Marine Pollution Bulletin 198 (2024) 115905

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

How we count counts: Examining influences on detection during shoreline surveys of marine debris

Hillary K. Burgess [a,c,*], Timothy T. Jones [b], Jacqueline K. Lindsey [b], Carlie E. Herring [c], Sherry M. Lippiatt [d], Julia K. Parrish [b], Amy V. Uhrin [c]

[a] *Genwest Systems, Inc., Edmonds, WA, USA*
[b] *School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA*
[c] *Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Office of Response and Restoration, Marine Debris Division, Silver Spring, MD 20910, USA*
[d] *Scoot Science, Santa Cruz, CA, USA*

ARTICLE INFO

Keywords:
Marine debris
Shoreline monitoring
Macrodebris
Detection rate
Observer bias

ABSTRACT

Shoreline surveys are a common approach for documenting loads of marine macrodebris (≥ 2.5 cm) loads. When surveys are conducted repeatedly over time and space, patterns in source, abundance, geographic distribution, and composition can be detected. Yet to realize their full potential, monitoring programs that rely on surveys must grapple with high variability in debris abundance, and appropriately manage uncertainty when reporting estimates of debris quantity. A potentially important source of bias in estimating debris loads from shoreline monitoring datasets is variability in debris detection rates. With this in mind, we conducted field experiments using common strip-transect marine debris survey protocols, designed to test detection of macrodebris. We quantified how protocol, shoreline, and debris characteristics influence the detectability of marine macrodebris. Detection rates varied according to debris distance from observer (0–5 m), number of observers, debris characteristics (size, color), and shoreline substrate. Our results highlight considerations for monitoring program design. Comparisons across datasets should be approached cautiously given differences in survey protocols and sources of bias that may affect debris density estimates should be quantified and addressed. We hope these results will inform marine debris monitoring efforts that are optimized for intended data use and impact.

1. Introduction

Plastic debris in the world's oceans has been described as a creeping crisis (Mæland and Staupe-Delgado, 2020). Shoreline debris loads are commonly monitored as an indicator of the state of the marine debris problem (GESAMP, 2019) as shorelines mark the interface between land and sea, are accessible, and are often in close proximity to land-based sources of debris. Shoreline debris loads can be measured repeatedly over time and space, allowing for detection of patterns in source, abundance, geographic distribution and composition (GESAMP, 2019) as well as associated drivers including disaster events (Murray et al., 2018), ocean currents (Van Sebille et al., 2020; Gennip et al., 2019), and human population density and activities (Schuyler et al. 2018a; Hardesty et al., 2017a; Hardesty et al., 2017b; Willis et al., 2017). Monitoring data can also support prioritization and evaluation of preventative actions and policies. When data are collected over time, they allow for "before and after" intervention comparisons (Harris et al., 2020; Uhrin et al., 2020; Blickley et al., 2016) and when collected over space, they can allow for comparisons under different policy or management regimes (Schuyler et al., 2018a). However, these comparisons are only possible when the protocols are similar enough that differences in search effort and approach are not confounding. Methods that are readily and consistently adoptable by many data collectors, and data that can be interoperable post-hoc (Serra-Gonçalves et al., 2019; Browne et al., 2015), are crucial for tracking progress of marine debris prevention globally.

As noted elsewhere (e.g. Uhrin et al., 2022; Serra-Gonçalves et al., 2019; GESAMP, 2019; Browne et al., 2015) there is no single, standard protocol for shoreline debris monitoring. However, most employ a variation on plot surveys (Lippiatt et al., 2013; Hanke et al., 2013; Wenneker et al., 2010; Ribic et al., 1992; Hong et al., 2014), or survey via strip transects serving as replicates within a larger area (Burgess

* Corresponding author at: 7600 Sand Point Way NE, Seattle, WA 98115, USA.
E-mail address: hillary.burgess@noaa.gov (H.K. Burgess).

https://doi.org/10.1016/j.marpolbul.2023.115905
Received 9 March 2023; Received in revised form 30 November 2023; Accepted 3 December 2023

et al., 2021; Schuyler et al., 2018b; Parrish and Burgess, 2017; Lippiatt et al., 2013) whereby observers count debris items from within a geographic sampling unit of known dimensions to establish marine debris items per unit length or area of shoreline (Burgess et al., 2021; GESAMP, 2019; Schuyler et al., 2018b; Cheshire et al., 2009; Ryan et al., 2009). Therefore, understanding the influence of protocol characteristics, debris characteristics, and shoreline characteristics on debris detection within and across these methods is important for making comparisons of debris loads over time and space, ensuring that prevention and management decisions are made accounting for uncertainties and with the best available data.

Prior analyses of marine debris monitoring methods and datasets have highlighted considerations for data quality, interpretation and interoperability that include characteristics of data collectors, search area and duration, debris characteristics, and the environment or conditions of the survey (Lavers et al., 2016; Hardesty et al., 2017b; van der Velde et al., 2017; McWilliams et al., 2018; Angelini et al., 2019; Uhrin et al., 2020). While there is evidence that data collected by students and adult volunteers is of quality equivalent to, or greater than that by professionals (van der Velde et al., 2017; Hidalgo-Ruz and Thiel, 2015), characteristics of participants such as height, education level, and experience correlate with varying detection of marine debris (Angelini et al., 2019; van der Velde et al., 2017; Lavers et al., 2016). Larger survey team size and observer fatigue, two aspects of search effort, correlate with detection of higher and lower debris loads, respectively (Uhrin et al., 2020; Hardesty et al., 2017b; Lavers et al., 2016). Debris size and its treatment within a protocol, i.e. inclusion of a lower size limit (Hardesty et al., 2017b) as well as debris color (Angelini et al., 2019) and substrate type (Lavers et al., 2016; McWilliams et al., 2018), have been shown to influence debris detection and load estimates.

With the goal of optimizing marine debris shoreline monitoring, for data use and impact we considered how data collection protocols, shoreline characteristics, and debris characteristics influence the detectability of marine debris during surveys. Specifically, we asked the following questions:

Q1. . Does debris detection improve as the number of observers increases (where an observer is a survey participant engaged in the search for marine debris)?

Q2. . How does the observer search pattern influence debris detection?

Q3. . Is debris detection influenced by substrate (sand, cobble)?

Q4. . Is debris detection influenced by shoreline zone (i.e., vegetation, wood, bare, wrack, surf)?

Q5. . Is debris detection influenced by debris item size and/or color?

Q6. . Is debris detection influenced by linear distance between the item and the observer?

2. Methods

We designed two field trial approaches (***fixed effort***, ***variable effort***) that collectively assessed the influence of protocol, debris, and shoreline characteristics on detection of marine debris. ***Fixed effort*** trials kept survey effort constant (single observer per transect), and were designed to assess individual item detection and how detection rate varied with shoreline and debris characteristics, as well as debris distance from the observer. The subsequent ***variable effort*** trials utilized paired analyses to examine gross detection rate, or the proportion of items detected during a survey, as a function of the number of observers (one, two, or three) and search pattern.

In the sections below, we describe the overall field set up that was common to both sets of trials, followed by the specifics of each trial, and end with the approach to data analysis for each trial.

2.1. Data collection location and timing, and participant recruitment

In the fall of 2018 and spring through fall of 2019 (Table 1), we conducted field trials at three sites in the greater Seattle area of Washington State USA: Port Townsend Marine Science Center (PTMSC), Carkeek Park, and Ocean City (Fig. 1). These sites were chosen for their public accessibility and presence of amenities including onsite parking, restrooms, and running water. Sites had driftwood and vegetated back barrier features as well as cobble (PTMSC, Carkeek Park) or sandy (PTMSC, Ocean City) shoreline substrates.

We recruited volunteer observers from the Coastal Observation and Seabird Survey Team (COASST) participant corps, University of Washington students, faculty and staff, and through advertisement with regional National Oceanic and Atmospheric Administration (NOAA) and COASST partner organizations ($n = 99$ participants). Attendance was variable across field trial days (Table 1). Many participants ($n = 18$) attended multiple field trial days but on any given day, no single person surveyed an individual transect more than once (Table 1).

2.2. General field setup

We define a *trial* as a site- and day-specific event in which shoreline marine debris surveys were carried out; *transect* refers to within-trial experimental units ($n = 3$–4 per trial; Table 1) that were established at fixed locations within a site on a given day; and *survey* refers to a search made by an observer or team of observers on a specific transect. In each trial, we established five meter wide transects oriented from the water's edge to the backshore, and five meters into the vegetation or up to the first physical barrier (e.g., parking lot, seawall). Transects of five meters width were chosen because they are the standard in U.S. shoreline monitoring programs (Lippiatt et al., 2013; Parrish and Burgess, 2017). Transect locations were chosen for diversity in shoreline characteristics, attempting to encompass areas with representative shoreline zones of *surf* (area from the water's edge to the tide line), *wrack* (flotsam accumulating area from the lowest tide line to the uppermost fresh tide line), *driftwood* (upper part of the shoreline that accumulates driftwood from storm surges in the Pacific Northwest, USA), and vegetation (back barrier marked by shrubs or dune grasses). High-traffic (pedestrian) areas of the shoreline were avoided. Because public access to the shoreline was not restricted during our surveys, we intentionally minimized the extent of the transect into the portion of the shoreline closest to the water's edge to allow beachgoers to pass by without disrupting the transect.

During a trial, two to four transects were established on the

Table 1
Summary of field trial location and dates, including number of transects (n trans.), number of observers (n obs.), and a breakdown of the number of each survey type performed.

Trial #	Site	Date	Trial type	n obs.[a]	n trans.	Surveys
1	PTMSC	Sep-2018	Fixed	19	3	45 (Left: 16, Mid: 16, Right: 13)
2	Carkeek	Apr-2019	Fixed	18	2	34 (Left: 10, Mid: 12, Right: 12)
3	PTMSC	June-2019	Variable	22	4	18 (1-Obs: 8, 2-Obs: 6, 3-Obs: 4)
4	Carkeek	Jul-2019	Variable	21	4	22 (1-Obs: 8, 2-Obs: 8, 3-Obs: 6)
5	Carkeek	Oct-2019	Variable	14	3	14 (1-Obs: 6, 2-Obs: 5, 3-Obs: 3)
6	Ocean City	Nov-2019	Variable	35	4	39 (1-Obs: 14, 2-Obs: 14, 3-Obs: 11)

[a] These numbers represent the number of individuals that took part in that day's field trial, but as individuals surveyed as a pair, or trio, the number of combinations of survey "team" was larger than this.

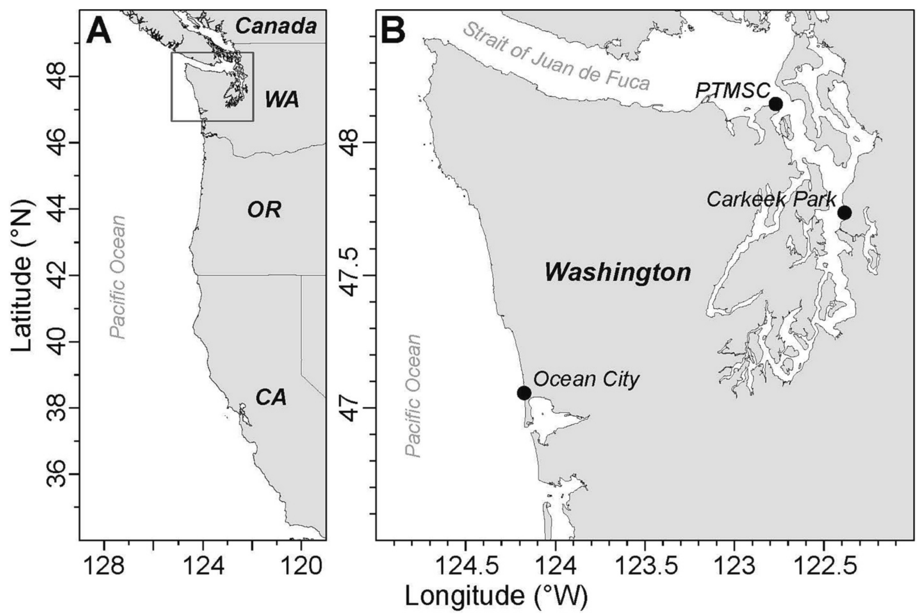

Fig. 1. An outline of the west coast of the United States and the State of Washington are provided for reference (A). Location of the three field trial sites (closed black circles) in Washington state (B). Fixed effort trials occurred at Carkeek Park and Port Townsend Marine Science Center, variable effort trials occurred at Carkeek Park, Port Townsend Marine Science Center, and Ocean City.

shoreline, each delineated by colored flags delimiting the corners, edges, and midline of the transect (Fig. 2). The number of transects was determined by the number of staff available to guide the volunteers and ensure quality control. Transects were thoroughly cleared of all existing marine debris by repeatedly passing over the entirety of the transect and removing all debris items by hand and disposing of them until no further items were found. Then, each transect was populated with 20 debris items ($\geq$ 2.5 cm) from one of four experimental debris kits (Table S1). This sample size fit the criteria of falling within the average range of debris quantities found on historical COASST surveys from within the

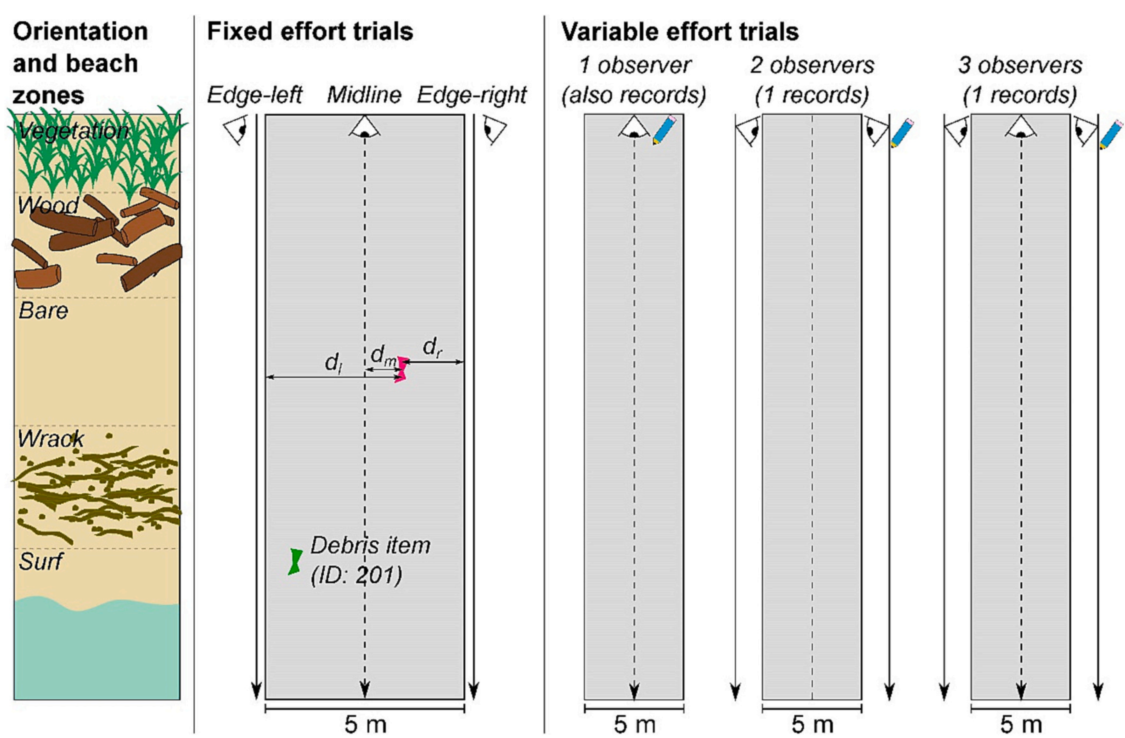

Fig. 2. Search pattern schematic for fixed effort (left panel) and variable effort trials (right panel). In fixed effort trials there were three search patterns: a single observer searched from the line at Edge-left, Midline or Edge-right. In variable effort trials there were three search patterns: 1 Observer - One lone observer walking down the midline of the transect records all information, 2 Observers - Two observers walking down the edges (opposite one another) of the transect while one of them records all information, and 3 Observers - Three observers where two search walking down the edges of the transect and a third searches down midline, one of which records information. Transects were seeded with debris, and dependent on the search pattern, measurements of distance from observer (d_l, d_m, and d_r, respectively) were made from the corresponding search line to the piece of debris found.

region (unpublished data, available by request), while being reasonable enough to manage and recover during a field trial. Cleaning the shoreline and using a known quantity, location, and type of debris allowed us to document which items were detected during each transect survey. The composition of items in each experimental debris kit was based on the proportion of different debris characteristics (type, material, size, and color) within the COASST dataset (characteristics are described in Parrish and Burgess, 2017, see supplementary materials Table S1). We directly labeled or tagged and inventoried each debris kit item with a serial number so that item detection within a survey could be uniquely attributed to item characteristics and location. Items ranged in size from 2.5 cm up to 50 cm in the longest dimension. To represent item size, we used maximum item surface area calculated as the product of the item's two longest orthogonal dimensions (i.e. length by width). Within each transect, we scattered debris on the surface haphazardly such that it was present in all shoreline zones. We buried the inventory tags in the substrate to prevent sighting as a function of inventory tag rather than debris item; tags also served as anchors, preventing debris movement.

For both fixed and variable-effort trials, observers were separated in space and/or time to minimize the chances that an incoming observer could see or hear others actively conducting surveys. To ensure an active observer did not bias the surveys of subsequent observers, transects were "reset" after each survey, smoothing the substrate and reburying tags as needed. Debris items and their placements remained unchanged among surveys in a given transect. Thus, any observer on a given transect experienced the same experimental debris kit and similar amounts of surface trampling.

2.3. Fixed effort trials

Fixed effort trials involved a single observer conducting each survey and were performed at PTMSC (September 2018) and Carkeek Park (April 2019; Table 1). These trials focused on research questions Q2, Q3, Q4, Q5, and Q6, and by assessing gross survey detection rate (proportion of items observed in a survey vs not; hereafter, gross detection rate) as a function of search pattern (defined below) as well as individual item detection rate as a function of item (size, color) and shoreline (substrate, zone) characteristics, and debris distance from observer (defined below). Observers were taught how to perform surveys using two basic search patterns: edge searches, where the observer looked into the transect from the right or left edge (a distance of 5 m); and midline searches, where the observer walked down the middle of the transect searching in both directions toward each edge (a distance of 2.5 m in either direction; Fig. 2). In both cases, observers were instructed to begin their survey from the back of the shoreline walking toward the water. During a survey, the observer remained in a standing position and scanned the transect area looking laterally and slightly forward for any "manmade items" >2.5 cm (roughly the size of a bottle cap). They were not aware that known quantities of debris had been distributed throughout the transect, or what the expected count of debris was in each transect. They were not permitted to move into the transect, backtrack or look backwards as this would have compromised the distance variable. If an item was sighted by an observer, they called out to research staff who then logged the item inventory number and measured the minimum orthogonal distance (i.e. at a right angle to the direction of travel) between the observer and the item in centimeters (distance from observer). This distance was chosen, rather than the distance the item was first sighted, to represent the minimum distance at which an item could be detected (or not). If the debris item's tag was exposed in the course of collecting data, staff reburied the tag before moving to the next item.

Observers performed between one and three surveys during a trial, moving among randomly assigned transects and shifting among search patterns (edge-left, edge-right, midline) such that search pattern diversity was maximized within observer, and equilibrated across transects within trial. Some attempt was made to increase the number of midline surveys, to ensure adequate sample size should subsequent analyses determine no effect of left or right edge search pattern, allowing edge surveys to be binned.

2.4. Variable effort trials

Variable effort trials addressed research questions Q1 and Q2 and were designed to assess gross detection rate as a function of the number of observers and their search pattern. During these trials, we varied the number and arrangement of observers according to one of three search patterns (Fig. 2). Variable effort trials were performed at PTMSC (June 2019), Carkeek Park (July and October 2019), and Ocean City (November 2019; Table 1). Because these trials had multiple simultaneous observers, distance from observer was not measured, and so we limit detection analyses from these trials to overall gross detection rate, as a function of item characteristics only (size, color), omitting analyses of item location (zone, substrate) and distance from observer. In variable effort trials, observers also recorded data themselves (item serial number), which necessitated leaving their search line to inspect the item. Thus, variable effort trials were more representative of a standard marine debris survey, albeit with different levels of effort.

2.5. Data analysis – fixed effort trials

To examine overall differences in detection rate among the fixed effort search patterns, we calculated gross detection rate as the proportion of items observed out of the 20 deployed debris items for each survey performed. We then calculated the mean and 95 % confidence intervals of detection rate via bootstrap resampling (1000 permutations), for each of the three fixed-effort search patterns (edge-right, edge-left, midline). To control for inter-transect variability in detection rate (i.e. among trials and among transects within trials) we also calculated a bootstrap mean (±95 % CI; 1000 permutations) for paired differences in detection rate between midline and edge search patterns conducted within the same transect.

A series of models were then used to estimate how item detectability varies with both item and shoreline characteristics, as well as search pattern and distance from observer (Table 2). Our dependent variable was individual item detection within a given survey, taking a value of 1

Table 2
Predictors used in detection rate models for fixed effort surveys. Reference value refers to the value at which that predictor was held constant when obtaining model-averaged predictions of detection rate.

Name	Type	N_{levels}	Levels/range	Effect type	Reference value
Distance	Continuous		edge: 0–5 m midline: 0–2.5 m	Fixed	2 m
Item size[a]	Continuous		1.5–18.5 cm	Fixed	3 cm
Site	Factor	2	Carkeek, PTMSC	Fixed	Ptmsc
Search pattern	Factor	2	Edge, midline	Fixed	Edge
Item color	Factor	4	Bright, dull, clear, white	Fixed	White
Substrate	Factor	2	Cobble, sand	Fixed	Sand
Zone	Factor	4	Bare, wrack, wood, vegetation	Fixed	Bare
Observer id	Factor	37		Random	
Survey id	Factor	79		Random	
Item-location id	Factor	100		Random	
Transect id	Factor	5		Random	

[a] Item size was included in the model as the square root of the item's maximum visible surface area (product of each item's longest two dimensions).

if the item was detected, and 0 otherwise. Because our response variable was binary, we used binomial Generalized Linear Mixed Models (GLMM; logistic link function) using the glmmTMB package (Brooks et al., 2017) to model the probability of item detection according to different combinations of predictor fixed and random effects (Table 2). Modelling detection at the item level allowed us to generate detection curves as a function of distance from observer as well as item and shoreline characteristics. Because transects were populated with a known debris quantity, modelled detection rate was absolute. We considered seven fixed effects (Table 2). Distance from observer and item size were both continuous variables. Categorical variables included search pattern (two levels), item color (4 levels), site (2 levels), substrate (2 levels), and zone (4 levels) (Table 2). Item size was square root transformed to reduce the potentially high leverage of the relatively fewer large sized items on fitted relationships. Because item color has many possible levels, we conducted a preliminary analysis to inform color groupings based on detection probability (see supplemental materials, Table S2, Fig. S2). This resulted in four color groups: bright (blue, yellow, red), dull (black, brown, grey), clear (including transparent with some colorful lettering as in a single-use cup with a brand imprinted), and white, which were used throughout all subsequent analyses. We also considered four random effects within our models (Table 2). Observer ID and transect ID account for consistently higher/lower detection rates for specific observers (i.e. some observers are more/less proficient than others) and transects (i.e. location-specific differences). Survey ID incorporates observer- and transect-level variability but potentially also accounts for differences in detection rate at the whole survey level that might be attributable to extrinsic (i.e. differences in lighting brought on by cloud cover) or intrinsic (i.e. relative observer fatigue) factors not accounted for by other model predictors. The random effect of item location accounts for multiple observations of the same item in a similar location within each transect, accounting for variability not accounted for by recorded item characteristics (size, color) and/or location (distance, substrate, zone). Search pattern (edge, midline) was included in our models as both a fixed effect, modelling differences in overall detectability between search patterns, and as an interaction term with distance from observer to model different changes in detectability with increasing distance for each search pattern. We also included quadratic terms for continuous variables (distance, item size) to allow for additional flexibility in the functional form (i.e. not a strict logistic function) of those fitted relationships.

Models representative of all possible predictor combinations were fitted to the data, and subsequently ranked based on small sample size corrected Akaike's Information Criterion (AICc). Model-specific AICc values were then used to calculate Akaike weights as a measure of the support in favor of a given model being the best within the full model set (Burnham et al., 1980; Galipaud et al., 2017). Akaike weights were then used to identify the 99 % confidence set of models as the ordered (lowest to highest AICc values) set of models whose cumulative Akaike weights sum to 0.99 (Symonds and Moussalli, 2011). Likelihood ratio tests and calculation of evidence ratios were utilized for specific inclusions/exclusions of predictors relative to the highest ranked model to identify the level of support for including those predictors within the model. Parameter estimates for all predictors were subsequently obtained via model-averaging (see Table S4).

To examine fitted relationships/differences among categorical predictor levels and to determine predictor importance based on estimated effect sizes, we calculated model-averaged detection rates for different search patterns, number of observers, and shoreline, and debris characteristics. Model-averaged values were obtained by holding all predictors constant, with the exception of the predictor whose effect was under examination (see Table 2 for predictor reference values). For continuous variables of item size and distance, model-averaged detection rates were calculated across the range of values for those variables (Table 2) to examine the functional form of those relationships. For categorical variables, model-averaged detection rates were calculated for each category level (Table 2). Due to the mix of categorical and continuous predictors we assessed individual predictor importance as the maximum difference in model-averaged detection rate among levels for categorical predictors, or across the range of observed values for continuous predictors. This gave a measure of maximal effect size that could be attributed to that predictor when changed within the constraints of the field trial design (i.e. maximum detection distance at 5 m, or minimum item size = 2.5 cm). Uncertainty in model-averaged detection rates was calculated using bootstrap resampling (bootMer function in the lme4 package in R; Bates et al., 2015), followed by model-averaging. For a given debris item (i.e. size, color, distance) and search pattern (right edge, left edge, midline), 10,000 bootstrap replicates of estimated detection rate were generated from each model within the model set, where variation among bootstrap replicates represents within model uncertainty. We then constructed a distribution of detection rates by drawing 10,000 values from the combined pool of bootstrap replicates across models, where the number of draws for a particular model was proportional to its standardized Akaike Weight (i.e. a model with $w = 0.3$ contributed 3000 bootstrap replicates to the combined distribution). The resulting distribution therefore reflects within (i.e. replicates from the same model) and among (i.e. replicates from different models) model uncertainty, and was subsequently used to calculate a model-averaged mean and 95 % confidence interval of detection rate for each debris item scenario. Because variable importance was constructed as the maximum difference in detection rate across a variable's range, paired differences for a particular predictor variable (i.e. detection rate at distance 0 m compared to 5 m, or the maximum compared to minimum detection rate across levels of item color) were calculated for each bootstrap set. The distribution of paired differences was then processed as described above to calculate the overall mean and 95 % confidence interval of effect size.

2.6. Data analysis – variable effort trials

For the variable effort trials, we were interested in overall debris detection (i.e. at the transect level) and how detection varies between survey effort (inter-effort), but also within survey effort (intra-effort) as a measure of survey consistency. As we did for the fixed-effort trials, we calculated bootstrapped 95 % confidence intervals of mean gross detection rate for each level of survey effort (1, 2, or 3 observers; Fig. 2). To control for variability among transects, we also calculated all possible pairwise differences in gross detection rate between (inter) and within (intra) search patterns, restricted such that pairwise comparisons were only calculated among surveys performed on the same transect. We then calculated bootstrapped 95 % confidence intervals of the mean pairwise difference in gross detection rates by generating 1000 randomly drawn (with replacement) sets of n pairwise differences, where n is the number of available pairwise differences for a given comparison (i.e. inter: 1 vs 2 observers; intra: 2 vs 2 observers), and calculating the mean value for each randomly drawn set. We then present the overall mean and 95 % CI as the mean and 95 % range of the resultant distribution of sample means. We also repeated these analyses on subsets of debris items based on size and color, to examine whether these differences were more/less pronounced for certain debris types. All analyses were performed in R version 4.2.1 (R Core Team, 2022).

3. Results

3.1. Fixed effort trials

A total of 79 fixed effort surveys (5 transects, 37 observers) were performed (Table 1). Three transects were surveyed by 19 observers at PTMSC in September 2018 while two unique transects were surveyed by 18 observers at Carkeek Park in April 2019 (Table 1). Because the 95 % confidence intervals overlap edge (right, left) for mean gross detection rate (right: 76.9 %, 95 % CI = 71.3–81.8 %; left: 70.0 %, 95 % CI =

63–76 %), we combined these results for a total of two search patterns: edge and midline. Edge surveys had a slightly lower mean gross detection rate at 73.2 % (95 % CI: 69.2–77.1 %) versus midline surveys at 77.0 % (95 % CI: 72.3–81.3 %), but mean pairwise differences (i.e. controlling for transect) between search patterns overlapped with zero (mean edge – midline difference = −3.3 %, 95 % CI = −9.0, 2.1), indicating that both search patterns performed similarly at the whole transect level.

3.2. Item detectability within fixed effort trials

Among all model permutations of item detection rate (probability of item detection), no particular model could be identified as unequivocally the best model as the highest ranked model had an Akaike weight of 0.044 (Table 3), and the 99 % confidence set of models contained 366 different models. Models included within the 99 % confidence set consistently included the predictors of distance, item size (linear and quadratic terms), and item color, as well as random effects of survey and item-location (Table S4). The high degree of model-selection uncertainty, as indicated by the low Akaike weight (0.044) of the highest ranked model, was driven by several predictors that had marginal effects on model likelihood, as well as competition between linear and quadratic functional forms of the distance relationship (see Table 3 for models with ΔAICc ≤2). The effect of distance was better modelled via a quadratic rather than a linear term, although differences in AICc (ΔAICc = 1.27) were small enough to suggest that the associated differences in functional form on model fit were marginal. The inclusion of shoreline zone and substrate were moderately supported based on likelihood ratio tests (zone: $\chi^2 = 9.05$, df = 3, $p = 0.029$, substrate: $\chi^2 = 4.14$, df = 1, $p = 0.042$) and evidence ratios (zone: ER = 4.37, substrate: ER = 2.86) between the highest ranked model and equivalent models excluding those predictors (Table 3). When added to the highest-ranked model, the additional variables of site (ΔAICc = 0.83), search pattern (ΔAICc = 1.8), and random effects of observer (ΔAICc = 0.33) all resulted in higher AICc values.

Predictor importance, as evaluated based on model-averaged detection rate effect sizes, were highest for distance and color (difference in detection rate, $\Delta_d = 0.5–0.6$), followed by item size and shoreline zone ($\Delta_d = 0.3–0.4$), whereas shoreline substrate ($\Delta_d = 0.12$), site ($\Delta_d = 0.05$) and search pattern assessed at 0 m ($\Delta_d = 0.00$) had the lowest effect sizes (Table 4). However, while shoreline substrate and zone had non-zero mean effect sizes, the confidence intervals for both of these factors encompassed zero, indicating relatively higher uncertainty in their importance compared to other predictor variables (Table 4).

Modelled detection rates as a function of size (distance held constant at 2 m) indicate that there is a decrease in detectability for items smaller than 6 × 6 cm (roughly equivalent to an extra large hen's egg), but that detection rate is high, and approximately constant for items above this size (Fig. 3A). Detection rate also varied with item color decreasing as follows: bright > white > clear > dull (Fig. 3B), albeit with considerable uncertainty for the latter two color groups. Detection rate as a function of distance (the minimum orthogonal distance between the observer and the item) are presented for two scenarios: a *small* (9cm^2, or 3 cm × 3 cm) white item, roughly a large bottle cap, and a *larger* (100cm^2, or 10 cm ×

Table 3
Models fitted to item detectability from fixed-effort trials in rank order (ΔAICc ≤2) (AICc = second-order bias adjusted AIC, ΔAICc = AIC differences, ω = Akaike weights, ER = evidence ratios).

#	Model	log-lik.	df	AICc	ΔAICc	w_{AICc}	ER
1	search:distance^2 + size + size^2 + color + substrate + zone + rand(survey) + rand(item-location)	−621.7	14	1271.6	0	0.044	1.00
2	search:distance^2 + size + size^2 + color + substrate + zone + rand(observer) + rand(survey) + rand(item-location)	−620.8	15	1271.9	0.33	0.037	0.84
3	site + search: distance^2 + size + size^2 + color + substrate + zone + rand(survey) + rand(item-location)	−621.1	15	1272.4	0.83	0.029	0.66
4	site + search: distance^2 + size + size^2 + color + zone + rand(survey) + rand(item-location)	−622.2	14	1272.6	1.04	0.026	0.59
5	site + search: distance^2 + size + size^2 + color + substrate + zone + rand(observer) + rand(survey) + rand(item-location)	−620.2	16	1272.7	1.06	0.026	0.59
6	site + search: distance^2 + size + size^2 + color + zone + rand(observer) + rand(survey) + rand(item-location)	−621.2	15	1272.7	1.13	0.025	0.57
7	search:distance + size + size^2 + color + substrate + zone + rand(survey) + rand(item-location)	−622.3	14	1272.9	1.27	0.023	0.53
8	search:distance + size + size^2 + color + substrate + zone + rand(observer) + rand(survey) + rand(item-location)	−621.3	15	1273.0	1.39	0.022	0.50

Table 4
Model-averaged predicted detection rate and effect sizes (maximal difference in detection rate attributable to that predictor) for predictor variables in the fixed-effort trials. Predicted detection rate and effect sizes are for a reference debris item (3 × 3 cm white piece of debris, located on bare sand 2 m away from an observer performing an edge survey), modified according to the presented level for each predictor variable.

Predictor	Level	Predicted mean detection rate		Predictor effect size	
		Mean	95 % CI	Mean	95 % CI
Site	PTMSC	0.83	[0.66, 0.95]	0.05	[−0.02, 0.28]
	Carkeek	0.78	[0.54, 0.94]		
Search [0 m]	Edge	0.91	[0.80, 0.98]	0.00	[−0.04, 0.06]
	Midline	0.91	[0.80, 0.98]		
Distance [Edge]	0 m	0.91	[0.80, 0.98]	0.59	[0.30, 0.76]
	5 m	0.32	[0.11, 0.61]		
Distance [Midline]	0 m	0.91	[0.80, 0.98]	0.49	[0.15, 0.77]
	2.5 m	0.42	[0.12, 0.77]		
Color	white	0.83	[0.66, 0.95]	0.55	[0.20, 0.83]
	bright	0.95	[0.88, 0.99]		
	clear	0.66	[0.43, 0.86]		
	dull	0.40	[0.14, 0.73]		
Size	1.5 cm	0.65	[0.34, 0.89]	0.34	[0.08, 0.68]
	18.5 cm	0.99	[0.97, 1.00]		
Substrate	sand	0.83	[0.66, 0.95]	0.12	[−0.01, 0.46]
	cobble	0.72	[0.36, 0.93]		
Zone	bare	0.83	[0.66, 0.95]	0.39	[0.00, 0.76]
	wood	0.81	[0.66, 0.93]		
	wrack	0.82	[0.64, 0.95]		
	veg	0.47	[0.13, 0.88]		

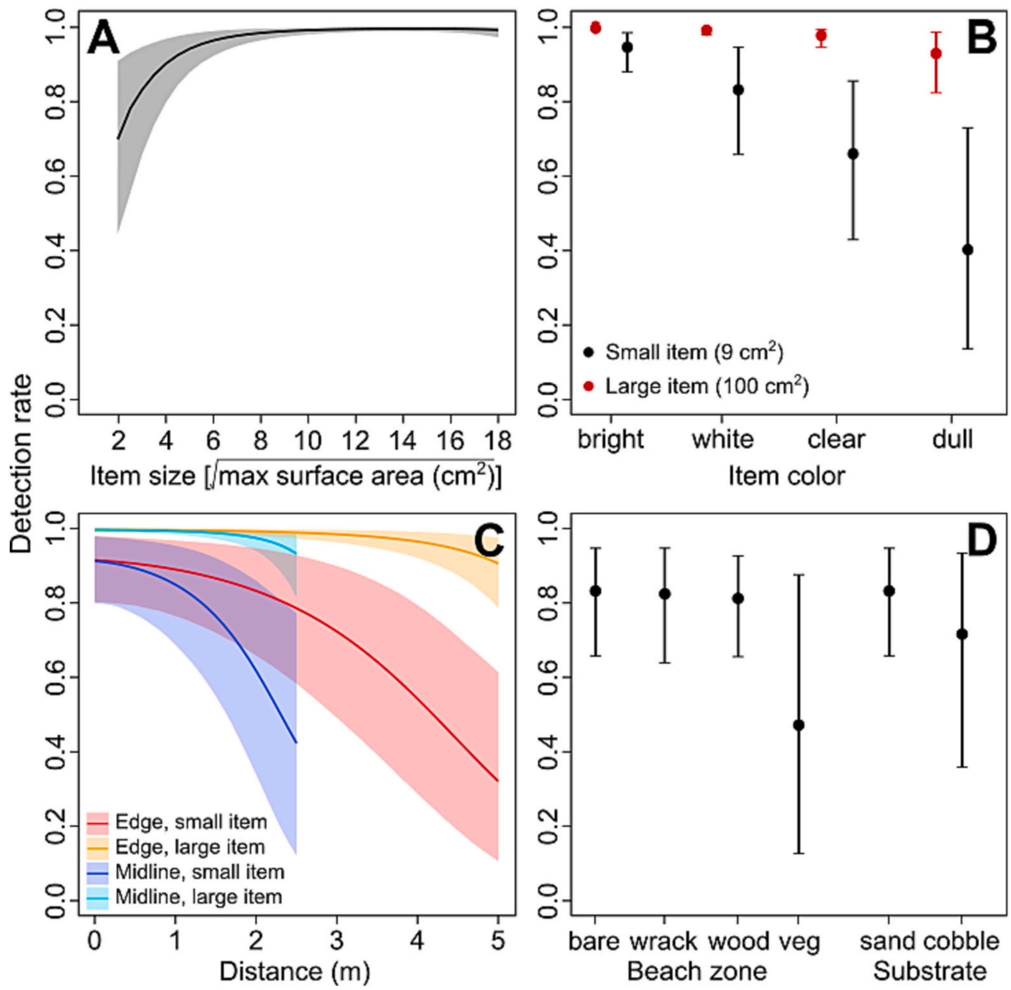

Fig. 3. Model-averaged mean item detection rates for effects of (A) item size, (B) item color, (C) orthogonal distance from the observer for edge and midline search patterns, and (D) beach zone and substrate where the item was located. Shaded areas and error bars indicate model-averaged 95 % confidence intervals of the mean predicted detection rate. Unless otherwise stated, predictions are made holding other model factors constant equal to: site = PTMSC, search pattern = edge, distance = 2 m, item color = white, item size = 9 cm^2 (3 cm × 3 cm), beach substrate = sand, beach zone = bare.

10 cm) white item, as would be the case for a half pint juice or milk carton. Increasing distance was predicted to result in an initially small decrease in detection rate within the first half of the transect width (1.25 m for midline, 2.5 m for edge search patterns) for each search pattern, followed by a more rapid decrease in detection for more distant items (Fig. 3C). The rate of change of detection rate was more extreme for midline surveys than for edge surveys (see also Table S3 for model parameter estimates), but due to the greater maximum search distance, edge surveys were predicted to have a lower detection rate at the opposite edge of the transect than for midline searches (Fig. 3C). Evaluated at half of the maximum distance for each search pattern (i.e. 1.25 m for midline surveys and 2.5 m for edge surveys) both were predicted to have similar detection rates for a small debris item (edge = 0.79, midline = 0.81; Fig. 3C).

Model-averaging suggested that detection rate was higher on sand than on cobble (Table 4, Fig. 3D). For small (e.g. bottle cap) items, detection rate on cobble was approximately 11 % lower compared to sand on average, although there was a high degree of uncertainty surrounding these estimates (Fig. 3D). Model-averaged detection rates were indistinguishable among the wrack, bare, and driftwood shoreline zones, but were higher than in the vegetation zone (Fig. 3D). However, the estimate of detection rate in the vegetation zone had a high degree of uncertainty.

Within our random effects, survey-level differences in detection rate were better modelled via the survey ID random effect, than via random effects attributable to observer and transect level variability (Table 3). However, observer random effects were included in the second ranked model in addition to survey random effects, which had almost equivalent support to the highest ranked model (evidence ratio of 0.84; Table 3). It is worth noting that observer variability may have been largely accounted for by the survey random factor, as surveys are necessarily nested within observers.

3.3. Variable effort trials

Variable effort trials were carried out on four separate dates among three locations, and data were collected from 93 surveys (15 unique transects, 92 observers), of which 39 %, 35 % and 26 % were carried out by single-, double-, and triple-observer teams, respectively (Table 1). Four transects were surveyed by 22 observers at PTMSC in June 2019, four unique transects were surveyed by 21 observers at Carkeek Park in July 2019, three unique transects were surveyed by 14 observers at Carkeek Park in October 2019, and four unique transects were surveyed by 35 observers at Ocean City in November 2019 (Table 1).

Increasing the number of observers improved gross detection rate of debris, with a single observer having an average detection rate of 80 % compared to 85 % for two observers, and 89 % for three (Fig. 4A). Increasing effort also decreased inter-observer (or inter-team) variability, measured as the average of all differences among teams within a site performing the same search pattern. Doubling the number of

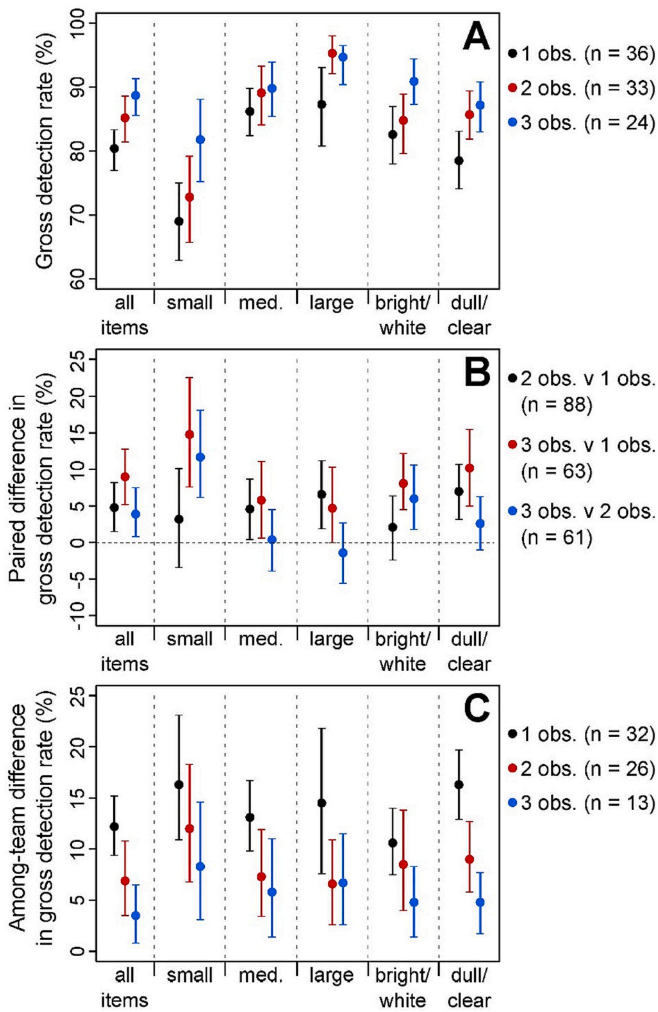

Fig. 4. Results of variable effort surveys (1–3 observers), expressed as (A) mean gross detection rate, (B) mean paired differences (controlling for transect) in gross detection rate between alternate team sizes and (C) average among team differences in gross detection rate. Results are presented for all debris items, as well as for items delimited by size (small: surface area < 16 cm^2, medium: surface area 16–64 cm^2, large: surface area > 64 cm^2) and color (bright or white items, versus dull or clear items), based on detection model results. All values were calculated via bootstrap resampling. Sample sizes (n) represent the number of surveys (gross averages, A), or the number of pairwise comparisons (pairwise differences, B—C) for that comparison.

observers reduced the mean among team difference in detection rate from ±12 % to ±7 %, and adding a third observer reduced this further to ±3.5 % on average (Fig. 4C).

Results were broadly similar when items were separated by size or color, but there were some notable differences (Fig. 4). The addition of a second observer had only a marginal effect on the detection of small (<16cm^2) items (+4.8 %), but adding a third observer resulted in a marked increase in gross detection rate (+12 % compared to 2 observers; Fig. 4A). However, for larger items (medium: surface area 16 – 64cm^2, large: surface area > 64 cm^2) the addition of a third observer had no effect on gross detection rate (Fig. 4A). Conflicting results were found for item color where bright/white items (pooled to boost item sample size) were only found more often after the addition of a third observer, whereas for dull/clear items, detection rates were the same for three versus two person teams (Fig. 4A). However, overall bright/white items were detected more often than dull/clear items, supporting our earlier findings (Fig. 4A). Consistency, as measured by among-team differences in gross detection rate showed a pattern of decreasing variability with increasing team size for all item classes, with the exception of detection rates of medium and large items which were equally variable between teams consisting of two and three observers, but lower than for a single observer (Fig. 4C).

4. Discussion

We examined shoreline macrodebris detection rates in real-world survey scenarios, documenting and quantifying potential sources of error in debris load estimates. We found that characteristics of the protocol, the debris itself, and of the shoreline can all influence the detectability of marine debris during shoreline surveys. Results are intuitive; items nearer to an observer are more likely to be detected than those farther away, larger team sizes detect more debris and have less variable detection among surveys, larger and more brightly colored items are more likely to be detected than smaller duller ones, and survey environments that are complex or have visual obstructions result in lower detection. The influences on detectability that we examined, and whose importance was supported by the model selection process, present considerations for marine debris shoreline survey design, analysis, and interpretation.

4.1. Contextualizing factors affecting detection: debris distance from observer

The largest effect size with high confidence was for orthogonal distance between an observer and a debris item. We found that debris items at the foot of an observer (0 m) were detected 91 % of the time on average but only 42 % of the time at 2.5 m and 32 % at 5 m. This finding suggests that monitoring protocols using strip transects of a given width with a prescribed search pattern (i.e. standardizing the maximum distance an observer is searching for debris) will have more consistent detection rates compared to surveys of larger plots without prescribed search patterns (e.g. most marine debris clean-ups where data are also collected). Standardized transect widths and search distances will maintain a consistent detection probability when other detection factors are consistent, and shorter distances from observers will have better detection rates than longer ones. Our results suggest that that the optimum search distance is no >1 m.

4.2. Contextualizing factors affecting detection: debris size and color

Debris size and color also had relatively large effect sizes and confidence, suggesting that shoreline debris counts for smaller, dull items likely result in underestimates compared to larger, brightly colored items. Our results suggest that a red playing card or blue bottle cap on a sandy shoreline located two meters from an observer will be detected close to 100 % of the time, and there is little variation around that average; whereas a brown bottle cap at the same distance from an observer will be detected less than half the time on average but the variability will range from ~15 to 75 %. In theory, our results could be used to apply a correction factor to debris estimates for surveys under similar conditions. However, without knowing the true relative frequency of debris colors and sizes, estimating the effect of variable detection rates on overall debris estimates poses a challenge.

While the size of debris that is encountered cannot be controlled, the size range that is counted as part of a protocol will inform the influence of detectability on count accuracy. We found that items equal to approximately 6 cm × 6 cm were reliably detected, but this threshold is notably larger than the target size range for macrodebris monitoring protocols. A common lower size limit for shoreline monitoring protocols is 2.5 cm, either as an absolute minimum (e.g. NOAA Marine Debris Monitoring and Assessment Project (Burgess et al., 2021), African Marine Litter Monitoring Method (Barnardo et al., 2020), Korea National Beach Litter Monitoring Program (Hong et al., 2014); this study) or as a cutoff when binning counts as in OSPAR's Guide to Monitoring Beach

Litter (Wenneker et al., 2010). However, even lower size cut-offs are also employed (e.g. 5 mm; Sustainable Coastlines, 2023), and some protocols have no lower size limit (Schuyler et al., 2018b). In these latter cases, underestimation and uncertainty will likely be amplified compared to our findings.

Two prior studies explored detection of items smaller than 2.5 cm (Lavers et al., 2016; Angelini et al., 2019) though did not test the effect of size on detection, both found that color was important. Lavers et al. (2016) reported variable detection probabilities over time for observers repeatedly examining 50 × 50 cm quadrats on sandy shorelines for variously colored plastic fragments (white, blue, black) and resin pellets (white, black) ranging in size from 2.5 to 60 mm. Raw detection probabilities were highly variable across each observer and plastic type ranging from 60 to 100 %. Angelini et al. (2019) tested detection of blue, clear, and white fragments between 1 and 2 m and found that white and clear plastics were under counted (50 % and 55 % detection, respectively) relative to blue plastic (95 %).

4.3. Putting factors affecting detection into context: shoreline characteristics

In our study, the effect of shoreline characteristics on detection was also supported. The presence of cobble or vegetation resulted in lower detection rates than for sand or other shoreline zones. These results are suggestive of at least two non-exclusive mechanisms by which debris items have lower detectability: 1) the level of contrast between the item and the substrate (marine debris items being camouflaged), and 2) visual obstruction (marine debris items being hidden behind features). Both explanations are supported by prior studies that examined detection of microdebris (Lavers et al., 2016, Angelini et al., 2019). Angelini et al. (2019) found that shell density and sand color influenced detection, similar to Lavers et al. (2016) who documented the effect of biological debris and coral rubble presence. They also identified the potential importance of false positives (e.g. mistaking a shell fragment for hard plastic), something we did not look at in our study, though this issue may be less prevalent for macrodebris (Lavers et al., 2016).

Most survey protocols in use today specify a transect extent that stops when vegetation or another barrier is encountered. However, backshore vegetation is a noted potential sink of debris. (Brennan et al., 2018; Olivelli et al., 2020). Given the variable presence of vegetation on shorelines, a relatively lower associated detection rate within vegetation compared to the rest of the beach, and the goal of data comparability over time, it may be beneficial to treat this area as a separate transect compartment if/when incorporated into future monitoring efforts.

4.4. Putting factors affecting detection into context: number of observers

Lastly, the variable effort trials showed an effect of the number of observers on gross detection rate as well as among-team variability. That is, teams of two or three consistently detected more debris than individuals surveying alone, and team surveys had lower inter-survey variability than surveys with one observer only. This could be due to an overlap in search area and shorter average distance from debris to an observeras the number of observers increases, thus increasing the “chances” for an item to be detected. It may also mitigate for differences in the proficiency of individual observers. Even when effort is controlled, some people are better at detecting debris than others (Baak et al., 2022). Furthermore, it may mitigate and/or prevent individual observer fatigue on overall debris detection rate (i.e. larger teams share the work load) (Lavers et al., 2016). Given intra- and inter-observer differences in detection, there is a potential interaction between this source of variability and resolution of underlying data trends (i.e., temporal shifts in debris load vs shifts in detection due to observer turnover), especially if the signal-to-noise ratio is low. Teams of two or three will buffer this effect. On the other hand, differences in debris counts introduced by varied survey team sizes might be tolerable and/or factored into analysis (e.g. Uhrin et al., 2020).

This study reflected local conditions in Washington State, USA – amount and type of debris, as well as shoreline characteristics. We kept the known debris quantity constant as a means of control, choosing 20 items as a manageable amount that was within the range of known local abundance, but actual debris concentrations vary (Serra-Gonçalves et al., 2019; Uhrin et al., 2022). The impact of debris concentration on detection is unknown, although it may factor into the potential for observer fatigue (Lavers et al., 2016), or perhaps higher detection of smaller items as observers bend over more frequently to assess and retrieve debris.

5. Conclusions

Through a series of field trials, we explored how characteristics of the protocol, the debris itself, and of the shoreline can influence the detection of marine debris during shoreline surveys. Debris detection rates varied according to distance from observer, number of observers, debris characteristics (size, color), and shoreline substrate. These results highlight considerations for the design of future shoreline surveys, as well as for analysis and interpretation of already extant data.

The benefits and approach to considering detection will depend on study goals. For study questions that do not require precise debris estimates or comparisons, our identified sources of bias may be of less consequence. For example, Murray et al. (2018) reported that the deposition of indicator debris items increased 10-fold following the Great Tōhoku Earthquake and Tsunami compared to baseline, a magnitude of change that would most likely be evident irrespective of potential differences in detection rates associated with survey protocols that differed before and after the event. In other cases, where smaller differences in amount or rates of change are of interest, uncertainty introduced by varied rates of detection could be an important consideration. That is, if factors influencing detection rates covary with spatial or temporal variables of interest, there is a risk of confounding detection rate with patterns in debris loads.

In addition to study and protocol design, our results raise considerations for combining and comparing datasets derived from different monitoring protocols. As monitoring efforts expand around the world, and given the identified need for regional to international indicators and metrics to track progress (e.g., United Nations Sustainable Development Goals, Sawarkar and Kodati, 2021, United Nations Environment Programme, 2021) integration among datasets should be approached cautiously. Bias introduced by different survey protocols has been a noted challenge to large scale debris assessments (Browne et al., 2015; Larsen Haarr et al., 2022; Fraisl et al., 2023). Our study showed that aspects of a protocol can potentially result in differing detection rates (e.g. in the presence or absence of a lower size limit, different search distances and numbers of observers) and in turn, different estimates of marine debris loads.

While statistical techniques can account for known biases in data, and the level of acceptable bias will vary according to the intended use, for marine debris datasets to be readily accessible and interpretable by a variety of stakeholders, mitigating sources of error will be beneficial and cost effective (Hardesty et al., 2017b; Uhrin et al., 2022). Best practices could include a combination of maximizing detection through protocols that optimize effort, identifying and promoting consistent detection probabilities through protocol standardization and training, adopting a minimum size threshold, and documenting shoreline and debris characteristics that influence detection, accounting for them in subsequent analyses.

Ethics and consent

The protocol for this study was reviewed by the University of Washington Human Subjects Division IRB, receiving activity approval under FW #00006878. Required documentation of consent was waived.

Staff participating in this project completed NOAA Contractor Sexual Assault and Sexual Harassment Prevention and Response Training on 10/2/2018.

Research permits were obtained and maintained from the National Park Service (OLYM-2018-SCI0058) and Washington State Parks (180503) for activities at Ocean City and Fort Worden. Although the Seattle Parks and Recreation (Carkeek Park) does not require a permit for citizen science activities, they were contacted and informed of this work.

CRediT authorship contribution statement

Hillary K. Burgess: Conceptualization, Data curation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Timothy T. Jones:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Jacqueline K. Lindsey:** Conceptualization, Writing – review & editing, Data curation, Methodology, Project administration. **Carlie E. Herring:** Conceptualization, Writing – review & editing. **Sherry M. Lippiatt:** Conceptualization, Writing – review & editing. **Julia K. Parrish:** Conceptualization, Methodology, Writing – review & editing. **Amy V. Uhrin:** Conceptualization, Funding acquisition, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: co-author A.V. Uhrin serves as a guest editor for the submission journal.

Data availability

Primary project data are available on the NOAA Marine Debris Clearinghouse. Analytical code and COASST data are available upon request.

Acknowledgements

Several undergraduate students and staff at the University of Washington were instrumental in data collection, including Abby Bratt, Ellie Davis, Allison DeKerlegand, Yurong He, Colin Piwtorak and Charlie Wright. Nir Barnea attended several field trials and took photos. The Olympic Coast National Marine Sanctuary, PTMSC, Olympic National Park and Washington State Parks graciously helped with participant recruitment and study locations. Comments provided by C. Wessel, S. Lundgren, N. Wallace, and two anonymous reviewers benefitted the manuscript. We would especially like to thank the many individuals who volunteered their time to help with this project. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

Funding information

This work was supported by the NOAA Marine Debris Program through the Joint Institute for the Study of the Atmosphere and Ocean hosted by the University of Washington (Project Number NA15OAR4320063). Indirect support was received from the NOAA Marine Debris Program (AVU), Genwest Systems Inc. (HKB), Lynker Technologies LLC and I.M. Systems Group (CEH and SML), the Joint Institute for the Study of the Atmosphere and Ocean (HKB, TTJ, JKL), and the Wakefield family (JKP).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2023.115905.

References

Angelini, Z., Kinner, N., Thibault, J., Ramsey, P., Fuld, K., 2019. Marine debris visual identification assessment. Mar. Pollut. Bull. 142, 69–75. https://doi.org/10.1016/j.marpolbul.2019.02.044.

Baak, J.E., Brown, Z.O., Provencher, J.F., Mallory, M.L., 2022. A rapid assessment technique for coastal plastic debris sampling: applications for remote regions and community science. Mar. Pollut. Bull. 178, 113641.

Barnardo, T., Ribbink, A.J., (Eds.)., 2020. African Marine Litter Monitoring Manual. African Marine Waste Network, Sustainable Seas Trust, Port Elizabeth, South Africa.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Blickley, L.C., Currie, J.J., Kaufman, G.D., 2016. Trends and drivers of debris accumulation on Maui shorelines: implications for local mitigation strategies. Mar. Pollut. Bull. 105, 292–298. https://doi.org/10.1016/j.marpolbul.2016.02.007.

Brennan, E., Wilcox, C., Hardesty, B.D., 2018. Connecting flux, deposition and resuspension in coastal debris surveys. Sci. Total Environ. 644, 1019–1026.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., Bolker, B.M., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 9 (2), 378–400.

Browne, M.A., Chapman, M.G., Thompson, R.C., Zettler, A.A., Jambeck, J., Mallos, N.J., 2015. Spatial and temporal patterns of stranded intertidal marine debris: is there a picture of global change? Environ. Sci. Technol. 49 (12), 7082–7094. https://doi.org/10.1021/es5060572.

Burgess, H.K., Herring, C.E., Lippiatt, S., Lowe, S., Uhrin, A.V., 2021. NOAA marine debris monitoring and assessment project shoreline survey guide. In: NOAA Technical Memorandum NOS OR&R, 56. https://doi.org/10.25923/g720-2n18.

Burnham, K.P., Anderson, D.R., Laake, J.L., 1980. Estimation of density from line transect sampling of biological populations. Wildl. Monogr. 1 (72), 3–202.

Cheshire, A.C., Adler, E., Barbiere, J., Cohen, Y., Evans, S., Jarayabhand, S., Westphalen, G., 2009. UNEP/IOC Guidelines on Survey and Monitoring of Marine Litter. UNEP Regional Seas Intergovernmental Oceanographic Commission.

Fraisl, D., See, L., Bowers, R., et al., 2023. The contributions of citizen science to SDG monitoring and reporting on marine plastics. Sustain. Sci. 18, 2629–2647. https://doi.org/10.1007/s11625-023-01402-4.

Galipaud, M., Gillingham, M.A., Dechaume-Moncharmont, F.X., 2017. A farewell to the sum of Akaike weights: the benefits of alternative metrics for variable importance estimations in model selection. Methods Ecol. Evol. 8 (12), 1668–1678.

Gennip, S.J.V., Dewitte, B., Garçon, V., et al., 2019. In search for the sources of plastic marine litter that contaminates the Easter Island ecoregion. Sci. Rep. 9, 19662 https://doi.org/10.1038/s41598-019-56012-x.

GESAMP, 2019. Guidelines or the monitoring and assessment of plastic litter and microplastics in the ocean. In: Kershaw, P.J., Turra, A., Galgani, F. (Eds.), IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP/ISA Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection. Rep. Stud. GESAMP, 99, p. 130.

Hanke, G., Galgani, F., Werner, S., Oosterbaan, L., Nilsson, P., Fleet, D., Liebezeit, G., 2013. Guidance on Monitoring of Marine Litter in European Seas: A Guidance Document Within the Common Implementation Strategy for the Marine Strategy Framework Directive.

Hardesty, B.D., Lawson, T.J., van der Velde, T., Lansdell, M., Wilcox, C., 2017a. Estimating quantities and sources of marine debris at a continental scale. Front. Ecol. Environ. 15 (1), 18–25.

Hardesty, B.D., Wilcox, C., Schuyler, Q., Lawson, T.J., Opie, K., 2017b. Developing a baseline estimate of amounts, types, sources and distribution of coastal litter - an analysis of US marine debris data (CSIRO: EP167399). In: Final Report to the NOAA Marine Debris Program in fulfillment of NOAA Award Number NA15NOS4630201. https://marinedebris.noaa.gov/sites/default/files/publicationsfiles/An_analysis_of_marine_debris_in_the_US_FINAL_REP.pdf.

Harris, L., Liboiron, M., Charron, L., Mather, C., 2020. Using citizen science to evaluate extended producer responsibility policy to reduce marine plastic debris shows no reduction in pollution levels. Mar. Policy 123. https://doi.org/10.1016/j.marpol.2020.104319.

Hidalgo-Ruz, V., Thiel, M., 2015. The contribution of citizen scientists to the monitoring of marine litter. In: Bergmann, M., Gutow, L., Klages, M. (Eds.), Marine Anthropogenic Litter. Springer, Cham. https://doi.org/10.1007/978-3-319-16510-3_16.

Hong, S., Lee, J., Kang, D., Choi, H.-W., Ko, S.-H., 2014. Quantities, composition, and sources of beach debris in Korea from the results of nationwide monitoring. Mar. Pollut. Bull. 84 (1–2), 27–34. https://doi.org/10.1016/j.marpolbul.2014.05.051.

Larsen Haarr, M., Falk-Andersson, J., Fabres, J., 2022. Global marine litter research 2015–2020: geographical and methodological trends. Sci. Total Environ. 820, 10. https://doi.org/10.1016/j.scitotenv.2022.153162.

Lavers, J.L., Oppel, S., Bond, A.L., 2016. Factors influencing the detection of beach plastic debris. Mar. Environ. Res. 119, 245–251.

Lippiatt, S., Opfer, S., Arthur, C., 2013. Marine debris monitoring and assessment. In: NOAA Technical Memorandum NOS-OR&R-46.
Mæland, C.E., Staupe-Delgado, R., 2020. Can the global problem of marine litter be considered a crisis? Risk Hazards Crisis Public Policy 11, 87–104. https://doi.org/10.1002/rhc3.12180.
McWilliams, M., Liboiron, M., Wiersma, Y., 2018. Rocky shoreline protocols miss microplastics in marine debris surveys (Fogo Island, Newfoundland and Labrador). Mar. Pollut. Bull. 129, 480–486. https://doi.org/10.1016/j.marpolbul.2017.10.018.
Murray, C., Maximenko, N., Lippiatt, S., 2018. The influx of marine debris from the 2011 Great Japan Tsunami to North American shorelines. Mar. Pollut. Bull. 132, 26–32.
Olivelli, A., Hardesty, B.D., Wilcox, C., 2020. Coastal margins and backshores represent a major sink for marine debris: insights from a continental-scale analysis. Environ. Res. Lett. 15, 074037 https://doi.org/10.1088/1748-9326/ab7836.
Parrish, J.K., Burgess, H.K., 2017. Coastal Observation and Seabird Survey Team Protocol. University of Washington, Seattle, WA.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Ribic, Christine A., Dixon, Trevor R., Vining, Ivan, 1992. Marine Debris Survey Manual. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service. NMS TR-108.
Ryan, P.G., Moore, C.J., van Franeker, J.A., Moloney, C., 2009. Monitoring the abundance of plastic debris in the marine environment. Philos. Trans. R. Soc. Lond. 364, 1999–2012.
Sawarkar, K., Kodati, M., 2021. Automated metadata harmonization using entity resolution and contextual embedding. In: Arai, K. (Ed.), Intelligent Computing, Lecture Notes in Networks and Systems. Springer, Cham. https://doi.org/10.1007/978-3-030-80126-7_10.
Schuyler, Q.A., Hardesty, B.D., Lawson, T.J., Opie, K., Wilcox, C., 2018a. Economic incentives reduce plastic inputs to the ocean. Mar. Policy 96, 250–255.
Schuyler, Q.A., Willis, K., Lawson, T.J., Mann, V., Wilcox, C., Hardesty, B.D., 2018b. Handbook of Survey Methodology – Plastics Leakage. CSIRO, Australia. ePublish. EP178700.
Serra-Gonçalves, C., Lavers, J.L., Bond, A.L., 2019. Global review of beach debris monitoring and future recommendations. Environ. Sci. Technol. 53 (21), 12158–12167.
Willis, K., Hardesty, B.D., Kriwoken, L., Wilcox, C., 2017. Differentiating littering, urban runoff and marine transport as sources of marine debris in coastal and estuarine environments. Sci. Rep. 7 (1), 44479.
Sustainable Coastlines (2023). Litter Intelligence Beach Monitoring. https://litterintelligence.org/about/beach monitoring. Accessed November 27, 2023.
Symonds, M.R., Moussalli, A., 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. 65 (1), 13–21.
Uhrin, A.V., Lippiatt, S., Herring, C.E., Dettloff, K., Bimrose, K., Butler-Minor, C., 2020. Temporal trends and potential drivers of stranded marine debris on beaches within two US National Marine Sanctuaries Using Citizen Science Data. Front. Environ. Sci. 8, 232. https://doi.org/10.3389/fenvs.2020.604927.
Uhrin, A.V., Hong, S., Burgess, H.K., Lim, S., Dettloff, K., 2022. Towards a North Pacific long term monitoring program for ocean plastic pollution: a systematic review and recommendations for shorelines. Environ. Pollut. 310, 119862 https://doi.org/10.1016/j.envpol.2022.119862.
United Nations Environment Programme, 2021. Understanding the State of the Ocean: A Global Manual on Measuring SDG 14.1.1, SDG 14.2.1 and SDG 14.5.1. Nairobi.
van der Velde, T., Milton, D.A., Lawson, T.J., Wilcox, C., Lansdell, M., Davis, G., et al., 2017. Comparison of marine debris data collected by researchers and citizen scientists: is citizen science data worth the effort? Biol. Conserv. 208, 127–138. https://doi.org/10.1016/j.biocon.2016.05.025.
Van Sebille, E., Aliani, S., Law, K.L., Maximenko, N., Alsina, J.M., Bagaev, A., Bergmann, M., Chapron, B., Chubarenko, I., Cózar, A., Delandmeter, P., 2020. The physical oceanography of the transport of floating marine debris. Environ. Res. Lett. 15 (2), 23003.
Wenneker, B., Oosterbaan, L., Intersessional Correspondence Group on Marine Litter (ICGML), 2010. Guideline for Monitoring Marine Litter on the Beaches in the OSPAR Maritime Area, Edition 1.0. OSPAR Commission, London, UK.