Trends, challenges, and opportunities in open biologging data

## Introduction

(499/500 words)

Biologging uses animal-borne sensors to remotely observe animals and their environments1. In the six decades since a biologist first attached a modified kitchen timer to a seal2, biologging has become an invaluable, multidisciplinary tool to address questions in ecology3, atmospheric science4,5, and oceanography6. As biologging data continue to grow in size and complexity (e.g., millions of new records daily in Movebank7 and novel sensors like infrasound recorders8), “big data” methods are a key direction for future developments9,10. Simultaneously, biologging data are increasingly used to inform conservation efforts across taxonomic groups11,12 and regions13. However, the envisioned impacts of aggregated biologging data on research and conservation are hindered by a lack of data standardization, discoverability, and accessibility14, which has become a growing matter of practical and ethical concern15. Currently, there is no systematic review of biologging studies, instruments used, species investigated, or open biologging datasets. Insights from such a review would support collaborative data sharing and curation efforts, and direct approaches to drive greater research and conservation impacts.

Open science practices, such as publishing data and code alongside manuscripts, improve research transparency and efficiency16,17. However, data “openness” is determined by multiple factors, and poor research data management and sharing practices are a major source of biodiversity data loss18. Fortunately, the FAIR principles19 ([Table 1](#tbl-fair)) provide a useful framework to facilitate open data, with FAIR data enabling groundbreaking advances in synthesis, biodiversity, and conservation science20–23 .

Numerous efforts within the biologging research community have sought to promote open data. For example, data standards14,24,25 and domain-specific data repositories (Movebank7, [seabirdtracking.org](https://www.seabirdtracking.org), Euromammal26) facilitate the adoption of the FAIR principles ([Table 1](#tbl-fair)). However, the overall state of open data in biologging research remains unknown, and there are indications that most tracking data remain inaccessible27,28.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: The FAIR principles19, as implemented in biologging cyberinfrastructure, support data sharing and improve research transparency and efficiency.

| FAIR principle | Definition | Application to biologging | Example |
| --- | --- | --- | --- |
| Findability | Data and metadata have a globally unique and persistent identifier (e.g., a digital object identifier, DOI) and are indexed in a searchable resource | Data repositories, like the Movebank Data Repository7, improve data discoverability and may assign DOIs to data (avoiding issues with broken hyperlinks, for example) | A tracking dataset deposited in the Movebank Data Repository is findable by its permanent DOI or by searching the repository |
| Accessibility | Data and metadata are retrievable by open and universal protocols, such as HTTP | Data repositories allow scientists to retrieve biologging data via a web browser or other open source tools | Publicly available data on Movebank may be downloaded via the website, API, or with the move2 R package29 |
| Interoperability | Data use formal and shared formats and vocabularies | Shared protocols like the Darwin Core standard30, Movebank data model31, and proposed bio-logging standardization framework14 reduce barriers to combining datasets and increase uptake within and across scientific disciplines | Location and environmental data from seal-borne biologgers harmonized to a standard netCDF format facilitated their reuse by oceanographers to study polar regions6 |
| Reusability | Data and metadata are richly described and reuse permissions are clearly defined | Data repositories and standards together capture essential context for biologging data and provide licensing options for data reuse | The Movebank Data Repository releases datasets under the CC0 license and the Movebank data model31 includes fields important for data reuse (e.g., whether the animal was relocated before release) |

 |

We will conduct a systematic review to quantify the diversity of instruments, species, and open data practices in biologging studies published between 2007-2023. 2007 marked the creation of Movebank, the largest biologging data platform. Our review will test the following three hypotheses:

**H1:** Open biologging data practices are increasing over time.

* Prediction: we expect open data increased slowly 2007-2016, then more rapidly 2016-2023 following the publication of the FAIR principles and many journals adopting open data policies32.

**H2:** Open biologging data practices vary by ecosystem.

* Prediction: we expect biologging data to be more open for studies in terrestrial than marine ecosystems, as many leading journals in marine science (e.g. *Marine Ecology Progress Series*, *Marine Mammal Science*) do not yet have open data requirements.

**H3:** Open biologging data practices are greater for spatial than aspatial (e.g., accelerometer) biologging data.

* Prediction: we expect spatial data to be more open than aspatial data, for which fewer repositories and data standards exist10.

Beyond hypothesis testing, our review will provide the first biologging bibliometric database, itself a valuable form of cyberinfrastrucure for future systematic reviews and meta-analyses. We will also provide a minimum information standard (e.g. MIAPE33, MIReAD34) as a resource for authors, journal editors, and funding bodies to facilitate FAIR sharing of biologging data.

## Methods

We will systematically review biologging studies of wild animals in the fields of ecology, evolutionary biology, and conservation science published between 2007-2023. We define biologging as animal-borne instruments with sensors and memory for data storage. Studies employing biotelemetry devices (defined as animal-borne instruments that emit a signal for an external receiver)1,35 will be excluded, because their analyses require data describing the configuration of the receiver network. We make an exception for satellite biotelemetry using the Argos network, because the Argos system resolves transmissions to timestamped locations, and thus provides researchers with data more similar to biologgers.

### Literature review

On 2023-08-21, we queried Web of Science Core Collection for papers published between 2007-2023 related to biologging with the following query:

animal AND (biologg\* OR bio-logg\* OR biotelemetry OR electronic tag\* OR satellite track\* OR GPS telemetry OR satellite telemetry OR satellite transmit\* OR GPS collar\* OR depth recorder\* OR accelerometer OR archival tag\*)

This yielded 6654 papers. We limited results to relevant Web of Science categories with 100+ papers, leaving 4799 papers. The categories we considered relevant were Ecology, Zoology, Marine Freshwater Biology, Biodiversity Conservation, Multidisciplinary Sciences, Environmental Sciences, Oceanography, Biology, Behavioral Sciences, Ornithology, Evolutionary Biology, and Fisheries. We excluded the following categories with 100+ papers because they typically involve domesticated or lab animals: Agriculture Dairy Animal Science, Veterinary Sciences, Engineering Electrical Electronic, Physiology, Food Science Technology, Engineering Biomedical, Agriculture Multidisciplinary, Computer Science Interdisciplinary Applications, Instruments Instrumentation.

### Initial classification

Every paper will be separately reviewed by two individuals to assess its relevance to this review, record the presence or absence of a data availability statement, and categorize the study by sensor type, taxonomy, and ecosystem. In cases where the reviewers disagree, a third reviewer will assess the study to reach a consensus. Papers will be considered relevant if they satisfy the following two characteristics:

1. They involved the collection of previously unpublished biologging data.
2. The biologgers were deployed on non-domesticated animals in unrestrained habitat. Studies involving domesticated animals (e.g., cattle), animals in managed care (e.g., a zoo or aquarium), or biomedical laboratories will be excluded.

We will consider a paper to have a data availability statement (DAS) if it states how to access the data or explicitly states why the data are not available. Any of the following examples would be considered a DAS: “Data were deposited in a repository” with an accompanying DOI, “Data are available at a website” with an accompanying URL, “Data are available upon request”, or “Data withheld for the safety of the tracked animals”.

We will assign sensor types to one or more categories following Williams et al.36: location (e.g., GPS, depth), intrinsic (e.g., accelerometer, internal temperature), or environment (e.g., ambient temperature, camera). Biologgers can combine sensors from multiple categories (e.g., a GPS-accelerometer collar). We will consider all sensors that were reported in the methods.

We will resolve taxonomy to the species level using the Integrated Taxonomic Information System (www.itis.gov). Each study’s ecosystem will be classified as one or more of aquatic (i.e., freshwater), marine (i.e., saltwater), and terrestrial according to the IUCN Red List (https://www.iucnredlist.org).

It was necessary for us to begin the initial classification to properly design the protocol for the open data assessment (next section). We have completed the initial classification for ~25% of studies in our sample. No data have been collected about studies’ open data practices, so our hypotheses are not informed by preliminary data.

### Assessing open data

Following our initial classification, we will randomly sample 600 studies to assess the openness of biologging data. We will stratify our sample by year, ecosystem, and sensor type. We will attempt to sample 50 studies in each combination of the following categories:

* Time period (early 2007-2015, middle 2016-2019, recent 2020-2023)
* Ecosystem (marine, terrestrial)
* Sensor type (spatial, aspatial)

Freshwater aquatic biologging studies will be excluded from the open data assessment. Our preliminary results from the initial classification indicate freshwater aquatic biologging studies are rare, as most tracking studies in this ecosystem use acoustic telemetry.

Studies may include both spatial (i.e., location) and aspatial (i.e., intrinsic or environmental) sensor data. Therefore, at the time of sampling, we will assign each study to either the aspatial or spatial category and limit our assessment to the corresponding sensor data. Our initial classification preliminary results indicate spatial sensors are deployed in the vast majority of studies (93%), whereas aspatial sensors are deployed less frequently (35% of studies).

For each sampled paper, we will attempt to retrieve the biologging dataset underlying the study. If data are available upon request, we will email the corresponding author to request the data and, if we do not receive a response within two weeks, we will send one follow-up email. Similarly, if we receive a summary of the biologging data, we will send one follow-up email requesting the underlying biologging data. We will record whether each dataset satisfies the following properties:

1. Findable ( $f$ ). Do the data have a permanent identifier, such as a DOI? I.e., FAIR principle F119.
2. Accessible ( $a$ ). Are the data retrievable via a standard communications protocol, such as HTTP? I.e., FAIR principle A119.
3. Reusable ( $r$ ). Do the data contain biologging data (raw or processed), as opposed to descriptive statistics or other summaries? I.e., FAIR principle R119.

If we cannot retrieve the dataset for a study, we will consider the study to not satisfy any of the three properties listed above.

To clarify our methods for spatial versus aspatial data, consider a hypothetical GPS-accelerometer biologging study sampled in the aspatial category. The authors deposited the GPS data in the Movebank Data Repository (which issues DOIs) and the accelerometer data are available upon request. We contact the author and they provide us with a summary table of the accelerometer data, where each row is a deployment and columns are characteristics of the deployment, such as duration. They do not respond to our follow-up email requesting the underlying biologging data. Then, because this study was sampled in the aspatial category, it would be scored as $f=0,a=0,r=0$ because ($f$) the accelerometer data do not have a DOI, ($a$) retrieving the accelerometer data required contacting the author, and ($r$) only summaries of the accelerometer data were made available. Had this same study been sampled in the spatial category, we would have scored it as $f=1,a=1,r=1$ because ($f$) the spatial data have a DOI, ($a$) the data were retrievable via HTTP, and ($r$) the GPS sensor data were available.

### Analysis

Our independent variables will be year of publication ($t$), ecosystem ($e$), and sensor type ($s$) [see “Initial classification”]. Our dependent variables will be findability ($f$), accessibility ($a$), and reusability ($r$) [see “Assessing open data”]. We will fit three Bayesian generalized linear models of the form:

$$\begin{matrix}DV&∼Bernoulli\left(p\right)\\logit\left(p\_{i}\right)&=α+β\_{e}×e\_{i}+β\_{s}×s\_{i}+β\_{t}×t\_{i}\\\end{matrix}$$

Where $DV$ is the dependent variable ($f$, $a$, or $r$). $e$ and $s$ encode ecosystem (0 for marine, 1 for terrestrial) and sensor type (0 for aspatial, 1 for spatial), respectively. $t$ is the publication year minus the beginning of the study period (i.e., $t=0$ in 2007, …, $t=16$ in 2023).

We use the following priors:

$$\begin{matrix}α&∼Normal\left(0,4\right)\\β\_{e}&∼Normal\left(0,2\right)\\β\_{s}&∼Normal\left(0,2\right)\\β\_{t}&∼Normal\left(0,0.5\right)\end{matrix}$$

We conducted prior predictive checks of our priors. Flatter priors, such as $Normal\left(0,10\right)$ for all coefficients, lead to unrealistic values for $p$ (i.e., the probability density of $p$ is concentrated almost entirely near 0 or 1).

We will use Stan37 and brms38,39 for model fitting.

We invite feedback from reviewers about statistical design. We considered more complex models, e.g., using splines for $β\_{t}$, treating year as an ordinal variable40, or allowing interactions between $e$, $s$, and $t$. However, we concluded that the model presented here would answer our research questions with the greatest clarity. If the reviewers think more complex models are essential, we will modify our design accordingly.

### Hypothesis testing

We will test hypotheses by testing if the 95% credible interval for the posterior probability distribution of the relevant model coefficients (in all three models) is greater than 0. We will quantify effect sizes using the 95% credible interval for the posterior probability distribution of contrasts in $p\left(f\right)$ , $p\left(a\right)$, and $p\left(r\right)$ across levels of the relevant category to the hypothesis. We calculate contrasts by sampling the posterior distribution of the probability of a dependent variable (e.g., $p\left(f\right)$) while varying the category of interest. For example, the contrast of $p\left(f\right)$ for hypothesis **H2** (open data vary by ecosystem, $e$), is $p\left(f|e=terrestrial\right)−p\left(f|e=marine\right)$, marginal across $s$ and $t$. See section “Analysis of simulated data” to see our implementation with simulated data.

#### H1: Open biologging data practices are increasing over time

We will test hypothesis **H1** using coefficient $β\_{t}$. We will quantify effect size by contrasting the dependent variable posterior probabilities for three eras: 2007-2015, 2016-2019, and 2020-2023 (see [Figure 2](#fig-h1) in section “Analysis of simulated data” below).

#### H2: Open biologging data practices vary by ecosystem

We will test hypothesis **H2** using coefficient $β\_{e}$. To quantify effect size, we will contrast posterior probabilities for marine and terrestrial ecosystems (see [Figure 3](#fig-h2)).

#### H3: Open biologging data practices vary by sensor class

We will test hypothesis **H3** using coefficient $β\_{s}$. To quantify effect size, we will contrast posterior probabilities for aspatial and spatial sensor data (see [Figure 4](#fig-h3)).

### Analysis of simulated data

We conducted a simulation to demonstrate how we will conduct our analysis. We simulated values for $t$, $e$, $s$ (time, ecosystem, and sensor class; independent variables), and $f$ (findability; dependent variable) according to our predictions, to demonstrate how we will perform our analysis. The full analysis will also include dependent variables $a$ (accessibility) and $r$ (reusability).

We set up our analysis by loading necessary packages and setting the seed.

library(brms)
library(cowplot)
library(tidybayes)
library(tidyverse)
set.seed(1234)

The following code simulates our independent and dependent variables. We simulate independent variables according to our stratified sampling plan (biolog\_grid). Then we choose values for our parameters $β\_{e}$ (beta\_E), $β\_{s}$ (beta\_S), $β\_{t}$ (beta\_T), and $α$ (alpha), which we use to simulate our dependent variable, $f$ (F). The probability of $f$, $p\left(f\right)$, increases over time, and $p\left(f\right)$ is greater for spatial sensors and terrestrial ecosystems than aspatial sensors and marine ecosystems ([Figure 1](#fig-sim-data)).

# Grid of independent variables
biolog\_grid <- expand\_grid(
 era = c("early", "middle", "recent"),
 E = c("marine", "terrestrial"),
 S = c("aspatial", "spatial")
) %>%
 slice(rep(seq(nrow(.)), each = 50)) %>%
 mutate(T = case\_when(
 era == "early" ~ sample(2007:2015, nrow(.), replace = TRUE),
 era == "middle" ~ sample(2016:2019, nrow(.), replace = TRUE),
 era == "recent" ~ sample(2020:2023, nrow(.), replace = TRUE)
 ))

# Parameters for simulation
# betas (E, S, T)
alpha <- -4
beta\_E <- c(marine = 0, terrestrial = 0.5)
beta\_S <- c(aspatial = 0, spatial = 1.5)
beta\_T = 0.2

# Run simulation
inv\_logit <- \(x) exp(x) / (1 + exp(x))
biolog <- biolog\_grid %>%
 mutate(T2007 = T - 2007,
 logit\_p = alpha + beta\_E[E] + beta\_S[S] + beta\_T \* T2007,
 p = inv\_logit(logit\_p),
 F = rbinom(nrow(.), size = 1, prob = p))

# Visualize data
e\_palette <- c(marine = "navy", terrestrial = "goldenrod")

biolog\_text <- filter(biolog, T == 2023) %>%
 distinct(E, S, p) %>%
 mutate(label = str\_to\_title(paste(E, S, sep = "\n")))

ggplot(biolog, aes(T, p, color = E)) +
 geom\_jitter(aes(y = F, shape = S),
 width = 0.25, height = 0.05,
 alpha = 0.5) +
 geom\_line(aes(linetype = S)) +
 geom\_text(aes(label = label), biolog\_text,
 x = 2023.1, hjust = 0) +
 scale\_color\_manual(values = e\_palette) +
 scale\_x\_continuous(limits = c(NA, 2025),
 breaks = seq(2010, 2020, by = 5)) +
 labs(y = expression(p(f))) +
 expand\_limits(x = 2025) +
 theme\_classic() +
 theme(legend.position = "none",
 axis.title.x = element\_blank())

|  |
| --- |
| Figure 1: In our simulation, probability of findability ($p\left(f\right)$) increases over time. Lines represent $p\left(f\right)$, points are simulated studies. $p\left(f\right)$ is greater for terrestrial (gold) than marine (blue) data, and greater for spatial (dashed lines, triangles) than aspatial (solid lines, circles) data. |

Here we specify our model, including priors, and fit it to our simulated data.

# Fit model
biolog\_prior <- c(
 set\_prior(prior = "normal(0, 4)", class = "Intercept"),
 set\_prior(prior = "normal(0, 2)", coef = "Eterrestrial"),
 set\_prior(prior = "normal(0, 2)", coef = "Sspatial"),
 set\_prior(prior = "normal(0, 0.5)", coef = "T2007")
 )
biolog\_mod <- brm(F ~ E + S + T2007,
 data = biolog,
 family = bernoulli(link = "logit"),
 prior = biolog\_prior,
 chains = 4,
 iter = 50000,
 seed = 6789,
 refresh = 0)
biolog\_draws <- as\_draws\_df(biolog\_mod)

We test our hypotheses using credible intervals of the parameter estimates and contrasts of the posterior probabilities. In our simulation, the 95% credible intervals for coefficients $β\_{t}$, $β\_{e}$, and $β\_{s}$ were greater than 0 ([Figure 2](#fig-h1)A, [Figure 3](#fig-h2)A, [Figure 4](#fig-h3)A). The contrasts quantify the effect sizes in the simulation. [Figure 2](#fig-h1)B, for example, shows the posterior distribution of the contrast of $p\left(f\right)$ from the early to middle periods (above) and from the middle to recent periods (below). The 95% credible interval from early to middle periods was (0.073 - 0.300), indicating the probability of findability increased 7.3% - 30%.

h1\_dens <- density(biolog\_draws$b\_T2007)
h1\_dens\_dat <- tibble(
 b\_T2007 = h1\_dens$x,
 dens = h1\_dens$y,
 ci95 = between(b\_T2007,
 quantile(biolog\_draws$b\_T2007, 0.025),
 quantile(biolog\_draws$b\_T2007, 0.975))
)

h1\_contrast <- map2(
 list("early", "middle", "recent"),
 list(2007:2015, 2016:2019, 2020:2023),
 function(era, years) {
 dat <- expand\_grid(
 E = c("marine", "terrestrial"),
 S = c("aspatial", "spatial"),
 T = years
 ) %>%
 mutate(T2007 = T - 2007)
 linpred\_draws(biolog\_mod,
 transform = TRUE,
 newdata = dat) %>%
 group\_by(E, S, .draw) %>%
 summarize(p\_F = mean(.linpred),
 .groups = "drop") %>%
 mutate(era = era)
 }
) %>%
 list\_rbind() %>%
 pivot\_wider(names\_from = "era", values\_from = "p\_F") %>%
 mutate(`Early to Middle` = middle - early,
 `Middle to Recent` = recent - middle) %>%
 pivot\_longer(c(`Early to Middle`, `Middle to Recent`),
 names\_to = "Eras",
 values\_to = "Contrast")

h1\_contrast\_dens <- h1\_contrast %>%
 group\_by(Eras) %>%
 group\_map(\(rows, keys) {
 dens <- density(rows$Contrast)
 lwr <- quantile(rows$Contrast, 0.025)
 upr <- quantile(rows$Contrast, 0.975)
 cross\_join(keys,
 tibble(Contrast = dens$x,
 Density = dens$y,
 ci\_95 = between(Contrast, lwr, upr)))
 }) %>%
 list\_rbind()

beta\_t\_ci\_lwr <- quantile(biolog\_draws$b\_T2007, 0.025)
beta\_t\_ci\_upr <- quantile(biolog\_draws$b\_T2007, 0.975)
early\_middle\_lwr <- with(
 h1\_contrast, quantile(Contrast[Eras == "Early to Middle"], 0.025)
)
early\_middle\_upr <- with(
 h1\_contrast, quantile(Contrast[Eras == "Early to Middle"], 0.975)
)
middle\_recent\_lwr <- with(
 h1\_contrast, quantile(Contrast[Eras == "Middle to Recent"], 0.025)
)
middle\_recent\_upr <- with(
 h1\_contrast, quantile(Contrast[Eras == "Middle to Recent"], 0.975)
)

p\_h1a <- ggplot(h1\_dens\_dat, aes(b\_T2007, dens)) +
 geom\_line() +
 geom\_area(data = filter(h1\_dens\_dat, ci95), fill = "grey50") +
 expand\_limits(x = 0) +
 labs(x = expression(beta[t]), y = "Density") +
 theme\_classic()

p\_h1b <- ggplot(h1\_contrast\_dens, aes(Contrast, Density)) +
 geom\_line() +
 geom\_area(data = filter(h1\_contrast\_dens, ci\_95), fill = "grey50") +
 facet\_grid(rows = vars(Eras), scales = "free\_y") +
 labs(x = "Contrast in p(f)") +
 expand\_limits(x = 0) +
 theme\_classic()

plot\_grid(p\_h1a, p\_h1b, nrow = 1, labels = "AUTO")

|  |
| --- |
| Figure 2: Evidence supporting H1 in our simulated data. A: The 95% credible interval for $β\_{t}$ for the model fit to our simulated data was greater than 0 (0.130 - 0.224). B: The 95% credible interval for the contrasts of $p\left(f\right)$ between the early (2007-2015) and middle (2016-2019) periods (0.073 - 0.300), and between the middle and recent (2020-2023) periods (0.085 - 0.205), were also greater than 0. |

h2\_dens <- density(biolog\_draws$b\_Eterrestrial)
h2\_dens\_dat <- tibble(
 b\_Eterrestrial = h2\_dens$x,
 dens = h2\_dens$y,
 ci95 = between(b\_Eterrestrial,
 quantile(biolog\_draws$b\_Eterrestrial, 0.025),
 quantile(biolog\_draws$b\_Eterrestrial, 0.975))
)

h2\_contrast <- map(
 c("marine", "terrestrial"),
 function(e) {
 dat <- expand\_grid(
 E = e,
 S = c("aspatial", "spatial"),
 T = 2007:2023
 ) %>%
 mutate(T2007 = T - 2007)
 linpred\_draws(biolog\_mod,
 transform = TRUE,
 newdata = dat)
 }
) %>%
 list\_rbind() %>%
 pivot\_wider(names\_from = "E", values\_from = ".linpred") %>%
 mutate(Contrast = terrestrial - marine)

dens <- density(h2\_contrast$Contrast)
h2\_contrast\_dens <- tibble(
 Contrast = dens$x,
 Density = dens$y,
 ci\_95 = between(Contrast,
 quantile(h2\_contrast$Contrast, 0.025),
 quantile(h2\_contrast$Contrast, 0.975))
)

beta\_e\_ci\_lwr <- quantile(biolog\_draws$b\_Eterrestrial, 0.025)
beta\_e\_ci\_upr <- quantile(biolog\_draws$b\_Eterrestrial, 0.975)
e\_contrast\_lwr <- quantile(h2\_contrast$Contrast, 0.025)
e\_contrast\_upr <- quantile(h2\_contrast$Contrast, 0.975)

p\_h2a <- ggplot(h2\_dens\_dat, aes(b\_Eterrestrial, dens)) +
 geom\_line() +
 geom\_area(data = filter(h2\_dens\_dat, ci95), fill = "grey50") +
 expand\_limits(x = 0) +
 labs(x = expression(beta[e]), y = "Density") +
 theme\_classic()

p\_h2b <- ggplot(h2\_contrast\_dens, aes(Contrast, Density)) +
 geom\_line() +
 geom\_area(data = filter(h2\_contrast\_dens, ci\_95), fill = "grey50") +
 expand\_limits(x = 0) +
 labs(x = "Contrast of p(f) [terrestrial - marine]") +
 theme\_classic()

plot\_grid(p\_h2a, p\_h2b, nrow = 1, labels = "AUTO")

|  |
| --- |
| Figure 3: Evidence supporting H2 in our simulated data. A: The 95% credible interval for $β\_{e}$ for the model fit to our simulated data was greater than 0 (0.416 - 1.176). B: The 95% credible interval for the contrast of $p\left(f\right)$ between terrestrial and marine ecosystems was also greater than 0 (0.026 - 0.251). |

h3\_dens <- density(biolog\_draws$b\_Sspatial)
h3\_dens\_dat <- tibble(
 b\_Sspatial = h3\_dens$x,
 dens = h3\_dens$y,
 ci95 = between(b\_Sspatial,
 quantile(biolog\_draws$b\_Sspatial, 0.025),
 quantile(biolog\_draws$b\_Sspatial, 0.975))
)

h3\_contrast <- map(
 c("aspatial", "spatial"),
 function(s) {
 dat <- expand\_grid(
 E = c("marine", "terrestrial"),
 S = s,
 T = 2007:2023
 ) %>%
 mutate(T2007 = T - 2007)
 linpred\_draws(biolog\_mod,
 transform = TRUE,
 newdata = dat)
 }
) %>%
 list\_rbind() %>%
 pivot\_wider(names\_from = "S", values\_from = ".linpred") %>%
 mutate(Contrast = spatial - aspatial)

dens <- density(h3\_contrast$Contrast)
h3\_contrast\_dens <- tibble(
 Contrast = dens$x,
 Density = dens$y,
 ci\_95 = between(Contrast,
 quantile(h3\_contrast$Contrast, 0.025),
 quantile(h3\_contrast$Contrast, 0.975))
)

beta\_s\_ci\_lwr <- quantile(biolog\_draws$b\_Sspatial, 0.025)
beta\_s\_ci\_upr <- quantile(biolog\_draws$b\_Sspatial, 0.975)
s\_contrast\_lwr <- quantile(h3\_contrast$Contrast, 0.025)
s\_contrast\_upr <- quantile(h3\_contrast$Contrast, 0.975)

p\_h3a <- ggplot(h3\_dens\_dat, aes(b\_Sspatial, dens)) +
 geom\_line() +
 geom\_area(data = filter(h3\_dens\_dat, ci95), fill = "grey50") +
 expand\_limits(x = 0) +
 labs(x = expression(beta[s]), y = "Density") +
 theme\_classic()

p\_h3b <- ggplot(h3\_contrast\_dens, aes(Contrast, Density)) +
 geom\_line() +
 geom\_area(data = filter(h3\_contrast\_dens, ci\_95), fill = "grey50") +
 expand\_limits(x = 0) +
 labs(x = "Contrast of p(f) [spatial - aspatial]") +
 theme\_classic()

plot\_grid(p\_h3a, p\_h3b, nrow = 1, labels = "AUTO")

|  |
| --- |
| Figure 4: Evidence supporting H3 in our simulated data. A: The 95% credible interval for $β\_{s}$ for the model fit to our simulated data was greater than 0 (0.867 - 1.638). B: The 95% credible interval for the contrast of $p\left(f\right)$ between aspatial and spatial ecosystems was also greater than 0 (0.056 - 0.354). |

## Planned tables and figures

Table 1: Description of the FAIR principles and their application to biologging data (see [Table 1](#tbl-fair) above).

Table 2: Results of our hypothesis tests for the three dependent variables. 95% credible intervals for posterior probability distributions of our model coefficients and contrasts. This table summarizes [Figure 2](#fig-h1), [Figure 3](#fig-h2), and [Figure 4](#fig-h3) from the section “Analysis of simulated data”, as well as results for the other two dependent variables.

Figure 1: Results of the initial classification. **A** Sankey diagram showing the frequencies of instrumented species by number of studies, with aggregations at higher taxa. **B** Frequencies of location, intrinsic, and environmental sensors used in studies. **C** Prevalence of data availability statements over time. **D** Cumulative biologging studies published over time in the ten journals with the most published biologging studies.

Figure 2: Results of the open data assessment. Mean and standard error of dependent variables ($f$ findability, $a$ accessibility, $b$ reusability) aggregated over: **A** sensor type (aspatial or spatial), **B** ecosystem (marine or terrestrial), and **C** time (year of publication).

Supplemental figures 1-3: Posterior probability distributions of model coefficients and contrasts. Each supplemental figure will show the full posterior probability distributions for three hypotheses and one dependent variable. I.e. supplemental figure 1 would combine [Figure 2](#fig-h1), [Figure 3](#fig-h2), and [Figure 4](#fig-h3) from the section “Analysis of simulated data” above. Supplemental figures 2 and 3 would be similar, showing the results for the other dependent variables accessibility ( $a$ ) and reproducibility ( $r$ ).

1. Rutz, C. & Hays, G. C. [New frontiers in biologging science](https://doi.org/10.1098/rsbl.2009.0089). *Biology Letters* **5**, 289–292 (2009).

2. Kooyman, G. L. [Maximum Diving Capacities of the Weddell Seal,*Leptonychotes weddelli*](https://doi.org/10.1126/science.151.3717.1553). *Science* **151**, 1553–1554 (1966).

3. Wilmers, C. C. *et al.* [The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology](https://doi.org/10.1890/14-1401.1). *Ecology* **96**, 1741–1753 (2015).

4. Weinzierl, R. *et al.* [Wind estimation based on thermal soaring of birds](https://doi.org/10.1002/ece3.2585). *Ecology and Evolution* **6**, 8706–8718 (2016).

5. Yonehara, Y. *et al.* [Flight paths of seabirds soaring over the ocean surface enable measurement of fine-scale wind speed and direction](https://doi.org/10.1073/pnas.1523853113). *Proceedings of the National Academy of Sciences* **113**, 9039–9044 (2016).

6. Treasure, A. *et al.* [Marine mammals exploring the oceans pole to pole: A review of the MEOP consortium](https://doi.org/10.5670/oceanog.2017.234). *Oceanography* **30**, 132–138 (2017).

7. Kays, R. *et al.* [The Movebank system for studying global animal movement and demography](https://doi.org/10.1111/2041-210x.13767). *Methods in Ecology and Evolution* **13**, 419–431 (2021).

8. Ouden, O. F. C. den, Smets, P. S. M., Assink, J. D. & Evers, L. G. [A Bird’s-Eye View on Ambient Infrasonic Soundscapes](https://doi.org/10.1029/2021gl094555). *Geophysical Research Letters* **48**, (2021).

9. Nathan, R. *et al.* [Big-data approaches lead to an increased understanding of the ecology of animal movement](https://doi.org/10.1126/science.abg1780). *Science* **375**, (2022).

10. Grémillet, D., Chevallier, D. & Guinet, C. [Big data approaches to the spatial ecology and conservation of marine megafauna](https://doi.org/10.1093/icesjms/fsac059). *ICES Journal of Marine Science* **79**, 975–986 (2022).

11. Beal, M. *et al.* [Global political responsibility for the conservation of albatrosses and large petrels](https://doi.org/10.1126/sciadv.abd7225). *Science Advances* **7**, (2021).

12. Queiroz, N. *et al.* [Global spatial risk assessment of sharks under the footprint of fisheries](https://doi.org/10.1038/s41586-019-1444-4). *Nature* **572**, 461–466 (2019).

13. Hindell, M. A. *et al.* [Tracking of marine predators to protect Southern Ocean ecosystems](https://doi.org/10.1038/s41586-020-2126-y). *Nature* **580**, 87–92 (2020).

14. Sequeira, A. M. M. *et al.* [A standardisation framework for bio-logging data to advance ecological research and conservation](https://doi.org/10.1111/2041-210x.13593). *Methods in Ecology and Evolution* **12**, 996–1007 (2021).

15. Rutz, C. [Register animal-tracking tags to boost conservation](https://doi.org/10.1038/d41586-022-02821-6). *Nature* **609**, 221–221 (2022).

16. Lowndes, J. S. S. *et al.* [Our path to better science in less time using open data science tools](https://doi.org/10.1038/s41559-017-0160). *Nature Ecology & Evolution* **1**, 1–7 (2017).

17. Hampton, S. E. *et al.* [Skills and Knowledge for Data-Intensive Environmental Research](https://doi.org/10.1093/biosci/bix025). *BioScience* **67**, 546–557 (2017).

18. Peterson, A. T., Asase, A., Canhos, D., Souza, S. de & Wieczorek, J. [Data leakage and loss in biodiversity informatics](https://doi.org/10.3897/bdj.6.e26826). *Biodiversity Data Journal* **6**, (2018).

19. Wilkinson, M. D. *et al.* [The FAIR Guiding Principles for scientific data management and stewardship](https://doi.org/10.1038/sdata.2016.18). *Scientific Data* **3**, 160018 (2016).

20. Halpern, B. S. *et al.* Ecological Synthesis and Its Role in Advancing Knowledge. *BioScience* (2020) doi:[10.1093/biosci/biaa105](https://doi.org/10.1093/biosci/biaa105).

21. Fredston, A. L. & Lowndes, J. S. S. [Welcoming More Participation in Open Data Science for the Oceans](https://doi.org/10.1146/annurev-marine-041723-094741). *Annual Review of Marine Science* **16**, (2023).

22. Roche, D. G. *et al.* [Closing the knowledge-action gap in conservation with open science](https://doi.org/10.1111/cobi.13835). *Conservation Biology* **36**, (2021).

23. Buxton, R. T. *et al.* [Avoiding wasted research resources in conservation science](https://doi.org/10.1111/csp2.329). *Conservation Science and Practice* **3**, (2021).

24. Dettki, H., Arlt, D., Bäckman, J. & Blanchet, M. [Key-value pairs and NoSQL databases: A novel concept to manage biologging data in data repositories](https://doi.org/10.3897/biss.7.111438). *Biodiversity Information Science and Standards* **7**, (2023).

25. Kranstauber, B. *et al.* [The Movebank data model for animal tracking](https://doi.org/10.1016/j.envsoft.2010.12.005). *Environmental Modelling & Software* **26**, 834–835 (2011).

26. Urbano, F. & Cagnacci, F. [Data management and sharing for collaborative science: Lessons learnt from the euromammals initiative](https://doi.org/10.3389/fevo.2021.727023). *Frontiers in Ecology and Evolution* **9**, (2021).

27. Campbell, H. A. *et al.* [Finding our way: On the sharing and reuse of animal telemetry data in Australasia](https://doi.org/10.1016/j.scitotenv.2015.01.089). *Science of The Total Environment* **534**, 79–84 (2015).

28. Scarpignato, A. L. *et al.* [Shortfalls in tracking data available to inform North American migratory bird conservation](https://doi.org/10.1016/j.biocon.2023.110224). *Biological Conservation* **286**, 110224 (2023).

29. Kranstauber, B., Safi, K. & Scharf, A. K. [*move2: Processing and analysing animal trajectories*](https://CRAN.R-project.org/package%3Dmove2). (2023).

30. Wieczorek, J. *et al.* [Darwin Core: an evolving community-developed biodiversity data standard](https://doi.org/10.1371/journal.pone.0029715). *PLoS ONE* **7**, e29715 (2012).

31. Animal Behavior, M. P. I. of. [Movebank Attribute Dictionary](http://dx.doi.org/10.1038/s41597-019-0042-5).

32. Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. [Public Data Archiving in Ecology and Evolution: How Well Are We Doing?](https://doi.org/10.1371/journal.pbio.1002295) *PLOS Biology* **13**, e1002295 (2015).

33. Taylor, C. F. *et al.* [The minimum information about a proteomics experiment (MIAPE)](https://doi.org/10.1038/nbt1329). *Nature Biotechnology* **25**, 887–893 (2007).

34. Rund, S. S. C. *et al.* [MIReAD, a minimum information standard for reporting arthropod abundance data](https://doi.org/10.1038/s41597-019-0042-5). *Scientific Data* **6**, (2019).

35. Cooke, S. J. *et al.* [Biotelemetry: a mechanistic approach to ecology](https://doi.org/10.1016/j.tree.2004.04.003). *Trends in Ecology & Evolution* **19**, 334–343 (2004).

36. Williams, H. J. *et al.* [Optimizing the use of biologgers for movement ecology research](https://doi.org/10.1111/1365-2656.13094). *Journal of Animal Ecology* **89**, 186–206 (2019).

37. Stan Development Team. [Stan modeling language users guide and reference manual, version 2.34](https://mc-stan.org). (2024).

38. Bürkner, P.-C. [Advanced Bayesian Multilevel Modeling with the R Package brms](https://doi.org/10.32614/rj-2018-017). *The R Journal* **10**, 395 (2018).

39. Bürkner, P.-C. [Bayesian Item Response Modeling in *R* with **brms** and *Stan*](https://doi.org/10.18637/jss.v100.i05). *Journal of Statistical Software* **100**, (2021).

40. Bürkner, P.-C. & Charpentier, E. [Modelling monotonic effects of ordinal predictors in Bayesian regression models](https://doi.org/10.1111/bmsp.12195). *British Journal of Mathematical and Statistical Psychology* **73**, 420–451 (2020).