y_{t_1+1}, \ldots, y_{t_2}\}$,[2] s is a known signal of length L,[3] and $\tau_{\ell,k} \in \{\widehat{\tau}_{k-1} + \widehat{T}_k - M, \widehat{\tau}_{k-1} + \widehat{T}_k - M + 1, \ldots, \widehat{\tau}_{k-1} + \widehat{T}_k + M\}$ is a candidate time of arrival of the kth beat.[4] The best model (or the best estimate of τ_k) will be the one that is associated with the model that has the highest a posteriori probability (MAP estimate), that is

$$\widehat{\tau}_k = \max \arg_{\tau_{\ell,k}} p(\mathcal{M}_{\ell,k} | y_{t_{1,k}:t_{2,k}}, s, \tau_{\ell,k}, \mathcal{D}^-_{t_{1,k}}), \tag{2}$$

where

$$t_{1,k} = \widehat{\tau}_{k-1} + \widehat{T}_k - M, \tag{3}$$

$$t_{2,k} = \widehat{\tau}_{k-1} + \widehat{T}_k + M + L - 1, \tag{4}$$

and $\mathcal{D}^-_{t_{1,k}}$ denotes past data with respect to $t_{1,k}$. We observe that all the models for the time of occurrence of the kth beat use the same vector of observations.

We write for the a posteriori probability of model $\mathcal{M}_{\ell,k}$,

$$P(\mathcal{M}_{\ell,k} | y_{t_{1,k}:t_{2,k}}, s, \tau_{\ell,k}, \mathcal{D}^-_{t_{1,k}}) \propto p(y_{t_{1,k}:t_{2,k}} | s, \tau_{\ell,k}, \mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k}) P(\mathcal{M}_{\ell,k} | \mathcal{D}^-_{t_{1,k}}), \tag{5}$$

where $p(y_{t_{1,k}:t_{2,k}} | s, \tau_{\ell,k}, \mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k})$ is the likelihood of the model $\mathcal{M}_{\ell,k}$, and $P(\mathcal{M}_{\ell,k} | \mathcal{D}^-_{t_{1,k}})$ is the a priori probability of that model. For the first factor

[1] The number of models may vary with time

[2] The interval $t_1 : t_2$ contains the kth beat.

[3] Without loss of generality, we assume that L is an odd number.

[4] This is the time instant of the first sample of s in the kth beat.

on the right hand side in (10) we have

$$p(y_{t_{1,k}:t_{2,k}}|s_{t-\tau_\ell}, \mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k}) = \int_0^\infty \int_{-\infty}^\infty p(y_{t_{1,k}:t_{2,k}}|s, \tau_{\ell,k}, c_{\ell,k}, \sigma_k^2, \mathcal{M}_{\ell,k}) \times p(c_{\ell,k}, \sigma^2|\mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k})\mathrm{d}c_{\ell,k}\mathrm{d}\sigma^2. \quad (6)$$

If we define the vector $\widetilde{s}_\ell \in R^{(2M+L)\times 1}$ by

$$\widetilde{s}_\ell = \begin{bmatrix} 0_{1:\ell} \\ s \\ 0_{\ell+L+1:2M+L} \end{bmatrix}, \quad (7)$$

we can write

$$p(y_{t_{1,k}:t_{2,k}}|s, \tau_{\ell,k}, c_{\ell,k}, \sigma_k^2, \mathcal{M}_{\ell,k}) = p(y_{t_{1,k}:t_{2,k}}|\widetilde{s}_l, c_{\ell,k}, \sigma_k^2, \mathcal{M}_{\ell,k}) = \frac{1}{\left(2\pi\sigma_k^2\right)^{\frac{2M+L}{2}}} \times \exp\left(-\frac{\left(y_{t_{1,k}:t_{2,k}} - c_{\ell,k}\widetilde{s}_\ell\right)^\top \left(y_{t_{1,k}:t_{2,k}} - c_{\ell,k}\widetilde{s}_\ell\right)}{2\sigma_k^2}\right), \quad (8)$$

where $y_{t_{1,k}:t_{2,k}} \in \mathbb{R}^{(2M+L)\times 1}$. We note that $\tau_{\ell,k}$ "decides" the position of the vector s in $\widetilde{s}_\ell$.

The factor in the integrand in (6), $p(c_{\ell,k}, \sigma_k^2|\mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k})$, is the joint prior of $c_{\ell,k}$ and σ_k^2 learned from previous samples. Here we have several options. One of them is to use the samples of y_t before the expected arrival of the kth beat (we use N of them) that do not have a beat, and for $p(c_k)$ to adopt the noninformative prior (i.e., $p(c_k) \propto$ const). Then we have

$$p(\sigma_k^2|\mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k}) \propto p(\sigma^2|y_{t_{1,k}-N:t_{1,k}-1})p(c_{\ell,k}) \propto p(\sigma^2|y_{t_{1,k}-N:t_{1,k}-1}). \quad (9)$$

For $p(\sigma_k^2|y_{t_{1,k}-N:t_{1,k}-1})$, we write

$$p(\sigma_k^2|y_{t_{1,k}-N:t_{1,k}-1}) \propto p(y_{t_{1,k}-N:t_{1,k}-1}|\sigma_k^2)\, p(\sigma_k^2)$$
$$= \frac{1}{\left(2\pi\sigma_k^2\right)^{\frac{N}{2}}} \times \exp\left(-\frac{y_{t_{1,k}-N:t_{1,k}-1}^{\top} y_{t_{1,k}-N:t_{1,k}-1}}{2\sigma_k^2}\right) \times \frac{1}{\sigma_k^2}, \tag{10}$$

where $p(\sigma_k^2) \propto \frac{1}{\sigma_k^2}$. Here, we recall that an inverse Gamma random variable X has a pdf given by

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}, \quad x > 0, \tag{11}$$

where $\alpha > 0, \beta > 0$ are the shape and scale parameters of the distribution. We recognize from (10) that $p(\sigma^2|y_{t_{1,k}-N:t_{1,k}-1})$ is the inverse Gamma distribution with parameters

$$\alpha = \frac{N}{2}, \tag{12}$$
$$\beta = \frac{y_{t_{1,k}-N:t_{1,k}-1}^{\top} y_{t_{1,k}-N:t_{1,k}-1}}{2}. \tag{13}$$

For simplicity in notation, in the next equations let $y = y_{t_{1,k}:t_{2,k}}$. Then we can rewrite (8) as

$$p(y|s, \tau_{\ell,k}, c_{\ell,k}, \sigma_k^2, \mathcal{M}_{\ell,k}) = p(y|\widetilde{s}_l, c_{\ell,k}, \sigma_k^2, \mathcal{M}_{\ell,k})$$
$$= \frac{1}{\left(2\pi\sigma_k^2\right)^{\frac{2M+L}{2}}} \times \exp\left(-\frac{\left(c - \widehat{c}_{\ell,k}\right)^2}{2\left(\widetilde{s}^\top \widetilde{s}\right)^{-1}\sigma_k^2}\right) \times \exp\left(-\frac{y^\top y - \widehat{c}_{\ell,k}\widetilde{s}^\top y}{2\sigma_k^2}\right), \tag{14}$$

where $\widehat{c}_{\ell,k} = \left(\widetilde{s}_\ell^\top \widetilde{s}_\ell\right)^{-1} \widetilde{s}_\ell^\top y$. First we integrate out c_k in (6) and obtain

$$p(y_{t_{1,k}:t_{2,k}}|s_{t-\tau_\ell}, \mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k}) \propto \int_0^\infty \frac{1}{(\sigma_k^2)^{\frac{2M+L-1}{2}}} \exp\left(-\frac{y^\top y - \widehat{c}_{\ell,k}\widetilde{s}^\top y}{2\sigma_k^2}\right) \times \frac{1}{\left(\sigma_k^2\right)^{\frac{N}{2}+1}} \times \exp\left(-\frac{\widetilde{y}^\top \widetilde{y}}{2\sigma_k^2}\right) \mathrm{d}\sigma_k^2, \tag{15}$$

where $\widetilde{y} = y_{t_{1,k}-N:t_{1,k}-1}$. From

$$\int_0^\infty x^{-\alpha-1} e^{-\frac{\beta}{x}} \mathrm{d}x = \frac{\Gamma(\alpha)}{\beta^\alpha}, \tag{16}$$

we deduce that

$$p(y|s_{t-\tau_\ell}, \mathcal{D}^-_{t_{1,k}}, \mathcal{M}_{\ell,k}) \propto \left(\frac{\widetilde{y}^\top \widetilde{y} + y^\top y - \widehat{c}_{\ell,k}\widetilde{s}^\top y}{2}\right)^{-\frac{2M+L+N-1}{2}}. \tag{17}$$

We use (17) to compute the probabilities of the various models. If we assume the uniform prior for the model, it is sufficient to compute the right hand side of (17) and pick the model with the highest value. Since $\widetilde{y}^\top \widetilde{y} + y^\top y$ is the same for the models, the best model is the one that is selected by

$$\begin{aligned} \mathcal{M}_{\ell,k} &= \arg\max_\ell \widehat{c}_{\ell,k}\widetilde{s}^\top y \\ &= \arg\max_\ell y^\top \widetilde{s}_\ell \left(\widetilde{s}_\ell^\top \widetilde{s}_\ell\right)^{-1} \widetilde{s}^\top y \\ &= \arg\max_\ell (\widetilde{s}_\ell^\top y)^2 \end{aligned} \tag{18}$$

because $\widetilde{s}_\ell^\top \widetilde{s}_\ell$ is not a function of ℓ. Thus, we have to use a simple correlator to pick the best model. However, we reiterate, we must use (17) to find the probabilities of all the models.

In summary, the approach can be implemented as follows:

1. Compute for all models $\widetilde{s}_\ell^\top y$.

2. Pick the model with the highest probability according to (17) (i.e., according to (18)).

3 An alternative approach

Next we describe a modified algorithm. This is the algorithm that was implemented. First, we form a sequence of differenced data by

$$z_t = y_t - y_{t-1}. \tag{19}$$

Now the model in (1) becomes

$$z_t = \sum_k c_k s'_{t-\tau_k} + w'_t, \tag{20}$$

where

$$s'_t = s_t - s_{t-1}, \tag{21}$$

$$w'_t = w_t - w_{t-1}. \tag{22}$$

With this preprocessing of the data we make z_t zero mean, but we add correlation to the noise (that is, w'_t and w'_{t-1} are correlated). When we downsample z_t by two, this correlation disappears. After downsampling z_t, we obtain $\widetilde{z}_t$, and we proceed with processing the data in the same way as described in the previous section. We keep in mind that now the signal template is the downsampled s'_t.

The method for processing the data has the following steps:

1. Apply a low pass filter on the data with a bandwidth of (0 Hz -150 Hz),
2. Perform local detrending (see the package),
3. Calculate the differenced data,
4. Downsample the data by two (the template should also be downsampled by 2).

Once the data are preprocessed, one estimates the time instants τ_k as described in Section 2.

Our last modification of the method allows for processing the data from both time directions, forward and backward. More specifically, we propose the following:

1. Implement the method from the first sample onward.
2. Repeat the process but starting with the last sample and moving backward (the signal template should also be reversed).
3. If there is an agreement in the estimated times of the peaks, the estimate is accepted. (First these estimates are accepted and then we proceed with the next step where the remaining estimates are accounted for.)
4. If there is a disagreement, delete the estimates that have probabilities less than 0.2. Otherwise, directly compare the probabilities of the two estimates and choose the one with a higher probability. An alternative is to use as an estimate the weighted average

$$\widehat{\tau}_k = p\widehat{\tau}_{f,k} + (1-p)\widehat{\tau}_{b,k}, \tag{23}$$

where $\widehat{\tau}_{f,k}$ and $\widehat{\tau}_{b,k}$ are the forward and backward estimates and p is the probability associated with $\widehat{\tau}_{f,k}$. Further details are provided in the program. The disagreement has to be smaller than the interval $\frac{1}{2}\widehat{T}$, where $\widehat{T}$ is the median of the period between two signal arrivals.

Classification of accelerometer data into behavior using a hidden Markov model

We used a previously developed and validated unsupervised hidden Markov model (HMM) to classify accelerometer-derived metrics into 30-second timeseries of three states ('flapping flight', 'soaring flight', and 'on-water'; Conners et al. 2021). In brief, we first derived two additional 30-s data streams ('features') as input for the HMM: 1) 'hf' - the highest dominant frequency in the dynamic heave acceleration signal and 2- 'p5' – the top 5th quantile in the static heave acceleration signal. These features effectively distinguish flight behaviors (flapping flight has high heave frequency while soaring flight has a low frequency) as well as on-water behavior (which has a lower static heave acceleration relative to flight). An optimization routine identified the best-fitting model from 25 runs where each iteration used a randomly generated set of starting values to fit each feature distribution. The best-fitted model was identified as the model with the largest maximum likelihood. A Viterbi algorithm then estimated the most likely sequence of states (behaviors) from the fitted model. The hidden Markov model for behavioral classification and was implemented in RStudio (v1.2.1335) with the R statistical language (v4.0.3) using the 'momentuHMM' package (McClintock & Michelot 2018).

Linear mixed models

A total of 3 models were built on 30-minute datasets using movement metrics as explanatory variables (Table S1): ODBA, VeSBA, and nFlaps). The interaction between percentage time soaring and ODBA and VeSBA was included to determine how flight mode influenced the relationships of these metrics with $\dot{V}O_2$. Initial models included species, sex, mass, and number of landings as fixed effects. Bird was included as a random effect, modeled with random intercepts, to control for individual variation and to account for uneven repeated sampling across birds. We used a stepwise deletion model selection technique to identify the most parsimonious final models by dropping non-significant terms and noting the change in Akaike's information criterion (ΔAIC). When running model selection procedures, we estimated mean and variance parameters using maximum likelihood (ML), while final models used restricted maximum likelihood (REML) to separate the likelihood estimates from fixed and random parameters. All model fits were evaluated by inspecting normality and heteroscedacity of residuals.

While a 30 min resolution accounts for physiological steady-state and is consistent with timescales that albatrosses use for behavioural decision making (e.g., 1-2 landings hr^{-1}, Weimerskirch & Guionnet 2002) other ecological questions around movement strategy and its consequences are better considered at longer-timescales. Here a daily time and energy budget is relevant as the accuracy of energetic estimates is likely to increase with increasing timescales (Green, 2011). To evaluate the effect of timescale on these estimates we ran a second set of LMMs ('validation models') to evaluate how well movement metrics predicted f_H-derived $\dot{V}O_2$ at increasing timescales (30-min, 12-hr, 24-hr; (Table 2). For example, to evaluate model performance at a daily (24-hour) scale, we used the mean values of 'model-predicted $\dot{V}O_2$' (predicted from the best-performing models described above) and 'f_H-derived $\dot{V}O_2$' within 24-hour windows along each bird's deployment. Predicted values were estimated both with and without bird identity as a random effect to evaluate how individual variability affected model predictions. Daily $\dot{V}O_2$ predicted from each model was then included as the explanatory variable in a series of LMMs, with f_H-derived $\dot{V}O_2$ as the response variable, and bird as a random effect to

account for the deployment durations. Equivalent procedures were performed at the 12-hr timescale and the 30-min. timescale. Predictive performance was evaluated by slope and R^2 (marginal and conditional) calculated with the 'MuMlm' R package (Bartón 2022).

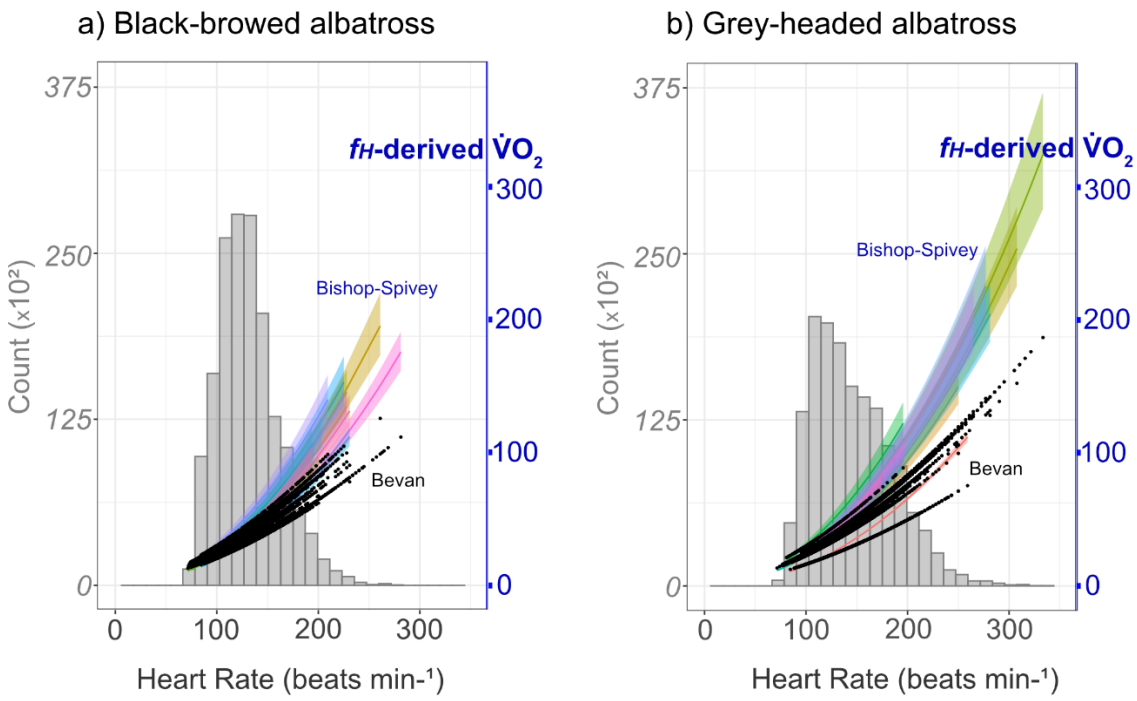

Fig. S1. Heartrate–$\dot{V}O_2$ curves from two methods: 1) Bevan et al., 1995 (black points) and 2) Bishop & Spivey (2013) (lines and confidence intervals colored by individual bird). Heartrate–$\dot{V}O_2$ curves from both methods are plotted over histograms representing the distribution of heartrates observed in the two study species. At higher heartrates, $\dot{V}O_2$ values estimated using the Bishop & Spivey method diverge from $\dot{V}O_2$ values estimated using the Bevan method; however, $\dot{V}O_2$ estimates from both methods overlap significantly for the majority of observed heartrates.

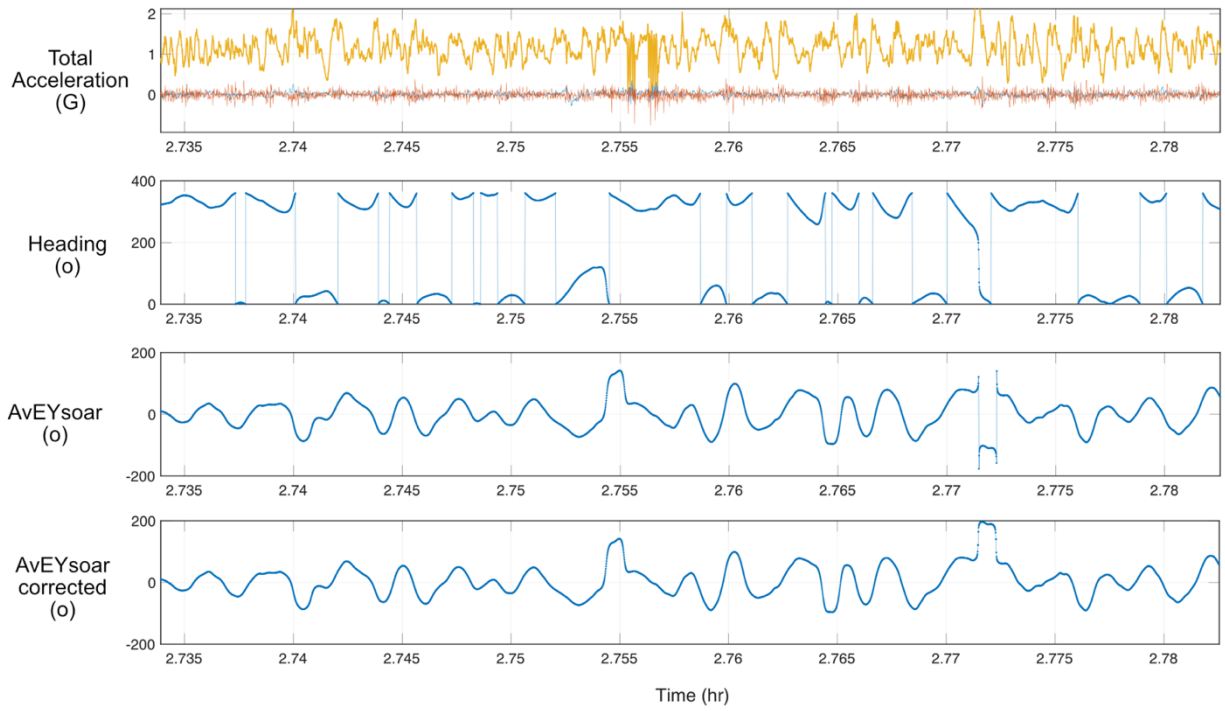

Fig. S2. An example of the derivation of $AvEY_{soar}$ and the post-derivation correction of $AvEY_{soar}$ used to obtain the size (in degrees) of soaring arcs performed by albatross. Total acceleration is provided to demonstrate that most of this timestamp contains soaring flight (Conners et al., 2021) except for a short flapping burst at 2.756 hours (as seen in the high frequency and high amplitude heave acceleration data stream (yellow)). Raw heading obtained from the magnetometer data (and corrected using pitch and roll data from the accelerometer) ranges from 0-360 degrees, with large jumps in the plot not representing changes in heading of that magnitude, but rather a radial crossing of the 0º|360º location on the circular axis. AvEYsoar was derived from the heading timeseries and represents the absolute change in degree headings over a 3-s moving window. In this example, a single large jump (>200º) in the data is observed around 2.78 hrs. This artifact was corrected (as detailed in the Methods) to produce the $AvEY_{soar}$ corrected timeseries.

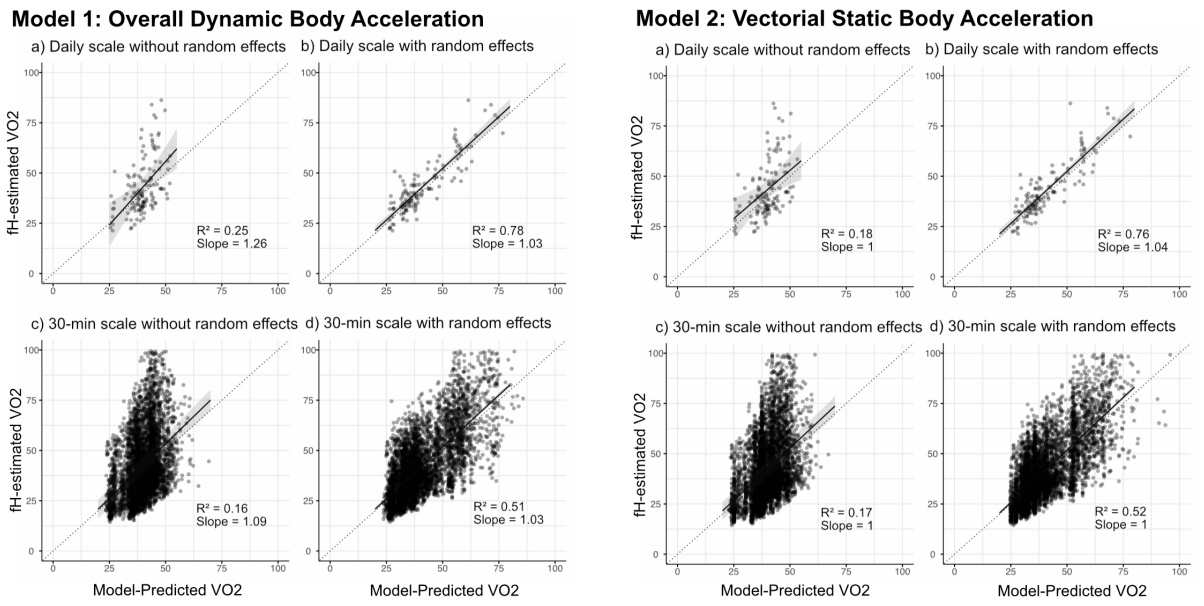

Fig. S3. Validation model results (with the exception of 'Model 3: nFlaps', displayed in Figure 4 of the main text) demonstrating the influence of timescale (daily vs. 30-min) and random effect of bird on model relationships. For each model, panels a) and b) represent the relationship between V O_2 and movement metrics at the daily scale while c) and d) represent the relationship at the 30-min scale. Panels b) and d) demonstrate the increase in predictive power of the model when the random effect of bird is included in the predictive function.

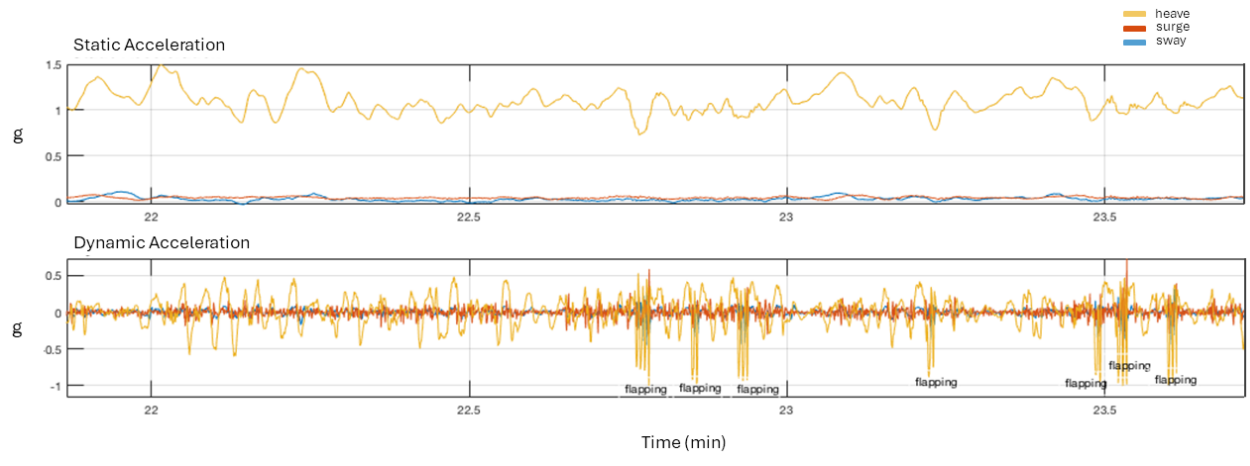

Fig. S4. Example of a ~2 sec timeseries of static and dynamic acceleration demonstrating patterns of acceleration during dynamic soaring and flapping flight. During a bout of dynamic soaring (~22-22.6 min), you can observe the large cyclical undulation of heave acceleration which occurs at a longer timescale than the large dynamic signals of flapping (starting at ~22.75)

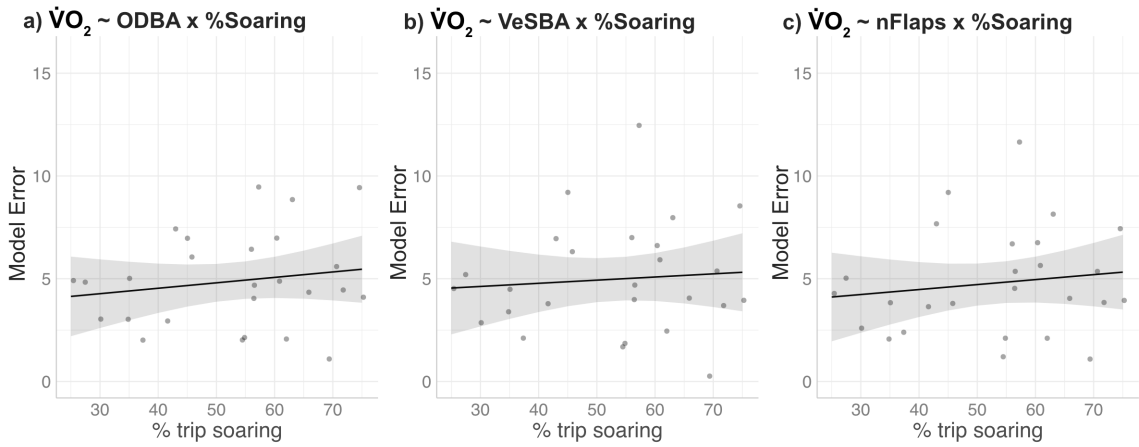

Fig. S5. For each bird's dataset, mean model divergence is plotted against percent of the trip spent soaring. A general trend of increasing divergence with increasing percent time soaring is observed; however, this trend was not significant ($p>0.05$) for all models. Model divergence was calculated by taking the absolute value of the difference between model-predicted V O_2 and f_H-derived V O_2, then taking the mean difference of each birds' dataset. Datasets used here were from the daily time scale.

Table S1. The final set of models evaluating relationships between movement metrics and *fH*-derived $\dot{V}O_2$ (the dependent variable). Estimates and confidence intervals were back-transformed from the log-scale, since *fH*-derived $\dot{V}O_2$ was log transformed in all models.

Model	Term	Estimate	conf.low - conf.high	Df	p-value
1. ODBA:pSoar + nLandings + mass	Intercept	7.74	3.620 - 16.540	5182	<0.001
	ODBA	4.16	3.606 - 4.787	5182	<0.001
	pSoar	1.01	1.005 -1.006	5182	<0.001
	nLandings	1.04	1.026 - 1.045	5182	<0.001
	Mass	1.45	1.132 - 1.85	24	0.0037
	ODBA:pSoar	0.98	0.980 - 0.985	5182	<0.001
2. VeSBA:pSoar + nLandings + mass	Intercept	0.94	0.437 - 2.015	5182	<.0001
	VeSBA	11.49	9.185 - 14.363	5182	0.0179
	pSoar	1.03	1.023 - 1.029	5182	<.0001
	nLandings	1.05	1.046 - 1.054	5182	<.0001
	Mass	1.45	1.143 - 1.836	24	0.0035
	VeSBA:pSoar	0.98	0.973 - 0.978	5182	<.0001
3. nFlaps + nLandings + mass	Intercept	11.36	5.438 - 23.737	5184	<.0001
	nFlaps	1.00	1.0003 - 1.0003	5184	<.0001
	nLandings	1.03	1.024 - 1.042	5184	<.0001
	Mass	1.45	1.140 - 1.840	24	0.0039

References

Bartón K (2022) MuMIn: Multi-Model Inference.

Green JA (2011) The heart rate method for estimating metabolic rate: Review and recommendations. Comp Biochem Physiol Part A 158:287–304

McClintock BT, Michelot T (2018) momentuHMM: R package for generalized hidden Markov models of animal movement. Methods Ecol Evol 9:1518–1530

Weimerskirch H, Guionnet T (2002) Comparative activity pattern during foraging of four albatross species. Ibis (Lond 1859) 144:40–50