# *Navigating Challenges in Spatio-Temporal Modelling of Antarctic Krill Abundance: Addressing Zero-Inflated Data and Misaligned Covariates*

André V. Ribeiro Amaral †
a.ribeiro-amaral@imperial.ac.uk

Imperial College London

July 24, 2024

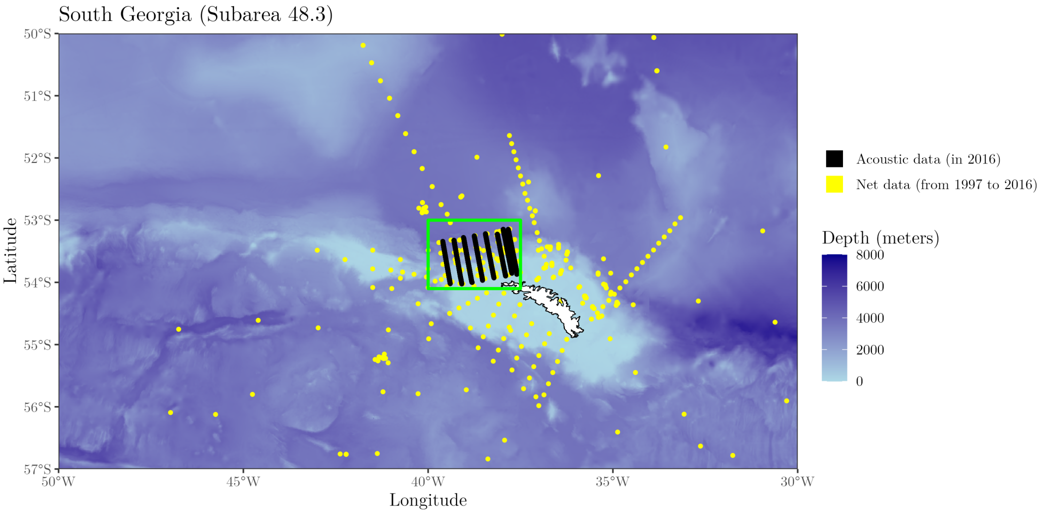
† Joint work with Sophie Fielding (BAS), Emma Cavan, and Adam M. Sykulski.

# Antarctic Krill (*Euphausia superba*)

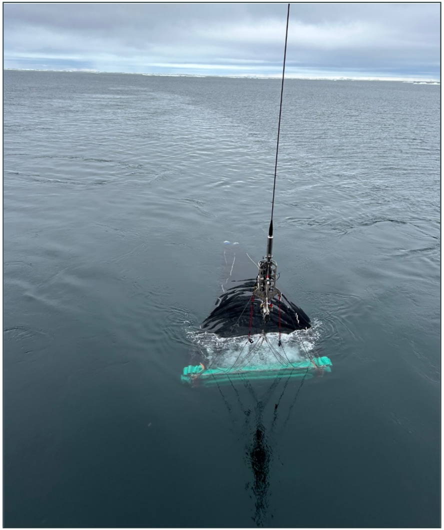
Antarctic Krill (*Euphausia superba*) factoids

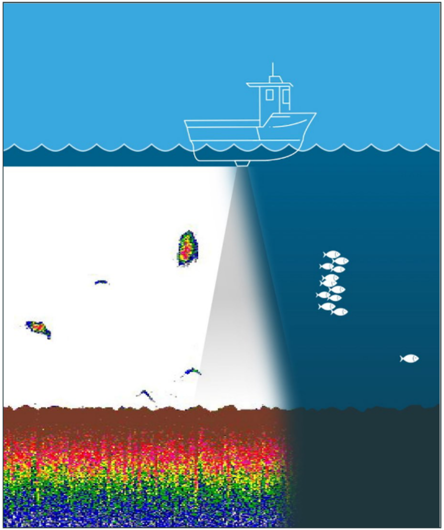
- Live up to 6 years.
- Grow up to 6 centimeters.
- Omnivorous, but can feed intensively on phytoplankton blooms.
- One of the most successful animals on the planet (approx. 500 million tonnes of biomass).
- Carbon cyclers, keystone species in the Antarctic food web, and the subject of fishery.

# Antartic Krill abundance



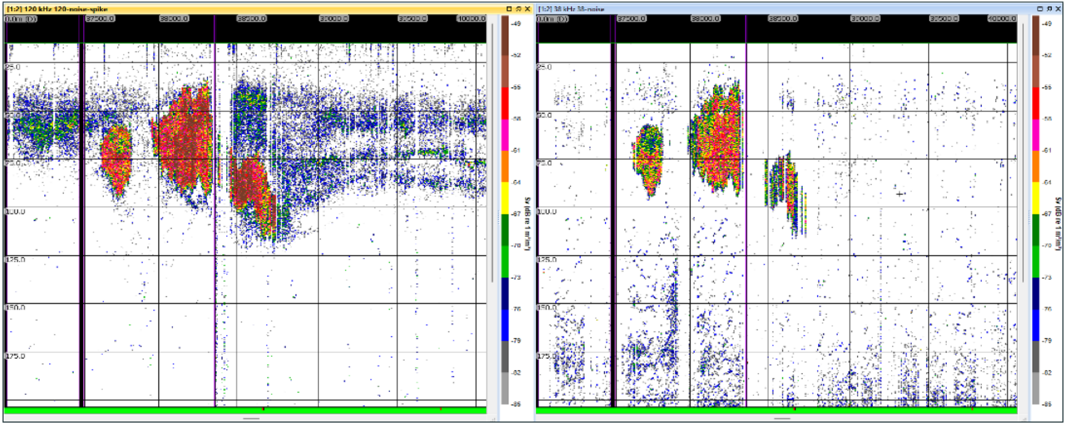
# **Antartic Krill abundance** (net and acoustic data)





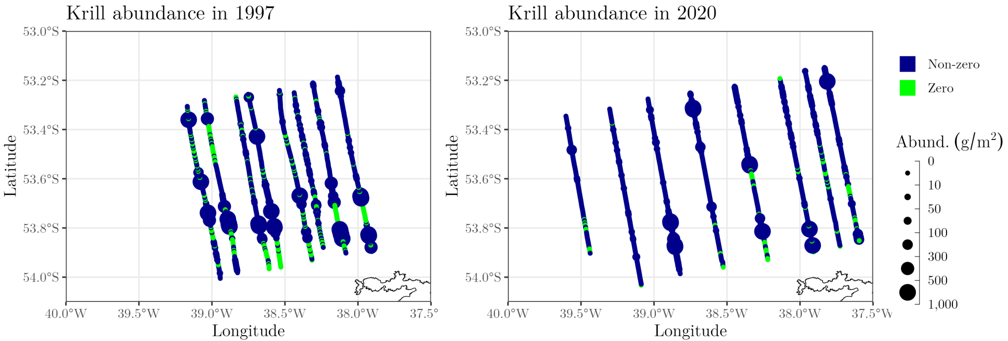
# Antartic Krill abundance (acoustic data)

Cleaned acoustic data (left: 120 kHz and right: 38 kHz). 3 krill swarms (largest is $\sim$ 500 m.).



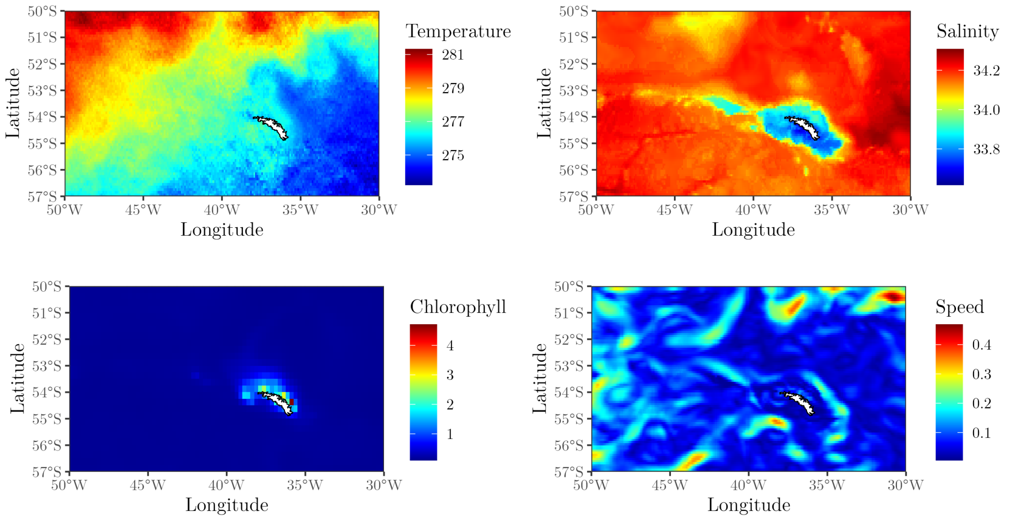
Collect frequencies → Cleaning and calibration → Identify krill swarms → Convert to biomass

# Western Core Box (WCB) ecosystem survey

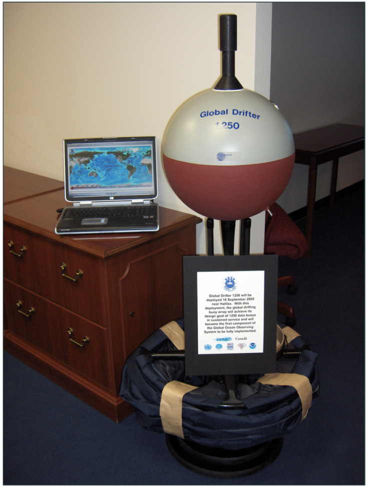



- Undertaken annually since 1997 (between December and February).
- 8 transects, each 40 nautical miles long (res. of 500 meters) and 10 nautical miles apart.
- Used to inform the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) on the ecosystem health.

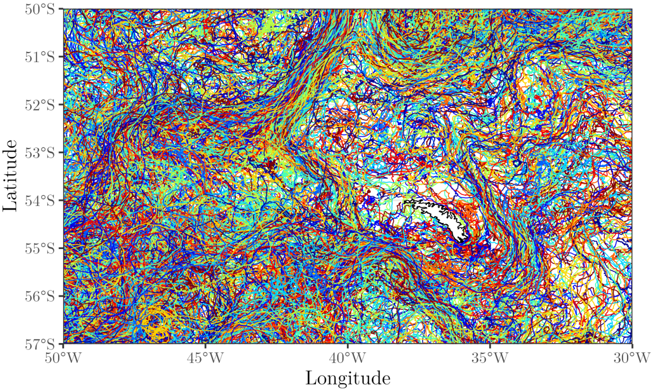
# Climate covariates obtained from satellite imagery



# Drifters (*Global Drifter Program*, by NOAA)




We use the drifter trajectories to derive other useful products.




This plot shows the trajectories of all drifters that floated through South Georgia from 1997 to 2020.

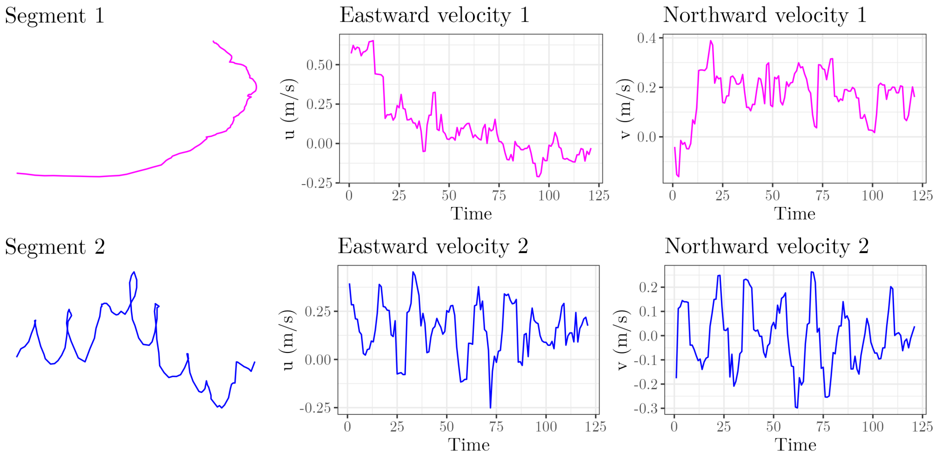
# All covariates

**Table 1:** Covariates for modelling the spatio(-temporal) distribution of krill. † indicates covariates obtained from satellite imagery, and ‡ indicates covariates derived as products from drifter trajectories. § denotes covariates observed only during the months of Dec., Jan., and Feb.

| Covariate | Spatial resolution (°) | Temporal resolution | Source |
|---|---|---|---|
| Bathymetry (depth)† | $0.01 \times 0.01$ | NA | NOAA (`10.25921/fd45-gt74`) |
| Slope | $0.01 \times 0.01$ | NA | Computed based on bathymetry |
| Chlorophyll† | $0.25 \times 0.25$ | Yearly§ | Copernicus Marine Service (`10.48670/moi-00019`) |
| Potential temperature† | $0.083 \times 0.083$ | Yearly§ | Copernicus Marine Service (`10.48670/moi-00021`) |
| Salinity† | $0.083 \times 0.083$ | Yearly§ | Copernicus Marine Service (`10.48670/moi-00021`) |
| Speed (satellite)† | $0.083 \times 0.083$ | Yearly§ | Copernicus Marine Service (`10.48670/moi-00021`) |
| Surface temperature† | $0.05 \times 0.05$ | Yearly§ | Copernicus Marine Service (`10.48670/mds-00329`) |
| Density of drifters‡ | $0.25 \times 0.25$ | 1997–2020 | Computed based on drifter trajectories |
| Expected frequency‡ | $0.01 \times 0.01$ | 1997–2020 | Computed based on drifter trajectories |
| Mass flux‡ | $0.01 \times 0.01$ | 1997–2020 | Computed based on drifter trajectories |
| Residence time‡ | $0.01 \times 0.01$ | 1997–2020 | Computed based on drifter trajectories |
| Speed (drifters)‡ | $0.01 \times 0.01$ | 1997–2020§ | Computed based on drifter trajectories |

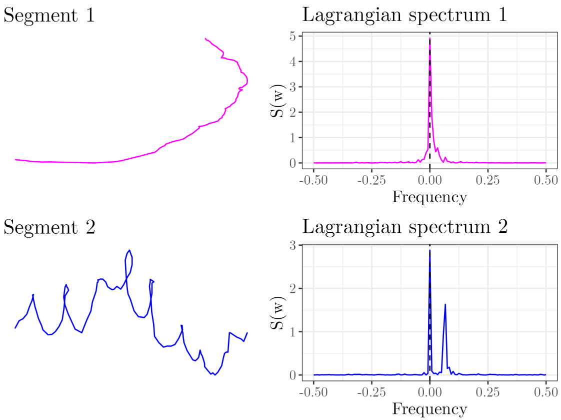
# Expected Frequency

Based on a drifter's positions, we can compute its Eastward ($u_t$) and Northward ($v_t$) velocities, such that $z_t$ can be represented as complex-valued time-series $z_t = u_t + iv_t$.



# Expected Frequency

Then, we can obtain an estimate of the Lagrangian spectrum $S(\omega) = |\mathcal{F}(\tilde{z})|^2/N$, where $\mathcal{F}(\tilde{z})$ is the discrete Fourier transform of the tapered time-series $\tilde{z}$, such that $\tilde{z} = (z_1, \cdots, z_N)$ after applying a tapering function to $z$.



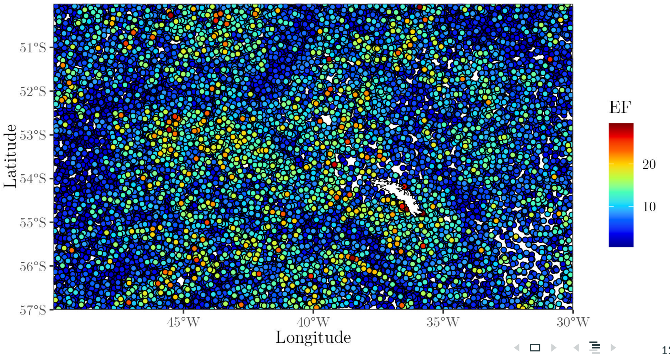
# Expected Frequency

In this setting, a possible estimator for the expected frequency (EF) $\phi$ of a particle in a location $s = (s_1, s_2)$ is given by

$$\phi(s) = \int_{\Omega} |\omega| S(\omega) d\omega \cdot k^{-1},$$

where $k = \int_{\Omega} S(\omega) d\omega$ and $s$ is the mid-point of a segment $\xi$.

From the previous two examples, the flux in $\xi_1$ and $\xi_2$ were 2.96 and 6.78, respectively.

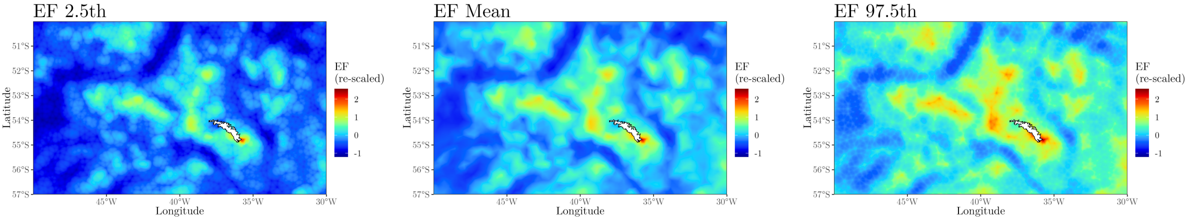


# Spatial Smoothing of Drifter Covariates

Let $y_i$ be the computed expected frequency at the location $s_i$, i.e., $y_i = \phi(s_i)$. Also, let $s = (s_1, \cdots, s_n)$. Thus, we can model the $\phi(s)$ in the following way

$$
\begin{aligned}
y_i &= \beta_0 + \phi(s_i) + \epsilon_i, \ \text{s.t.} \ \epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \sigma^2_\epsilon), \forall i \\
\phi(s) &\sim \text{Gaussian Process}(0, r_\phi(h; \theta)) \\
(\beta_0, \sigma^2_\epsilon, \theta)^\top &\sim \text{priors},
\end{aligned}
$$

where $r_\phi(h; \theta)$ is the Matérn kernel.



# Hurdle-Gamma Model

Now, let us model the krill abundance (with acoustic data) based on the obtained covariates (including "flux"). Given the many observed zeros, we will do it with a Hurdle-Gamma model.

Following the notation in Krainski et al. (2018), let

$$z_{it} = \begin{cases} 1, & \text{if the krill biomass is non-zero at location } s_i \text{ at time } t \\ 0, & \text{otherwise} \end{cases}$$

and

$$y_{it} = \begin{cases} \texttt{NA}, & \text{if the krill biomass is zero at location } s_i \text{ at time } t \\ \text{krill biomass}, & \text{otherwise} \end{cases}$$

where $z_{it} \sim \text{Binomial}(\pi_{it}, n_{it} = 1)$ and $y_{it} \sim \text{Gamma}(a_{it}, b_{it})$. Also, let $\mu_{it} = \mathbb{E}(y_{it})$, such that $\mathbb{E}(y_{it}) = a_{it}/b_{it}$.

# Hurdle-Gamma Model

The linear predictors in $z_{it}$ and $y_{it}$ are, respectively,

$$\text{logit}(\pi_{it}) = \beta_0^z + \beta_1^z \texttt{cov}_1^z + \cdots + \beta_{\ell_1}^z \texttt{cov}_{\ell_1}^z + \psi_{it}$$

and

$$\log(\mu_{it}) = \beta_0^y + \beta_1^y \texttt{cov}_1^y + \cdots + \beta_{\ell_2}^y \texttt{cov}_{\ell_2}^y + \gamma \cdot \psi_{it} + \nu_{it},$$

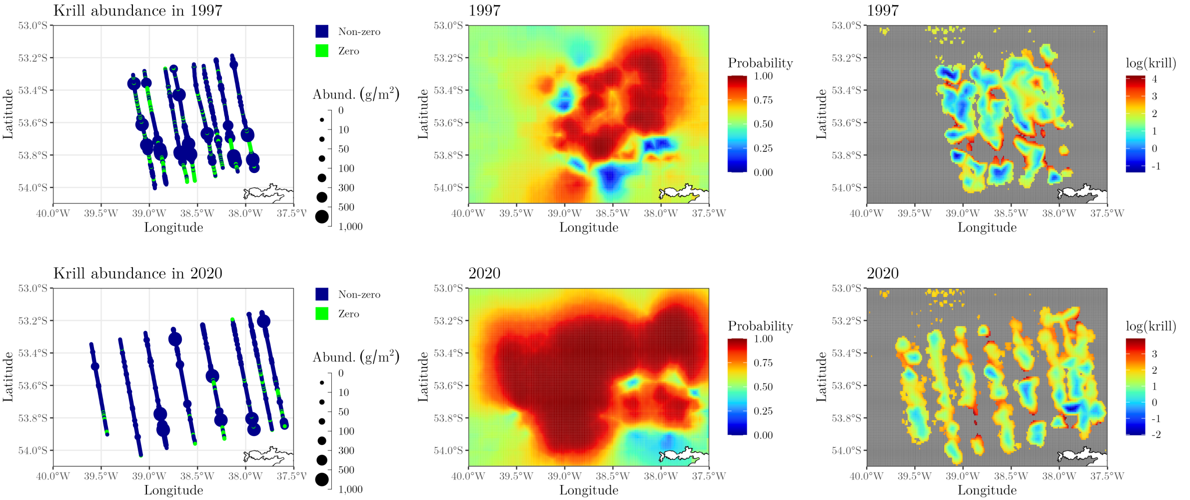
where the $\ell_1$ and $\ell_2$ covariates may be common between the two terms, $\gamma$ is a "copy" factor, and $\psi_{it}$ and $\nu_{it}$ are spatio-temporal random effects.

In particular, $\psi_{it}$ (similarly, $\nu_{it}$) is defined as

$$\psi_{it} = \alpha\psi_{i,(t-1)} + \phi_{it},$$

where $|\alpha| < 1$, $\psi_{i1} \sim \text{Normal}(0, \sigma_\phi^2/(1-\alpha^2))$, and $\phi_{it}$ is a temporally independent but spatially dependent GP at each year with a Matérn kernel (the model was fitted using `R-INLA`).

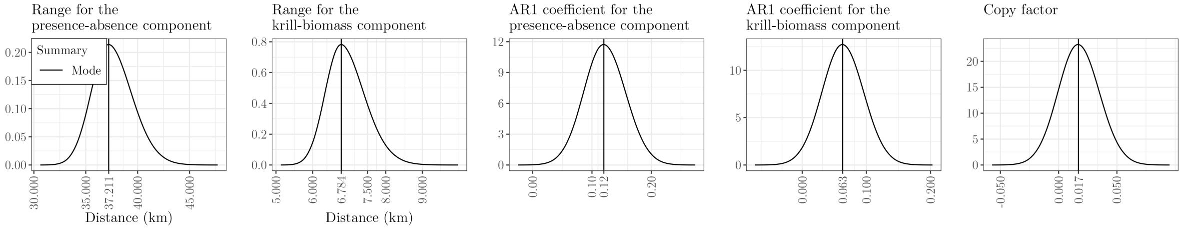
# Krill Abundance




The right-most plots mask out predicted values in locations where we have high uncertainty.

# Krill Abundance

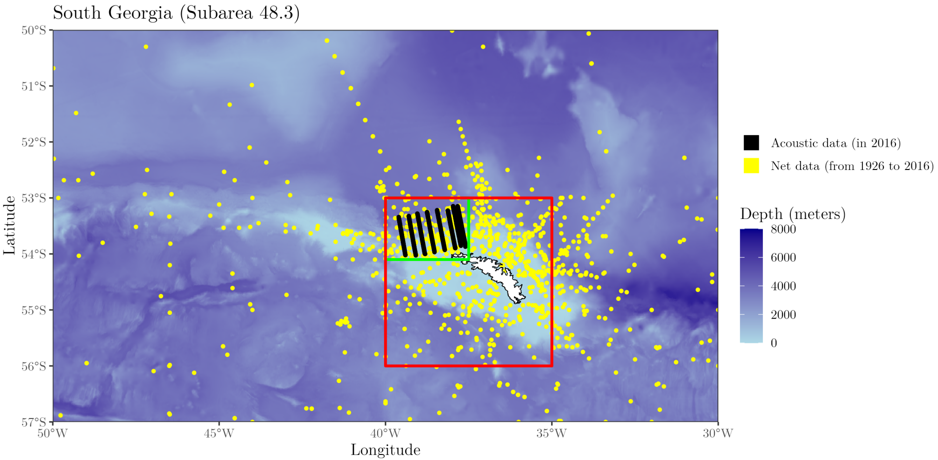
Regarding hyperparameter estimation, we obtained the following




Focusing on the two left-most plots, we can see that the posterior range for the presence-absence ($z_{it}$) process is 37.271 km (approx. 20 nautical miles), and posterior range for the krill-biomass ($y_{it}$) process is 6.784 km (approx. 3.7 nautical miles). This is useful for determining the sampling routes.

# Krill Abundance (aggregated data)

Alternatively, we can fit a model to the aggregated data (as in II)—both temporally and spatially (in $\sim 4$ km $\times$ 4 km cells). The focus is on extrapolation.



# Krill Abundance (aggregated data)

After performing stepwise forward variable selection based on the WAIC, the linear predictors in $z_i$ and $y_i$ are, respectively,

$$\begin{aligned}\mathrm{logit}(\pi_i) = \beta_0^z + &\\ &\beta_1^z \texttt{chlor} + \beta_2^z \texttt{pot\_temp} + \beta_3^z \texttt{speed\_sat} + \beta_4^z \texttt{surf\_temp} + \\ &\beta_5^z \texttt{mass\_flux} + \beta_6^z \texttt{res\_time} + \psi_i,\end{aligned}$$

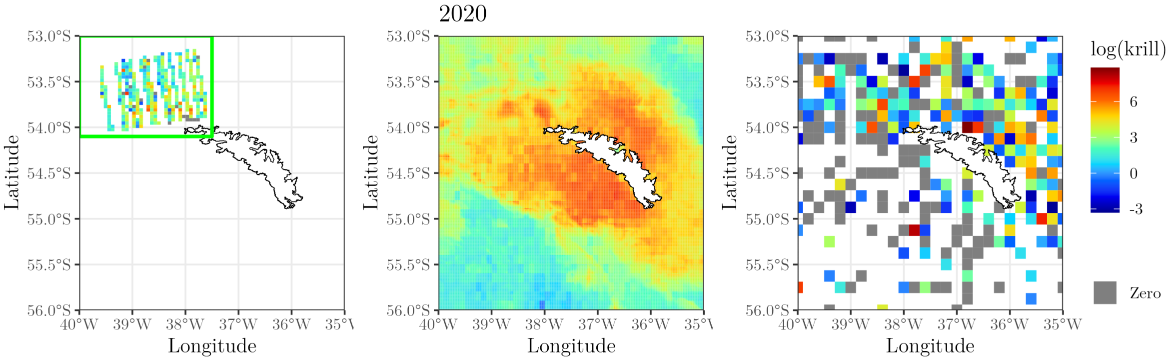
and

$$\log(\mu_i) = \beta_0^y + \beta_1^y \texttt{depth} + \beta_2^y \texttt{salinity} + \beta_3^y \texttt{surf\_temp} + \gamma \cdot \psi_i + \xi_i,$$

where $\gamma$ is the "copy" factor and the spatial random effects $\psi_i$ and $\nu_i$ are defined as GPs with a Matérn kernel.

*As a remark, we also tested different random structures for the same Hurdle-Gamma model, and the above choice performed best (according to WAIC and DIC).*

# **Krill Abundance** (aggregated data)




The left-most plot shows the *aggregated acoustic data*, the centre plot shows the *predicted process* in 2020, and the right-most plot shows the *aggregated net data* (for validation[1]).

[1] Although the net data do not represent the same quantity as the acoustic data—they only serve as a proxy.

# Final remarks

We modeled krill abundance using a novel dataset comprising acoustic *in situ* data of krill swarms.

To achieve this, we integrated climate covariates obtained from satellite imagery with information gathered by floating buoys (also known as drifters).

From the drifter trajectories, we derived an estimator for expected frequency (and others) that could be useful in our model.

We fitted a Hurdle model for acoustic data in both the (I) original form and (II) aggregated form. Each provided different insights.

We hope this work contributes to informed decision-making for krill management and fishing regulations.

# Thank you!

We also thank the "Turner Kirk Trust" foundation for their support of this project.

Turner Kirk Trust