Appendices

A A random walk as a special case of the ESF

The simple isotropic random walk (SRW) is unbiased and uncorrelated, with diffusion coefficient D, such that the density of a step starting at $\boldsymbol{x}$ and ending at $\boldsymbol{y}$ over time interval t is normally distributed around the start point: $y \sim N(x, \sigma^2)$, where $\sigma^2 = 4Dt$ (Codling et al., 2008). The likelihood of a step ending at $\boldsymbol{y}$ given that it started at $\boldsymbol{x}$ is $f(\boldsymbol{y}|\boldsymbol{x}) = \psi(\boldsymbol{y}|\boldsymbol{x}, \sigma^2)$, where ψ is the probability density function (PDF) of a normal distribution of $\boldsymbol{y}$ with mean $\boldsymbol{x}$ and variance σ^2.

We can write the SRW as a special case of the ESF. The ESF likelihood of a step ending at $\boldsymbol{y}$ given that it started at $\boldsymbol{x}$ is

$$f(\boldsymbol{y}|\boldsymbol{x}) = S^{-1} \exp[\beta_1 G(\boldsymbol{x}, \boldsymbol{y}) - \beta_2 C(\boldsymbol{x}, \boldsymbol{y})] \tag{1}$$

where S^{-1} is a normalization constant that ensures it is a PDF of $\boldsymbol{y}$. To write it as an SRW, we can set $\beta_1 = 0$ to represent no effect of energetic gains and $C(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x})^2$. The likelihood then becomes $f(\boldsymbol{y}|\boldsymbol{x}) = S^{-1} \exp[-\beta_2(\boldsymbol{y} - \boldsymbol{x})^2]$, which can be rewritten as

$$f(\boldsymbol{y}|\boldsymbol{x}) = S^{-1} \exp\left[-\frac{(\boldsymbol{y} - \boldsymbol{x})^2}{2\sigma^2}\right] \tag{2}$$

where $\beta_2 = \frac{1}{2\sigma^2}$. We recognize this as the PDF of a bivariate normal distribution with variance σ^2, mean $\boldsymbol{x}$, and $S = 2\pi\sigma^2$. This shows that an ESF with no gains and $C(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x})^2$ (i.e. costs formulated as the step length squared) is equivalent to an SRW model.

B Radius size, R

To approximate Equation 2, we generate controls on a disc (Section 2.3). By using this approximation method, we therefore assume that the probability density function of a step ending at $\boldsymbol{y}$ given that it started at $\boldsymbol{x}$ over the area of the disc, is

$$f(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} \exp\{\beta_1 G(\boldsymbol{x}, \boldsymbol{y}) - \beta_2 C(\boldsymbol{x}, \boldsymbol{y})\} & \text{if } l_{xy} \leq R \\ 0 & \text{if } l_{xy} > R \end{cases} \tag{3}$$

Therefore, for this approximation to be accurate, the disc needs to be large enough so that the probability of a step longer than R is very small (Figure S1). If we define the radius as $R = l_m \times \gamma$, where l_m is the maximum observed step length, the approximation will improve as γ increases. However, as the size of the disc becomes larger, so does the number of controls needed for the approximation. There is no straightforward way to assess this trade-off (i.e. the optimal size of R), but we can use importance sampling, based on where we expect the ESF to take large values. We explore the effect of the size of R on the approximation using simulated data, as well as comparing individual bear estimates approximated with different values of γ.

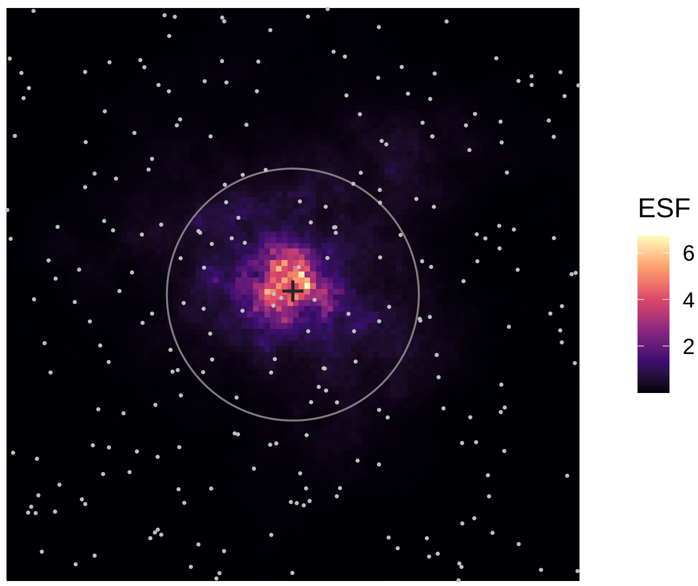

Figure S1: Plot illustrating importance sampling in the ESF. Ideally, we wish to sample uniformly over the entire study area (white dots). However, the ESF will decay with distance from the start point (+), due to the effect of step length on costs, and controls generated outside the disc will contribute very little to the approximation (i.e. their ESF is nearly zero). Therefore, for computational convenience, we can just sample within the disc, as long as the radius is large enough.

Simulated data We simulated 250 movement tracks $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ of length $n = 250$, as described in Section D. For each step, we generated 50 controls on a disc with a radius of the size $R = l_m \times \gamma$, where $\gamma = 0.5, 1.1, 2$. We fit the ESF for each movement track. As expected, β_2 was estimated with the lowest precision with the smallest radius ($\gamma = 0.5$). It was estimated correctly when $\gamma \geq 1.1$ (Figure S2). However, this represents a simplistic example, where the costs are *only* dependent on step length and $\beta_2 = 15$ is fairly strong selection against costs (i.e. the step length distribution should quickly decay to 0).

Real data We checked the effect of radius size on our polar bear telemetry data. We generated controls on a disc with $R_1 = 1.1 \times l_m$ and $R_2 = 2 \times l_m$ for each individual, fit the models separately, and then

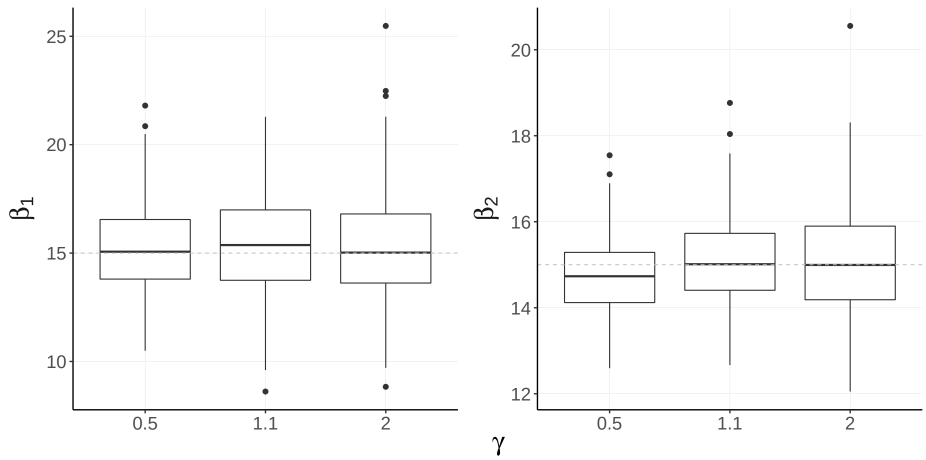

Figure S2: Estimates of β_1 and β_2 with $R = \gamma \times l_m$, where l_m is the maximum observed step length and $\gamma = 0.5, 1.1, 2$. Dashed line represents the true parameter value.

compared parameter estimates. Estimates varied up to ± 0.15 for β_2 (Figure S3), but followed the same general pattern. There was no evidence of systematic bias (i.e. underestimation or overestimation), and variation may be explained by the random generation of controls (which varied between the two trials). β_1 also varied between the two radius sizes, but this is likely attributable to high uncertainty in the estimates (i.e. no clear selection for gains).

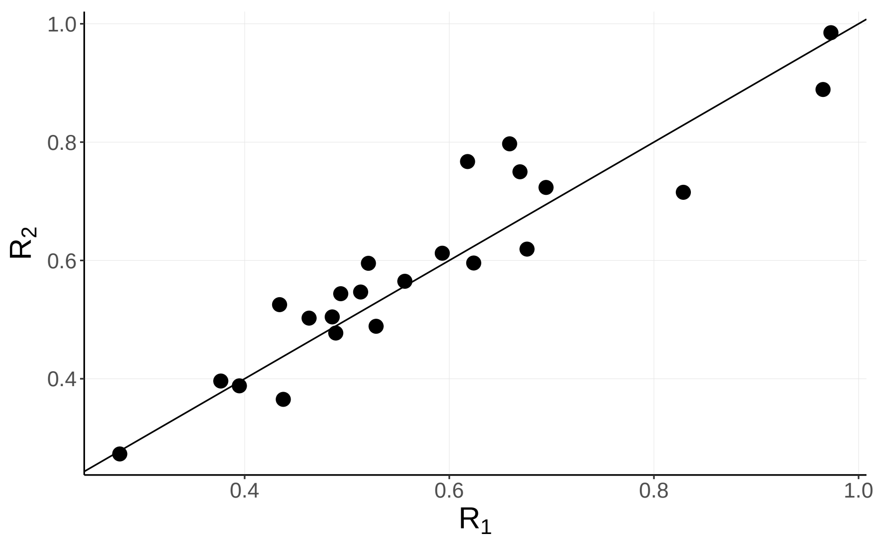

Figure S3: Individual estimates of β_2 with $R_1 = 1.1 \times l_m$ and $R_2 = 2.0 \times l_m$. Each point is an individual polar bear and the straight line represents a 1:1 relationship.

C Examples of gains and costs formulations

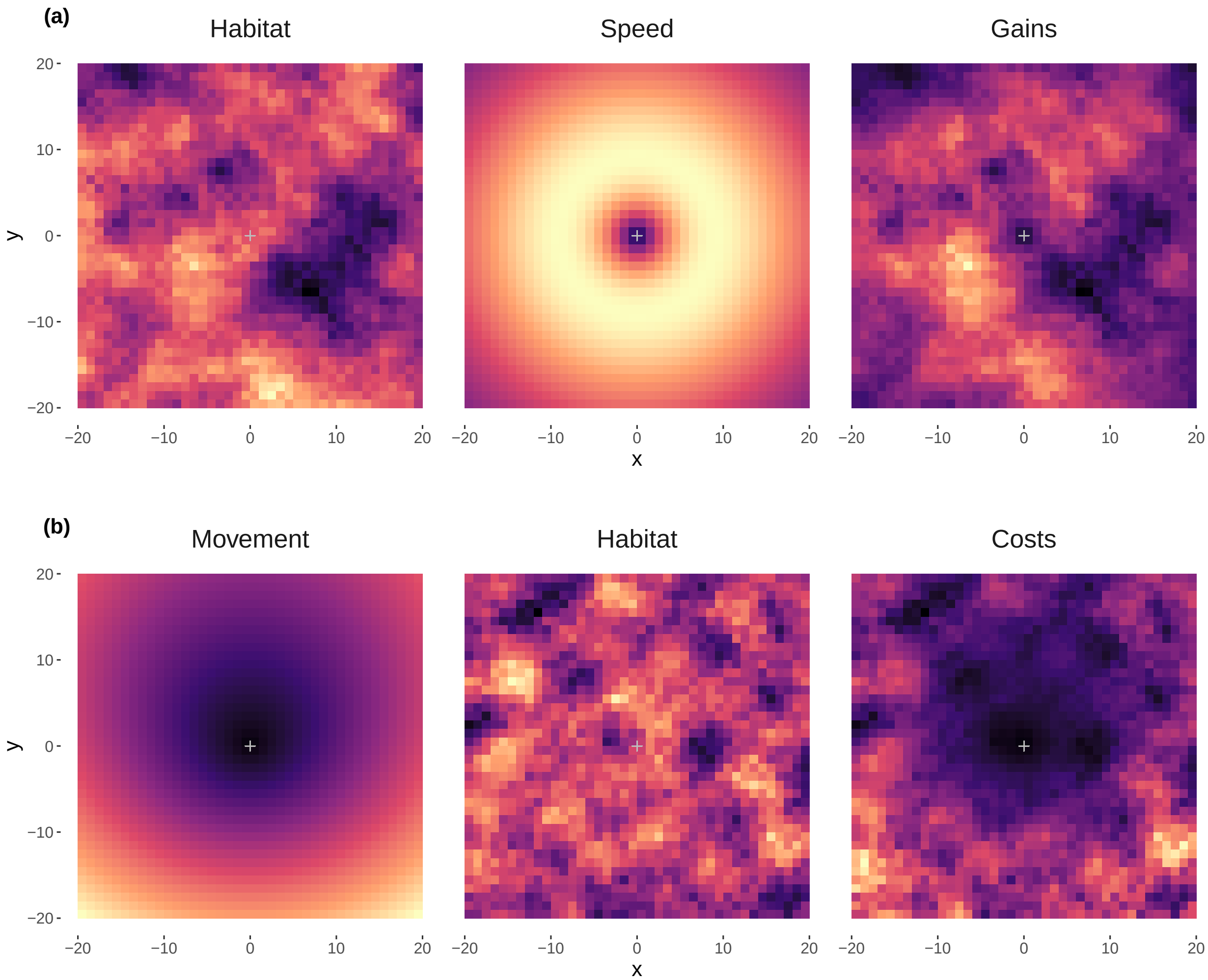

Figure S4: Example energetic gain G (a) and energetic cost C (b) formulations. In all panels, higher values are lighter in colour. In both (a) and (b), the third panel is a product of the first two panels, which represent movement and habitat components. In (a), energetic gains are composed of an energetically beneficial habitat covariate (e.g. forage biomass) scaled to the speed travelled. In this case, the effect of movement speed is gamma distributed ($k = 2$, $\theta = 2.2$) about the start point to represent decreased foraging potential at low and high speeds. In (b) the energetic costs are defined by the distance and turning angle from the start point (+; assuming movement up the y-axis), combined with a habitat covariate in which higher values increase energy expenditure.

D Simulation study

We ran simulations to assess the performance of the ESF inference method under different scenarios (i.e. to test the accuracy of the approximation; Section 2.3). The main objective was to recover model parameters from movement tracks simulated directly from the ESF, with known parameter values. For all simulations, G was defined as a random covariate field and C was calculated as the step length, both from $[0, 1]$ and assumed to be in the same units.

Algorithm We generated n locations $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$, with $\boldsymbol{x}_1$ selected randomly from the study area Ω. For each iteration $i = 1, \ldots, n-1$, we followed these steps to generate $\boldsymbol{x}_{i+1}$:

1. Simulate possible endpoints $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_K\}$ uniformly on a disc centred on $\boldsymbol{x}_i$, with a radius $R = 1$ and $K = 10,000$. We chose a very large value of K to ensure any bias was the result of the approximation, rather than the simulation itself.

2. Evaluate G and C at each endpoint.

3. For $k \in 1, 2, ...K$, sample x_{i+1} from $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_K\}$, with probabilities defined by

$$p_k = \frac{w(\boldsymbol{x}_i, \boldsymbol{z}_k)}{\sum_{j=1}^{K} w(\boldsymbol{x}_i, \boldsymbol{z}_j)}, \tag{4}$$

   where w is the ESF (Equation 2 of the main text).

Scenarios First, we assessed whether the selection strength affected our ability to estimate the parameters, and subsequently, whether certain foraging strategies may be harder to identify. For both β_1 and β_2, we considered 15 as a high parameter value. We considered low parameter values to be 0 for β_1 (no selection for gains), and 5 for β_2 (very weak selection against costs). We could not use 0 for β_2, as the ESF simulation algorithm would artificiality constrain the step length through the radius model. We tested different values of β_2 and found that 5 was the lowest parameter value where the size of the radius no longer affected the simulated step lengths. We combined these parameter values to represent the following movement patterns: i) optimal movement (high values of both parameters), ii) intake maximization (high β_1, low β_2), iii) cost minimization (low β_1, high β_2), and iv) movement nearly free of energetic considerations (low values of both parameters). Next, we altered the level of spatial autocorrelation in G. We simulated the study area Ω as a 1000×1000 raster with a resolution of 0.25, and assigned each grid cell a random value [$\sim U(0, 1)$]. We calculated the covariate field for G by using a circular moving average window with diameter ρ (measured in grid units) to control the degree of spatial autocorrelation (Avgar et al., 2016; Michelot, 2019). We created random rasters of G with $\rho = 1, 5, 10, 25$ to reflect four levels of spatially autocorrelated habitat. For each of the 16 scenarios (parameter sets and spatial autocorrelation), we generated 250 movement tracks $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ of length $n = 250$. For each track, we tested the inference method using 20 and 200 control locations in the Monte Carlo integration procedure (Section 2.3). All parameters were estimated using MLE.

Results In most cases, the parameters were estimated accurately, although β_2 was generally estimated more precisely than β_1 (Figure S5). The median (min, max) difference between estimated and known parameter values was -0.04 (-65, 27) for β_1 and 0.04 (-4.8, 4.7) for β_2. Spatial autocorrelation in G had a noticeable effect on the precision of β_1 estimates, but not β_2. When β_1 was high, there was a pattern of decreased precision with increased autocorrelation. When β_1 was low, precision was lowest when spatial autocorrelation was very low ($\rho = 1$) and very high ($\rho = 50$). Spatial autocorrelation is a documented issue in resource selection analyses, which can lead to biased parameter estimates (Northrup et al., 2013). In SSFs and ESFs, high spatial autocorrelation decreases the range of the covariate space that may be explored for each movement step, which may decrease the ability to infer selection, particularly when the number of control locations is low (Northrup et al., 2013). However, in our simulations, the number of control locations used in Monte Carlo integration had negligible effects on the precision or accuracy of parameter estimations. Therefore, in most cases, 20 control locations should be adequate to approximate the likelihood, although we still recommend caution when working with highly spatially autocorrelated environmental covariates. As noted in Fortin et al. (2005), more controls may also be necessary when covariates are rare.

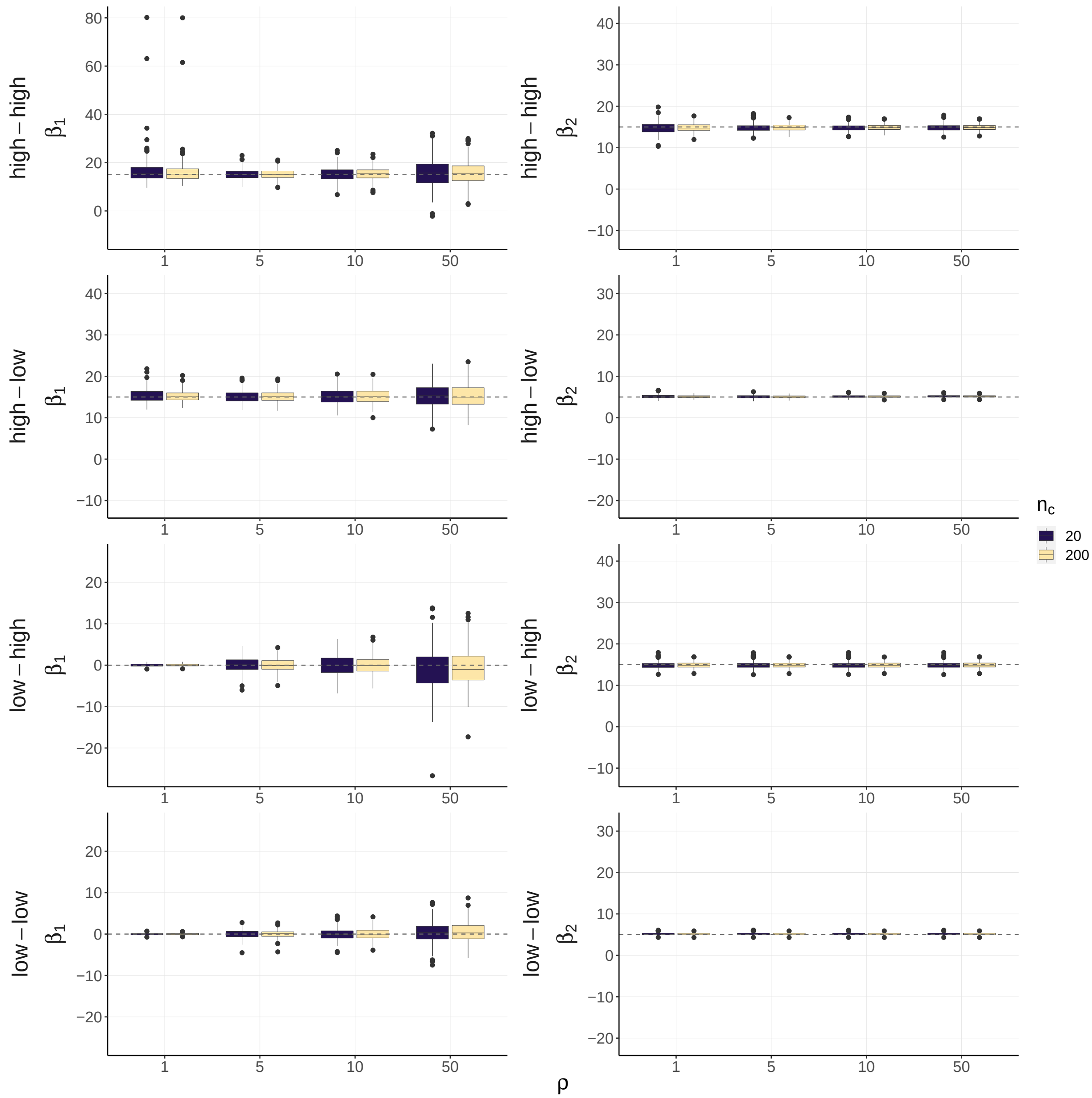

Figure S5: Parameter estimates from the simulations, under 32 different scenarios. Tracks were simulated either of four sets of parameters: "high-high" ($\beta_1 = 15$, $\beta_2 = 15$), "high-low" ($\beta_1 = 15$, $\beta_2 = 5$), "low-high" ($\beta_1 = 0$, $\beta_2 = 15$), and "low-low" ($\beta_1 = 0$, $\beta_2 = 5$). ρ refers to the level of spatial autocorrelation in the energetic gain covariate G, and n_c is the number of controls used in Monte Carlo integration. Dashed line is the true parameter value.

E Case study

E.1 Study Area and Field Sampling

Field sampling was done in Beaufort Sea, Canada (Figure S6). Sea ice in the area is mostly annual, with a flaw lead that separates near-shore areas of stable landfast ice and off-shore drifting pack ice (Carmack & Macdonald, 2002). The lead widens in spring and forms an active sea ice zone with high productivity (Pilfold et al., 2014), before most ice disappears by mid-summer (Stern & Laidre, 2016). Sea ice drift is characterized by the clockwise Beaufort Gyre, which is strengthening with climate change (Hutchings & Rigor, 2012; Petty et al., 2016), and increasing the energetic expenditure of polar bears in the area (Durner et al., 2017). Following standard capture procedures (Stirling et al., 1989), polar bears were sighted and immobilized in April-May of 2007-2011. Bears were fitted with GPS collars (Telonics, Mesa, AZ) set to a collect locations at a 4-hour resolution (relayed via the Argos satellite system; CLS America, Lanham, MD), and programmed to release after 1-2 years. The age of each bear was determined by analysing cementum growth layers of an extracted vestigial premolar (Calvert & Ramsay, 1998), and sex was determined in the field. Capture and handling was approved by the University of Alberta BioSciences Animal Care and Use Committee following guidelines from the Canadian Council on Animal Care.

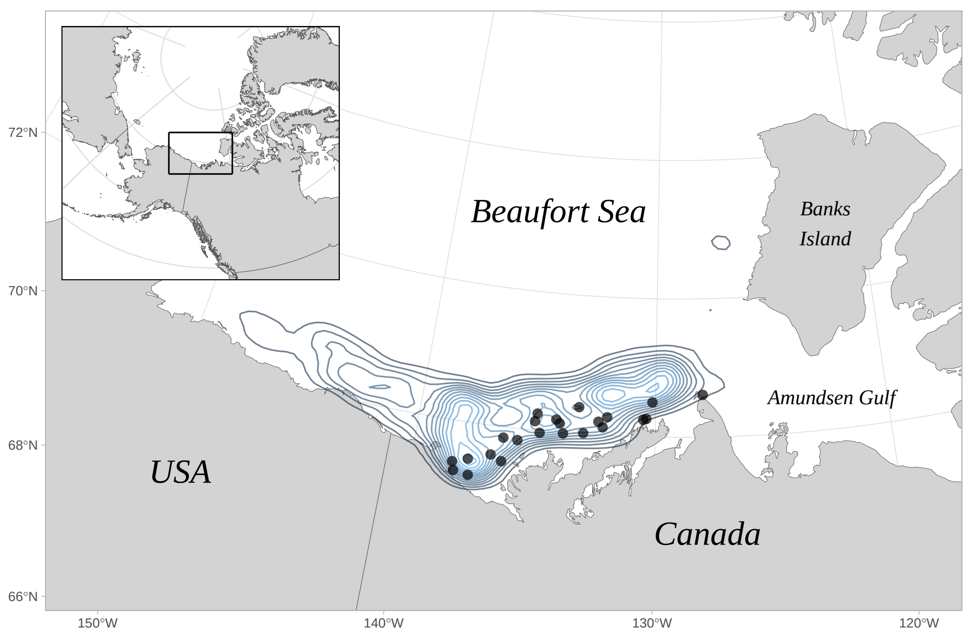

Figure S6: Study area in the Beaufort Sea, Canada. Circle points are polar bear collar deployment locations, and contour lines show the density of satellite telemetry data for all individuals (once regularised and limited to the spatiotemporal extent of the energetic gains raster).

E.2 Polar bear cost modelling

In this appendix, we present high-resolution complements to Figure 3 of the main text (Figures S7, S8), as well as a comparison of our cost model to doubly labelled water (Figure S9). In figure S9, we compared the estimated movement costs from our cost model (described in Section 3.2) with estimates of energy expenditure from doubly labelled water (Pagano & Williams, 2019). We estimated mean daily movement speeds and energy expenditure for adult female polar bears with > 6 locations per day for ≥ 25 days. We used linear regression to estimate the relationship between mean bear speed and daily energetic costs and compared this to the same relationship modelled in Pagano & Williams (2019). Our modelled costs were closely related to those from field DLW measurements, although slightly underestimated (Figure S9).

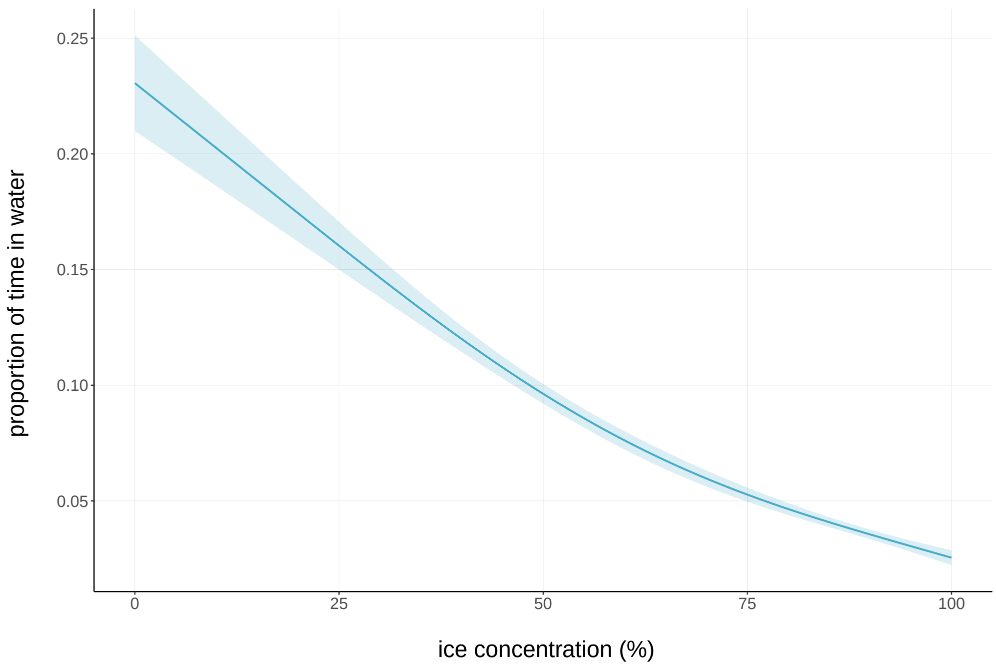

Figure S7: Estimated proportion of time spent in water relative to ice concentration, modelled with a generalized additive model (GAM). Data from Lone et al. (2018). Shaded area represents the standard error of the model fit.

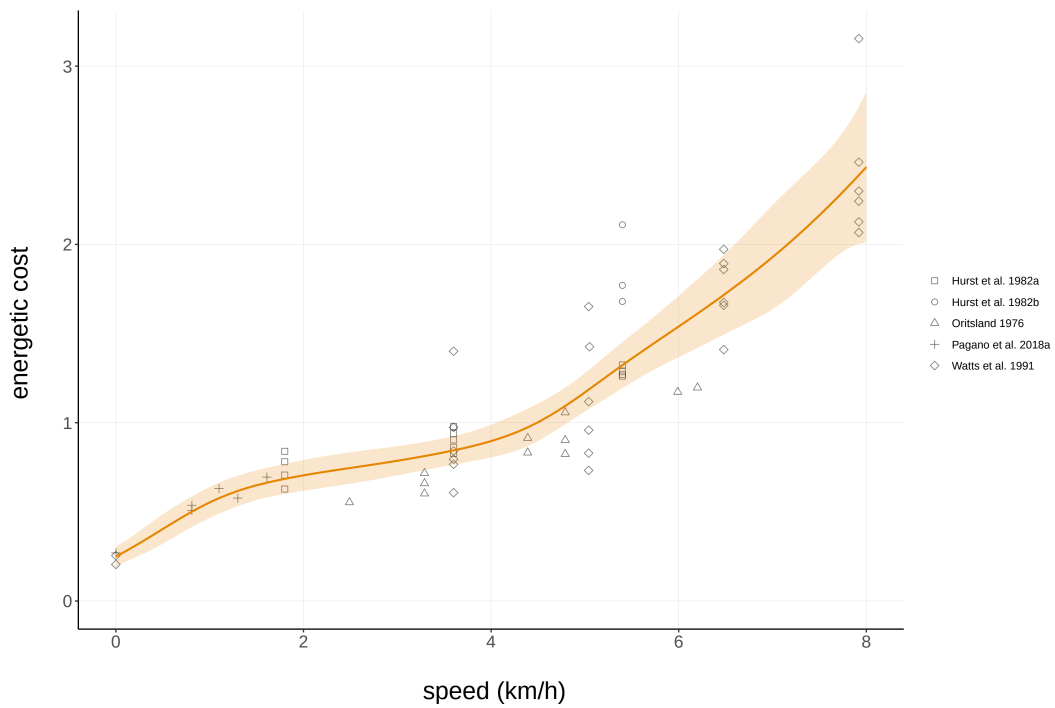

Figure S8: Relationship between polar bear walking speed (km/h) and energy expenditure (oxygen consumption; mL O2/g/hr) from six reference studies. Solid line is the predicted relationship from a monotonically constrained generalized additive model (gamma distribution, logit link function). Shaded area represents the standard error of the model fit.

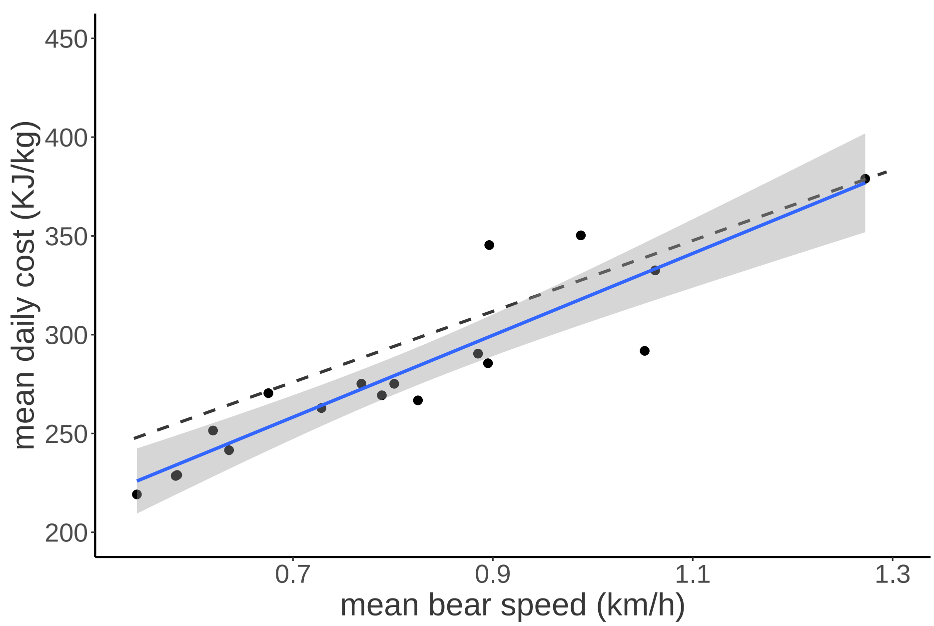

Figure S9: Relationship between mean daily movement speed (km/h) and daily energy expenditure (kJ/kg) for individuals with more than 25 days of locations with 6 locations (blue line) compared to the estimated relationship from doubly-labelled water (dashed line; Pagano & Williams, 2019).

E.3 Comparison to a simple random walk model

Model Fitting We compared the ESF to a null model, given by the simple random walk (SRW) described in Appendix A. Following Sections 2.2 and 4.1, we fit the SRW separately for each individual using *optim*. Costs were defined as l^2, where l is the ice drift-corrected bear step length (km). We compared models using AIC scores, where $AIC = 2 \times nllk + 2v$ where $nllk$ is the negative log-likelihood and v is the number of parameters in each model ($v_{ESF} = 2$ and $v_{SRW} = 1$).

Results SRW costs ranged from $0 - 509$ km^2 and ESF costs ranged from $3.3 - 161$ MJ. Since costs are in different units between models, β_2 estimates are on different scales (Figure S10). However, this does not affect the likelihood or AIC scores. The ESF was a better fitting model in almost all cases: $AIC_{ESF} < AIC_{SRW}$ for 20 out of 23 individuals (Figure S11). Based on guidelines from Burnham & Anderson (2002), there was little to no support for the competing model ($\Delta AIC > 6$) in all but one case (bearID = E; $AIC_{ESF} = -326.3$, $AIC_{SRW} = -325.5$). All but 3 cases had $\Delta AIC > 10$, which indicates essentially no support for the competing model (Burnham & Anderson, 2002).

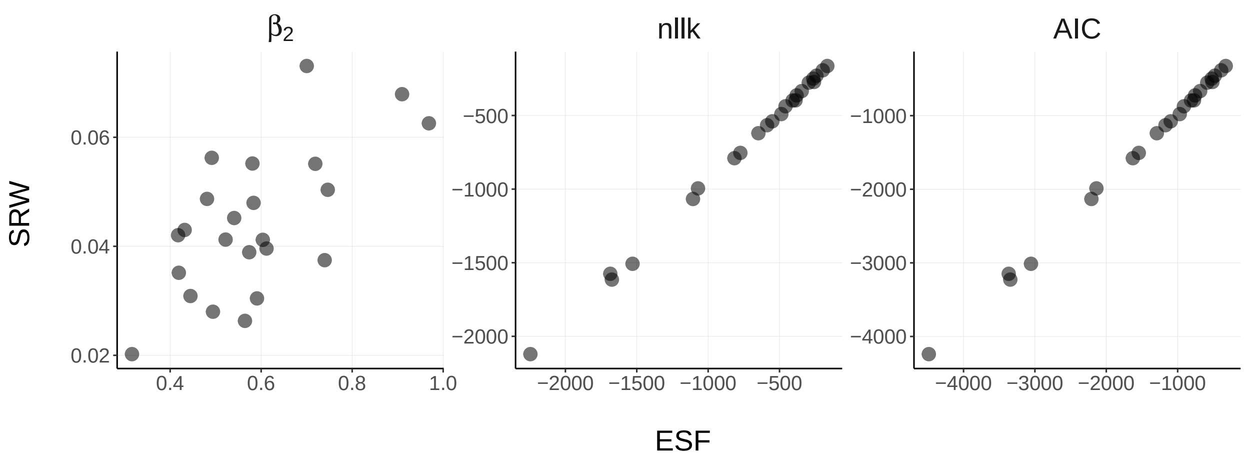

Figure S10: Parameter (β_2) estimates, negative log-likelihoods (nllk) and AIC scores for the ESF and SRW.

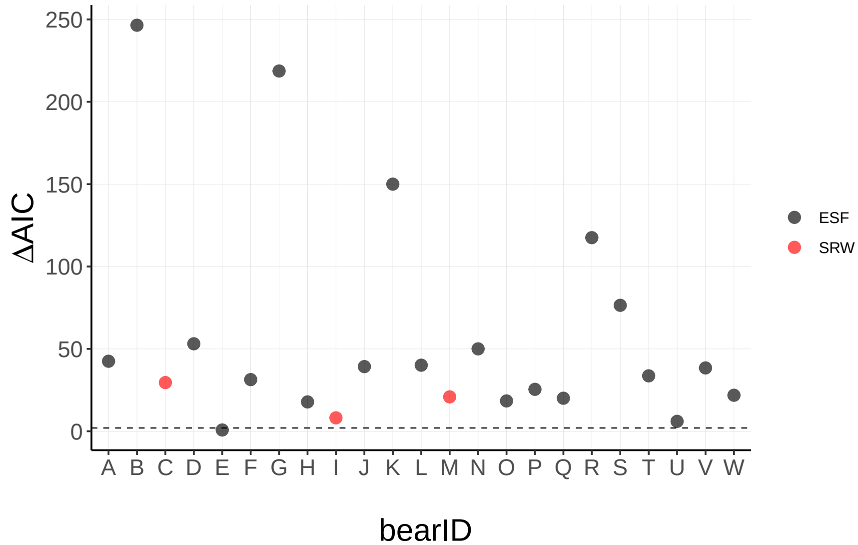

Figure S11: Comparison of AIC scores for bears with lower AIC_{ESF} (black dots; $\Delta AIC = AIC_{SRW} - AIC_{ESF}$) and lower AIC_{SRW} (red dots; $\Delta AIC = AIC_{ESF} - AIC_{SRW}$). The dashed line is at 2, which is a threshold to indicate considerable support for the model.

References

Avgar, T., Potts, J.R., Lewis, M.A. & Boyce, M.S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*, 619–630. https://doi.org/10.1111/2041-210X.12528

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, second edition.

Calvert, W. & Ramsay, M.A. (1998). Evaluation of age determination of polar bears by counts of cementum growth layer groups. *Ursus*, *10*, 449–453.

Carmack, E.C. & Macdonald, R.W. (2002). Oceanography of the Canadian shelf of the Beaufort Sea: A setting for marine life. *Arctic*, *55*, 29–45. https://doi.org/10.14430/arctic733

Codling, E.A., Plank, M.J. & Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal Society Interface*, *5*, 813–834. https://doi.org/10.1098/rsif.2008.0014

Durner, G.M., Douglas, D.C., Albeke, S.E., Whiteman, J.P., Amstrup, S.C., Richardson, E.S., Wilson, R.R. & Ben-David, M. (2017). Increased Arctic sea ice drift alters adult female polar bear movements and energetics. *Global Change Biology*, *23*, 3460–3473. https://doi.org/10.1111/gcb.13746

Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005). Wolves influence elk movements: behaviour shapes a trophic cascade in Yellowstone National Park. *Ecology*, *86*, 1320–1330. https://doi.org/10.1890/04-0953

Hutchings, J.K. & Rigor, I.G. (2012). Role of ice dynamics in anomalous ice conditions in the Beaufort Sea during. *Journal of Geophysical Research*, *117*, C00E04. https://doi.org/10.1029/2011JC007182

Lone, K., Kovacs, K.M., Lydersen, C., Fedak, M., Andersen, M., Lovell, P. & Aars, J. (2018). Aquatic behaviour of polar bears (*Ursus maritimus*) in an increasingly ice-free Arctic. *Scientific Reports*, *8*, 9677. https://doi.org/10.1038/s41598-018-27947-4

Michelot, T. (2019). *Stochastic models of animal movement and habitat selection*. Ph.D. thesis, University of Sheffield.

Northrup, J.M., Hooten, M.B., Anderson, C.R.J. & Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use-availability design. *Ecology*, *94*, 1456–1463. https://doi.org/10.1890/12-1688.1

Pagano, A.M. & Williams, T.M. (2019). Estimating the energy expenditure of free-ranging polar bears using tri-axial accelerometers: A validation with doubly labeled water. *Ecology and Evolution*, *9*, 4210–4219. https://doi.org/10.1002/ece3.5053

Petty, A.A., Hutchings, J.K., Richter-Menge, J.A. & Tschudi, M.A. (2016). Sea ice circulation around the Beaufort Gyre: the changing role of wind forcing and the sea ice state. *Journal of Geophysical Research: Oceans*, *121*, 3278–3296. https://doi.org/10.1002/2015JC010903

Pilfold, N.W., Derocher, A.E. & Richardson, E.S. (2014). Influence of intraspecific competition on the distribution of a wide-ranging, non-territorial carnivore. *Global Ecology and Biogeography*, *23*, 425–435. https://doi.org/10.1111/geb.12112

Stern, H.L. & Laidre, K.L. (2016). Sea-ice indicators of polar bear habitat. *Cryosphere*, *10*, 2027–2041. https://doi.org/10.5194/tc-10-2027-2016

Stirling, I., Spencer, C. & Andriashek, D.S. (1989). Immobilization of polar bears (*Ursus maritimus*) with Telazol in the Canadian Arctic. *Journal of Wildlife Diseases*, *25*, 159–168.