1 Introduction

This document provides supplemental results to the main text, *Risk Factors for Fouling Biomass: Evidence from Small Vessels in Australia*.

2 Distribution of Biomass by Geographical Location

As discussed in Section 2.1 of the main paper, data for this paper were collected from five geographical locations. Figure 1 shows the measured (log) wet weight of sampled biomass by the geographical location of the vessel. There are no clear differences in the distribution of measured biomass between geographical locations, hence our models (Section 2.2) do not contain an adjustment for geographical location.

3 Convergence of Multiple Imputation

Figure 2 displays the observed and imputed vessel-level data for both the number of days since the vessel was last used, and the median number of trips (per year) by the vessel. Figure 3 displays the observed and imputed vessel-level type of paint. Both Figures 2 and 3 show that the distribution of imputed data is broadly consistent with the distribution of the observed data, from which we conclude that the imputation has been successful. Note that in Figure 3, a random subset of imputations has been selected for display due to space limitations.

---

1Note that all numeric predictors were scaled to have mean 0, standard deviation 1, prior to modelling.
Figure 1: Measured (log) wet weight (gm) of sampled biomass by geographical location of the vessel. Vessels from the Hobsons Bay Yacht Club (4) and the Royal Yacht Club of Victoria (2) were removed due to the low vessel numbers. Samples that measured less than the limit of detection (1.5gm) have been removed.
Figure 2: Comparison of observed and imputed data for the number of days since the vessel was last used, and the median number of trips by the vessel. The leftmost boxplot in each figure is the observed distribution, all others include the imputed values. Imputation has been successful as the distribution of imputed data is broadly consistent with that of the observed data.

Figure 3: Comparison of observed and imputed data for anti-fouling paint type. A random selection from the imputed datasets are shown. The top-left panel (labelled 0) shows the observed distribution. Imputation has been successful as the distribution of imputed data is broadly consistent with that of the observed data.
4 Outcome Model

The t-distribution observation model (O2) outperform the Normal model (O1). The difference in the leave-one-out information criterion (LOOIC, analogous to AIC, see Vehtari et al., 2016) was 47 (standard error 13) in favour of the t-distribution observation model (O2). A full comparison of the observation models for latent-process model (M0)–(M2) is provided in Table 1.

As a further check on model performance we examined if the models produce posterior predictions that are consistent with the observed data. An obvious choice of posterior predictive check (PPC) is the proportion of posterior predictions that lie below the limit of detection; we also compare the median posterior predictions and the interquartile range (IQR) of the posterior predictions. The last two PPCs test the model’s ability at generating data with a similar location and spread as the observed data. Figure 4 compares the distribution of the PPCs between (O1, top row) and (O2, bottom row); the specific PPCs are shown in the columns of the figure, with the observed value shown as the solid vertical line. Both models capture the spread of the data well, with the observed interquartile range falling in the middle of the distribution of the posterior predicted values. Neither model reproduces data consistent with the median or the proportion of measurements below the limit of detection, but nonetheless (O2) still appears to outperform (O1). For this reason, all latent-process model coefficients were subsequently estimated assuming the t distribution observation model (O2).

Table 1: LOOIC-based comparison of analysis models (M0)–(M2), for outcome models (O1) and (O2). The model with the smallest LOOIC (M2) is shown first, with subsequent rows ordered by increasing LOOIC (smaller LOOIC values are preferred) within outcome model (O2). ∆LOOIC shows the difference in LOOIC between analysis models fit using the outcome models (O2) and (O1), positive values indicate (O2) fits better; se(∆LOOIC) shows the estimated standard error of the difference. Eff. P gives the estimated effective number of parameters; se(Eff. P) shows its standard error.

| Model | t outcome model (O2) | Normal outcome model (O1) | Difference (O2 - O1) |
|-------|----------------------|---------------------------|---------------------|
|       | LOOIC               | se(LOOIC)                  | LOOIC              | se(LOOIC)                     | ∆LOOIC | se(∆LOOIC) |
| M3    | 1102                | 43.4                      | 1149               | 36.4                           | 47.1   | 12.7       |
| M1    | 1103                | 43.3                      | 1149               | 36.3                           | 46.9   | 12.6       |
| M0    | 1106                | 43.4                      | 1150               | 36.5                           | 44.7   | 12.6       |
Figure 4: Comparison of posterior predictive checks from analysis model (M0) between outcome models (O1, top row) and (O2, bottom row). From left to right the panels show posterior predicted histograms of: interquartile range (IQR); median; and the proportion of values below the limit of detection (1.5gm). The vertical lines in each panel show the observed values.
Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2016). “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”. en. In: Statistics and computing, pp. 1–20. ISSN: 0960-3174, 1573-1375. DOI: 10.1007/s11222-016-9696-4.