Developing a baseline estimate of amounts, types, sources and distribution of coastal litter – an analysis of US marine debris data

A final report for Ocean Conservancy and NOAA

Britta Denise Hardesty, Chris Wilcox, Qamar Schuyler, TJ Lawson and Kimberley Opie

Version 1.2

23 October 2017

CSIRO Oceans and Atmosphere

Citation

Hardesty, Britta Denise, Chris Wilcox, Qamar Schuyler, TJ Lawson and Kimberley Opie. (2017) Developing a baseline estimate of amounts, types, sources and distribution of coastal litter – an analysis of US marine debris data. Version 1.2. CSIRO: EP167399

Copyright

Acknowledgments

This work has been supported by The Ocean Conservancy, the National Oceanic and Atmospheric Administration of the United States of America and CSIRO Oceans and Atmosphere. We thank Louise Bell for assistance with graphics and Camille Couzi for formatting assistance. We also thank the many volunteers, citizen scientists, state, federal and government employees and interested members of the public who are committed to reduce litter, marine debris, and illegal dumping in the environment.

Table of Contents

FIGURES

TABLES

Executive Summary

Marine debris is a growing environmental issue. It is a local, national, regional and globally relevant topic. Within the United States, there are clean-up campaigns and repeated surveys aimed to quantify the amount of debris on the coastline, but previously, a synoptic summary of the amount of debris and the relative contribution of particular items had not been carried out. To address this knowledge gap, this project set out to estimate the amount of debris on the coastline, based on coastal clean-up data collected by the International Coastal Cleanup (ICC), as well as two different NOAA debris surveys.

Presently, there are numerous approaches to monitoring marine debris including coastal surveys, coastal clean-ups, surface trawl sampling, sediment core sampling and deep ocean sampling. Most sampling and recording of debris or litter takes place on land, where surveys are easier to carry out for a number of reasons. However, land-based surveys are conducted using numerous different approaches, with different objectives, aims, outcomes and inferences that can be made from the data. Given the increasing interest and concern about the amounts, types and sources from which debris, and plastic pollution in particular, are entering the ocean, there has been acknowledgement within the global community of a need for data harmonization and standardization of data collection approaches. Different survey types are suited for addressing different questions, and to be in a better position to not only estimate the quantities, types and sources of debris but to enact change to reduce the inputs (and their impacts in the marine environment), it is important to collect information in a way that will address the target questions.

This project had two specific objectives:

1) To determine the density and distribution of marine debris within the United States and
2) To compare and contrast different survey/debris collection methods that have been used in collecting and reporting debris data.

NOAA and Ocean Conservancy's respective programs have different goals and thus use different approaches to collect data on marine debris. In particular, NOAA implements a comprehensive sampling regime with a group of trained volunteers who collect trash and quantify materials gathered per unit area at a relatively small set of representative beaches at regular time intervals. In contrast, Ocean Conservancy's Cleanup is an annual citizen event, at which people with no formal training count individual items of trash they collect while they clean a stretch of shoreline over the course of a 1.5-2 hour community participation event at thousands of essentially randomly selected sites (since clean-ups are held where and whenever volunteers are willing to organize one) each autumn. The CSIRO method implemented in this project differs again, focusing on designed surveys conducted by trained professionals. While it is possible to use data from these various monitoring programs to understand debris baselines, drivers, and changes, combining them is a challenging problem. They each suffer from different sampling biases due to design and implementation differences. It is likely that they can be used in a complementary way, as illustrated here. When we compared sampling at the same sites and times using the methods, in some cases they give similar pictures. However, in other cases the underlying differences in the methods give conflicting conclusions. Moving forward, it will be important to consider the purpose of the investment these programs, and the relative weight among their value for engagement, attitude change, publicity, developing baselines, detecting changes, and understanding system dynamics.

The national and regional picture of debris loads that we built in this project can be used as a baseline for future comparison, and to understand both the general load and the distribution of specific items at regional and national scales. In the process of developing this picture we were able to identify a number of variables that could be used to remove sampling bias in the various datasets, improving their accuracy. Building on this, we extended the analysis originally proposed in the project to explore possible driving factors affecting debris loads. We found a wide range of factors make significant contributions to debris loads, including land use, land cover, urbanization, access, socio-economic levels, and drainage from nearby watersheds. These factors were important across the datasets, suggesting there are a consistent set of driving variables that determine the debris load at a site.

Exploring the patterns in distribution of types of items found at a site, we were confronted by the complexity of these patterns. Using datamining tools such as regression trees and cluster analysis, we elucidated some patterns in the data. However, a number of key issues make transforming patterns into decision-ready information a challenge. First, items can have multiple sources and multiple sources can contribute items to one site. For instance, illegal dumping, tourism, marine transport, and littering by commuters can all contribute items to a single site. Thus, while the presence of tires might indicate dumping, it is confounded by the high abundance of cigarette butts left by beach users and commuters. Hence, identifying patterns in the distribution of items at a site is complex. Two paths forward are clear. Question driven analyses can help focus the process, and will likely be able to deliver useful answers. Second, much can be gained by using single item analysis in concert across a number of items suspected to be from a shared source or of common relevance.

We make five specific recommendations based upon the analyses and field campaign undertaken. First, we suggest conducting a nationwide survey to provide a useful reference against which to compare trends, interventions, and geographic patterns. We suggest that the west coast surveys done as part of this study could be replicated on the Gulf and Atlantic coasts of the US at relatively minimal cost. Such a survey would be most usefully done on a periodic basis, perhaps every 5 to 10 years, to track major changes in debris with time. Second, we recommend continuing the volunteer clean-up efforts undertaken by ICC participants. This dataset has extensive coverage, and in spite of some challenges with data analyses from clean-up events, this data has also provided a rich source of information. Next, we note that a further expansion of the analysis of drivers would provide useful information for understanding both the drivers of debris loads on the US coast and inland waterways and the data, and for targeting interventions through clean-ups, outreach, incentives, and regulation. The ICC data in particular could support a national analysis of bag ban, container deposit and other interventions, evaluating both their effectiveness and cost.

Another key opportunity would be to link the available coastal data to information on debris loads and transport processes on land, building a better picture of the important processes and possible intervention points. In this project, we investigated the role of outflows from rivers and estuaries in driving local debris densities. There is clearly a relationship, however, from preliminary analyses the relationship has some complexities. Finally, designing a national monitoring system that provides high-quality data at a range of investment levels would mean NOAA could periodically put delivery of a national dataset out to tender, but with a clear picture of the likely person-hours required, the expected data structure and sampling design, and with a pre-existing analytical design and data management system. This would allow NOAA to establish a national baseline and implement periodic monitoring in a very cost-effective manner, and allow NOAA to deliver these outcomes in an inter-operable manner over time using the most cost-effective public or private sector providers.

1 Introduction

Marine debris or coastal litter is a burgeoning environmental issue with myriad sources and impacts. This environmental issue is garnering attention around the world, with the United Nations general assembly focusing on marine debris in 2015-2016. Furthermore, recent and upcoming changes in legislation have demonstrated significant changes in the public's relationship with plastic as plastic bag bans are enacted and microbead bans come into effect.

As evidence of the public's increasing interest in addressing and responding to the issue, In 2014 alone, more than half a million volunteers in 91 countries picked up more than sixteen million pounds of trash as part of the International Coastal Cleanup (ICC) efforts. This annual effort led by the Ocean Conservancy, has long been supported by and has been working hand in hand with NOAA's Marine Debris Program. The NOAA program, initiated in 2006, is authorized by the United States Congress and is the Federal government's main program aimed at addressing marine debris related issues, with a mission of investigating and preventing the negative impacts of marine debris.

Presently, there are numerous approaches to monitoring marine debris including coastal surveys, coastal clean-ups, surface trawl sampling, sediment core sampling, deep ocean sampling. Various survey types are suited for addressing different questions, but there has been acknowledgement within the global community of a need for data harmonization and standardization of data collection approaches if we are to be in a position to better estimate the amounts, sources and impacts of marine debris on people, communities, industry, tourism and wildlife.

This work delivers directly to NOAA's strategic goals to create healthy oceans and resilient coastal communities and economies. It further target's NOAA's objective to improve the coastal water quality that supports human health and coastal ecosystems services by providing an analysis of NOAA's monitoring data with an aim of determining the baseline estimate of marine litter.

This project consisted of two main objectives. Objective 1 was to develop a baseline estimate of the amounts, types and distribution of coastal litter, with a focus on US beaches. For this work, we focused our analyses on two types of coastal debris surveys, NOAA's standing stock and accumulation monitoring and coastal clean-ups as carried out under the International Coastal Cleanup (ICC) activities. The goal within this main objective was to address a suite of questions about the state of marine debris along US coasts and waterways including:

1) How much marine debris occurs on U.S. shores?
2) Are there specific littered items that are most (and least) abundant? Do these change locally or regionally?
3) Does the "diversity" (types and relative frequency) of marine debris vary spatially or temporally?
4) Where are the "hot spots" or regions where marine debris is most prevalent?
5) Do patterns of distribution and abundance of marine debris change over time (i.e., are there discernible temporal differences in characteristics of debris in regions with sufficient sampling)?

Objective 2 was to compare data collection methods to improve data collection protocols. In this part of the project, we compared NOAA, ICC and CSIRO-developed protocols to estimate, quantify and compare debris amounts and types. Comparing between survey methods provided an opportunity to evaluate the relative power of each method to uncover pattern and process in marine debris at local, regional and national scales. Based on analyses, we discuss the relative strengths and weaknesses of different sampling/clean-up methods, and we include recommendations to improve data collection protocols for NOAA and Ocean Conservancy going forward.

We compared and contrasted the NOAA protocols with the CSIRO method by collecting new data during a field effort to take place in July 2016. Due to logistical constraints, ICC clean-ups were not conducted so we cannot directly compare all three methods based upon data collected simultaneously. However, we do discuss clean-ups and statistically designed surveys with respect to data quality, concerns and constraints and extent of data. Based on experience and analyses, we provide a suite of recommendations to improve statistical power, reduce data collection effort and associated costs, improve scientific inference, and maximize scientific and policy insights related to marine debris monitoring and clean-up efforts going forward. Using this approach, we:

1. Attempted to develop a conversion factor to allow Cleanup data from more than 2,500 U.S. Cleanup sites to be used to calculate indices that are consistent with NOAA's sampling regime;
2. Determined if we can develop a conversion factor between CSIRO and NOAA methods so that statistical inferences can be drawn beyond the NOAA sampling sites; and
3. Undertook a power analysis of the NOAA monitoring protocol to determine the minimum sampling intensity required to estimate temporal changes in debris at the 95 percent confidence level.

To better address the marine debris issue, we clearly need more and better data on the sources, fate and impacts of marine debris to enable the design of solutions to match the scale of the problem. To date, however, a thorough analysis of both NOAA's data and of the data collected by OC's volunteers has been lacking.

As we highlight in this report, the challenges of analysis of the three data sets used here points to the need for harmonization or standardization of methodologies. Of course, data are collected for a variety of purposes and different types of data have different utility, are aimed at different target audiences and can yield insights to different components of the problem. Nonetheless, clearly stating program goals and objectives and designing surveys to address questions about the sources, identify hotpots or accumulation sites, and to determine how effective policies and local efforts are reducing litter inputs to the coastal and marine environment will be likely to result in the highest level of success in answering such questions.

The first section of the report (Chapter 2) focuses on the primary objective of presenting analysis of the amounts, types and distribution of debris within the US. Within this objective, we address several questions that were highlighted as priority by NOAA and the Ocean Conservancy. We present the major findings for each question, with details regarding the analytical approach taken to address those questions presented in appendices. The next part of the report (Chapter 3) focuses on the second objective of comparing survey types, challenges, and recommendations for data harmonization. We then provide a conclusions and recommendation section (Chapter 4), followed by an analytical appendix in which we provide detail regarding data, site characterization, and analytical details (Section 6).

2 Density and distribution of debris within the United States

2.1 How much debris occurs on U.S. shores?

We analysed data from four different sources to estimate the amount of debris on the shore of the United States. These included NOAA's accumulation and standing stock survey data, along with clean-up data from the International Coastal Cleanup, and a survey of the west coast of the continental United States, using methods developed by CSIRO. On average, there were 16.5 (95% confidence interval: 9.8, 23.1) items per meter of beach based on the NOAA accumulation data, 0.2 (95% confidence interval: 0.20 0.24) items per meter based on the NOAA standing stock data, 1.21 (95% confidence interval: 1.12, 1.30) items based on the International Coastal Cleanup, and 12.1 (95% confidence interval: 7.1, 17.1) items per meter based on the CSIRO data.

These numbers differ widely for several reasons. First, the data sets cover different areas, with varying amounts of overlap. The ICC data is most widely distributed, covering much of the coast of the continental US (Figure 6.1.2.a). The two NOAA datasets are more limited in scope (Figure 6.1.3.a, Figure 6.1.4.a). The CSIRO dataset is further limited, covering only the US west coast (Figure 6.1.5.a). Second, the surveys differ in their sampling strategy. The CSIRO survey incorporates any items visible at the surface, from standing height. This generally equates to a lower detection limit on the order of 1 mm. The NOAA surveys specify a lower size limit for inclusion of 25 mm. The ICC data has no lower detection limit, but in practice appears to be inversely related to the number of participants on clean-up events. Thus, the more people participating, the higher the count at a site, likely in part due to a reduction in the minimum size of particles people collect. Third, the methods differ in their control of searching behaviour. At one extreme, the CSIRO method controls the search effort per unit area per unit time, which is always held at 1 person per square meter at a walking pace. At the other, the ICC method records only the length of the beach, with no control on total area searched, search pattern, or the number people involved. The NOAA methods are somewhere in between, with some control on sampling effort and sampling pattern, but likely some double searching due to the search pattern used. Fourth, two of the methods are based on removing items repeatedly at sites (NOAA accumulation and ICC), while the other two survey but do not remove items (CSIRO and NOAA standing stock).

Given these survey differences, it is not possible to combine the data sets directly to estimate the amount of debris on the coastline. It is still possible to use the data from each survey method to compare between sites or over time, or to investigate factors that are related to higher or lower loads at survey sites. Some progress can be made toward unifying the data sets, for instance the CSIRO method records the size distribution of items. This can be used to subset the data to match NOAA's larger size categories. Similarly NOAA and ICC datasets both include types of items, which could be used to cross-reference the data between datasets. However, the fundamental differences in control of sampling effort remain, and thus direct amalgamation of the datasets is likely to remain difficult. Controlling for the removal of items in two of the methods is also difficult, as removing that effect would require estimating the time between surveys required for the site to reach equilibrium between deposition and resuspension of items. When we examined this relationship it appears to be shorter than the survey interval, and may even vary by site.

An alternative approach is to use each dataset separately to estimate the amount of debris on the coastline, after standardizing the data for effects of area sampled, sampling effort, and other factors that affect the density observed during the surveys. We used generalized additive models to do this task, as described in sections 6.4 and 6.5 below. These models can be used for exploring driving variables for hotspots, as described in section 2.2, but here we focus on using a simplified version focused on just controlling for sampling effort. For more details on the models and parameter estimates, see section 6.3, 6.4. If we standardize for the time since the last clean-up to be 1 year and the total number of people on a survey to be 29 (the median value), the ICC data gives an average of 1.22 items per meter of coast (95% confidence interval: 1.129564, 1.317660). Standardizing the NOAA accumulation data for the number of participants, the time since the last survey, and the organization involved, gives a density of 1.49 items per meter of coast (95% confidence interval: 0.62, 2.37). Standardizing the NOAA Standing Stock surveys for the number of participants gives an average density of 0.13 items per meter (95% confidence interval: 0.07, 0.18). The CSIRO data is largely standardized, as there are a constant number of participants on a survey, a constant survey width, the samples were taken at a single time by one group, and there is no removal of items for which to adjust. Thus, the density from the CSIRO survey method is 12.1 items per meter of coast.

Adopting NOAA's estimate of 95,471 miles for the total length of the coastline of the US at the tidal line (www.oceanservice.noaa.gov/facts/dshorelength.html), we have a total coastal length of 153,649,118 meters. The linear concentrations of debris yield estimates of the total number of items on the US coast, ranging from 19,974,385 items (based on NOAA Standing Stock data) to 1,859,154,328 items based on the CSIRO data. It is important to consider the biases involved in the various sampling methods in evaluating these numbers. First, the NOAA Standing Stock data likely underestimates the amount of debris, due to a mixture of some issues with organizations conducting the sampling which we were not able to address (see Figure 6.4.2.b and Figure 6.4.2.c and associated text; e.g. number of people participating in surveys is not always reported) and the survey guidelines which specify sampling only items larger than 2.5 cm in diameter. Similarly, interpreting the upper bound based on the CSIRO survey method, it is important to keep in mind that this includes items down to rough 1mm in diameter, but is only based on samples from the continental US west coast.

The US coast adjoins 4 major ocean basins, the Atlantic with 32% of the total coast, the Gulf with 19%, the Pacific with 46%, and the Arctic with 3%. The datasets we examined vary substantially in their coverage of these coastal regions. The Pacific coast is included in all of the data sets. The Atlantic is included in the ICC data, and to a limited extent both NOAA datasets. The Gulf coast is represented only by the ICC data. There are no samples on the Arctic coast. Taking this into consideration, the load estimates are most reliable for the Pacific, which forms roughly half of the US coastline. Encouragingly, the Arctic, which is relatively little of the US coast makes only a small contribution to the estimate. Using the standardized concentration from the ICC data, one can make reasonably reliable estimates for the Gulf and Atlantic coasts. Importantly though, the ICC method has an unknown lower bound in particle size, and has only recently included the small fragments which are generally by far the most abundant, so the national estimate of 187,451,924 items on the coastline from the ICC data should be considered a lower bound.

2.2 Where are the "hot spots" or regions where marine debris is most prevalent?

We tackled the problem of hotspots in the marine debris data at two different scales, based on the data we analysed. Identifying hotspots requires a reasonably even distribution of samples over the region where the analysis is to be made. The ICC data provides enough coverage to tackle the question of hotspots at a national scale, although with low resolution in some parts of the interior of the country. The CSIRO data covers the west coast of the continental US. The NOAA accumulation data provides similar coverage, although more limited in the southern portion of the west coast. The NOAA standing stock data is widely distributed, but covers relatively few locations and thus is difficult to use in this spatial context.

Hotspots can be addressed at three levels analytically, in something of a step-wise manner. First, one can ask where the raw densities are particularly high or low. This approach is straightforward, although it suffers from a lack of standardization. Thus, surveys with more effort may yield higher debris densities due to higher search efficiency or bias in sampling, and not necessarily a higher plastic load. The next level of analysis is to use a statistical model that takes account of sampling biases, effort, and other factors that can affect the sampling process. We estimated the spatial distribution of debris in this context by also including a spatial component in these models, either a line feature to represent the US west coast or a two dimensional plane to represent the continental US. The next step in the analysis is to include additional factors that might be driving hotspots, such as the presence of river mouths or local population density. These factors explain some of the spatial pattern in the distribution of the debris, and by incorporating their influence in the model, we can evaluate the remaining spatial pattern which is unexplained by either sampling error or other driving variables. This spatial pattern can help identify areas of uncertainty and suggest possible additional variables that could be important to include.

The detailed models used to address hotspots are presented below in section 6.4. Here we primarily focus on the spatial pattern in the data at the three levels of analysis, focusing on identifying areas with particularly high debris concentrations. We first present the results for the US West Coast, and subsequently for the national scale. Figure 2.2.a illustrates the spatial pattern in the raw debris density data for the west coast, based on the NOAA accumulation dataset, and the estimated distribution of debris along the west coast based on these samples, after correcting for sampling bias, and the remaining spatial pattern after fitting the best overall model (Table 6.4.1.c). The sampling bias variable in this model are the number of people and days since last survey, while the remaining spatial pattern model was also standardised for state, year, land use, watershed area, month, the distance to the nearest river, rail, and road, as well as the population, poverty fraction, housing, and the number of roads within 50km. It is interesting to note that there are both sampling sites with consistently high and consistently low densities, for instance see the green and red strips of dots at sites 13 and 17, both along the northern Washington coast. In contrast there are also sites that are highly variable, such as site 19 in the same area. Interestingly, there is also no clear strong pattern of the influence of major urban areas visible in the raw data, northern Washington with relatively low populations has similar values to areas near San Francisco and Los Angeles, with very high populations.

After standardizing the data for factors that can affect debris estimates such as sampling effort, area, state and survey date , the spatial pattern in the data is much more reflective of what one would assume drives debris loads. The coastal area near San Francisco, a major population center, has relatively high debris loads. It is important to note, that although the remote portions of central California are shown as having relatively high loads, this is a function of having only one survey point in Southern California, and thus interpolating the pattern between that site and northern California. Interestingly, some remote sites still have relatively high levels, such as on the northern Washington coast and near sampling site 34 in southern Oregon and 36 in northern California.

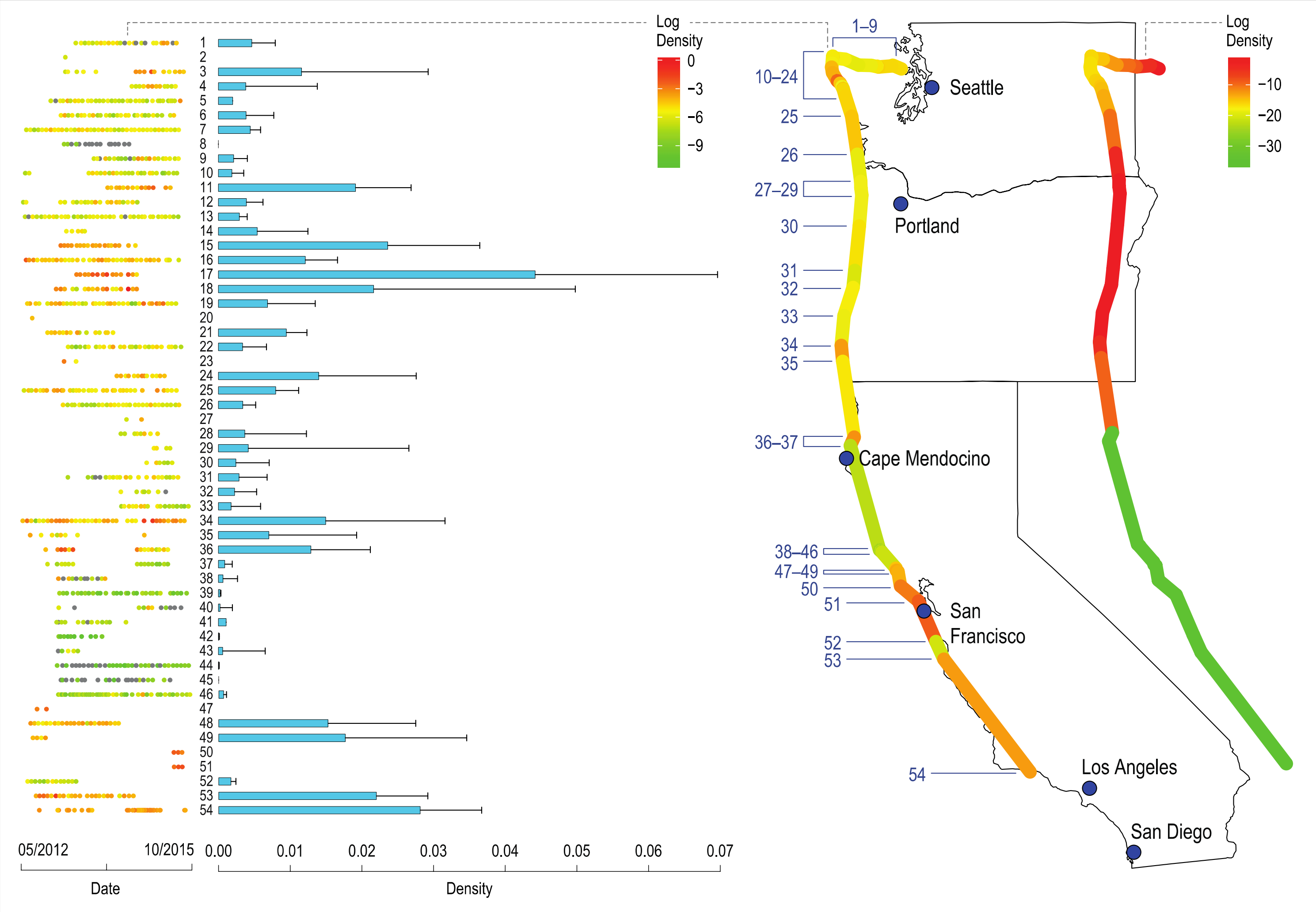


Figure 2.2.a Debris density from NOAA Accumulation data (2012-2015, west coast of USA).

Green values represent lower debris levels, red indicates higher debris levels. The series of points on the left side of the plot are raw debris density data from each survey. The points are ordered by year from oldest surveys on the left to those that have taken place most recently (left-right). Dots in grey indicate surveys with zero debris found. Note that the debris density values are on a log scale. Bar plots next to each set of dots indicate the median density (not on the log scale) and 95% confidence interval around the median for each survey site. The ribbon plot along the coastline is the spatial pattern in these data, after correction for sampling error using the statistical model described in Table 6.4.1.cb, to account for sampling effort. See Table 6.4.1c for data associated with ribbon plot on right hand side of figure. The ribbon plot to the east of the coastline is the spatial pattern in the data after accounting for local population density, access, input from nearby watersheds, and a variety of other factors that drive debris loads. This ribbon represents the spatial pattern in the unexplained variation in the best fitting statistical model, as described in Section 6.4.

As described in section 6.4, there are a number of important driving variables that affect the debris load at a site in the NOAA accumulation data, including level of urbanization, land use, vegetation type, access, population density, socio-economic status, and inputs from local watersheds. If we include these in the model, allowing us to control for their influence, we can examine the remaining spatial pattern in the data (Figure 2.2.a). This points to a clear pattern of an area of high debris, unexplained by the sampling bias or driving variables included in the model. This area is in the Cape Mendocino region of the west coast. In contrast, the area from just south of the California border southward is lower than expected in terms of debris densities. The California Current is the dominant ocean current system in this region, and moves north to south along the US west coast. Interestingly, the area that is lower than expected has a slightly south-westerly orientation, while the region with a higher load has a north-westerly orientation. Given the strong component of northerly winds on the US west coast, these spatial differences could be due to slightly differing levels of onshore transport driven by the interaction between coastal orientation and wind direction.

Examining the spatial pattern in the data collected according to the CSIRO transect method during July 2016 (Figure 2.2.b, Table 6.4.4.a), one can see that there is significant variation among the raw debris densities at the site level. Again, there does not appear to be a clear relationship between the raw debris densities and heavily urbanized areas or major estuarine/river systems.

However, after standardization for sampling biases (in this case, substrate, backshore, and aspect), some patterns in the interpolated distribution on the west coast emerge. Coastal areas near the major urban centers, particularly in Los Angeles and the Strait of Juan de Fuca are significantly high in debris. In these two areas in particular, predicted debris densities were so high that we had to plot them on a separate scale. Although it is not unexpected for sites in the Strait of Juan de Fuca to be high, given that the populations of both Seattle and Vancouver would be included in the larger watershed, the extremely high levels seen in the model results may also be caused by an anomaly in the data. One of the significant factors in the CSIRO model is the aspect of the beach. The northernmost site, adjacent to the Strait of Juan de Fuca, happens to be the only northern facing site in the data set. When we standardised the data for the site-level variables, we choose the most common factor and set all levels at that factor. The model thus predicts what the debris levels at the site would be if it were a west-facing beach.

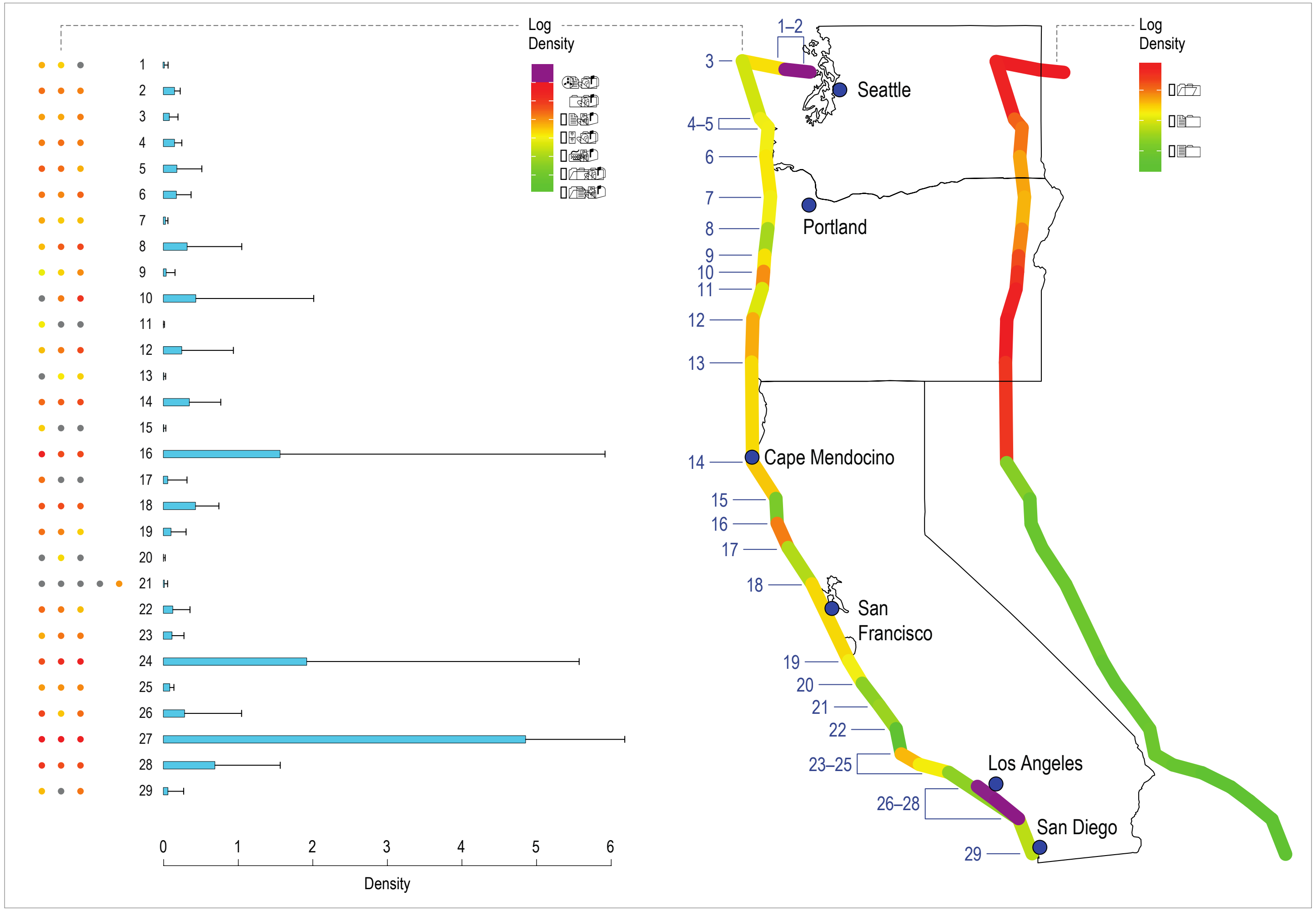


Figure 2.2.b Debris density ribbon plot from CSIRO data (2016, west coast of USA).

Green values represent lower debris levels, red indicates higher debris levels. The series of points on the left side of the plot are raw debris density data from each survey location. Dots in grey indicate surveys with zero debris found. The individual surveys at a site are offset horizontally to allow visualization of the variation in the samples. Note that the debris density values are on a log scale. Bar plots next to each set of dots indicate the mean density (not on the log scale) and 95% confidence interval around the median for each survey site. The ribbon plot along the coastline is the spatial pattern in these data, after correction for sampling error using the statistical model described in Section 6.4, to account for sampling effort, search time, coastal slope, substrate, backshore vegetation and other relevant variables. The purple segments are where predictions are significantly higher than the scale. The ribbon plot to the east of the coastline is the remaining spatial variation not explained by environmental variables.

After incorporating important driving variables for debris loads into the model (state, distance to nearest road and river, roads within 50km, watershed population and area, housing, work fraction, unemployed fraction, and median age within 50 km, and poverty fraction within 10 km), along with variables accounting for sampling bias (see above, and Table 6.4.4.a), we find a spatial pattern in the data (Figure 2.2.b) that is similar to that in the NOAA accumulation data (Figure 2.2.a). Again the region north of Cape Mendocino tends to have relatively high debris loads, which were not captured by either the variables for sampling bias or those for factors driving load. Given the consistency of this pattern, despite major differences in the two datasets and sampling protocols, it does suggest that there is an underlying mechanism creating the pattern. One clear candidate is marine transport, as discussed above.

Turning to the national picture, based on a similar staged analysis of the ICC data, we can examine the spatial patterns in the ICC data at three levels: using the raw concentrations observed during the clean-ups, using those raw values, standardized for sampling bias due to sampling effort, and looking at the remaining spatial pattern after accounting for sampling bias and other factors such as local population density that might drive debris concentrations at a site. Based on the raw concentrations, Texas stands out as having particularly high debris loads (Figure 2.2.c, Table 6.4.3.c). Urbanized states along the central eastern seaboard also had relatively high debris loads. Several states on the Gulf coast and Illinois also had relatively high loads. In some cases, such as Illinois, this is driven by the coastal portion of the state (on the shore of Lake Michigan) also being heavily populated (by the Chicago metro area). In other cases, there is a significant contribution from inland waterways and lakes. This effect can be seen in the analysis of specific items, such as fishing waste, presented in Section 2.3. Based on the raw data, the west coast of the US appears to have relatively low densities of debris (Figure 2.2.c).

Using a statistical model to standardize the raw data for effort and other variables that could introduce sampling bias, we find that the map of high and low debris load locations shifts significantly (Figure 2.2.d). Texas is still estimated to have relatively high loads. There is a consistent pattern of relatively high loads on the Gulf coast. The heavily populated central Atlantic coast states have higher loads after controlling for sampling effort. California also emerges as having relatively high loads. In considering these patterns, it is important to be conscious of the density of sample sites in the dataset. The interior of the US has relatively sparse sampling, and thus estimates in those regions should be treated with some caution (Figure 6.1.2.a). Extending the statistical model to take account of urbanization, access, population density, land use, vegetation type, and other variables driving debris loads, in addition to the sampling bias variables, we find a slightly different spatial pattern (Figure 2.2.e). Arkansas emerges as having a relatively high load at the sample sites, however, based on the very low number of sampling sites (Figure 6.1.2.a), this is likely to be an unreliable estimate. Debris densities in California and the states along central Atlantic coast now appear to be captured by the parameters for population, access and other driving variables, leaving little pattern in the spatial surface. Texas again has significantly higher debris loads than other states.

This analysis suggests that most debris on California and central Atlantic coasts can be explained by variables such as population, suggesting they are driven by domestic sources. However, Washington and Texas have relatively high loads, even after accounting for local factors. This points to an interesting pattern, there are particularly high loads at several sites in Texas near the border with Mexico. However, there is no similar pattern at sites in southern California near the Mexican border. This may be due to an interaction between the directions of currents in these two regions, with a coastal current in the Gulf of Mexico moving material from the US Gulf coast south-westerly along the coast. This could lead to concentration of debris at the southern portion of the Texas coastline. Similarly, Washington is subject to both a current moving southward, potentially bringing material from the northern Pacific Ocean and drainage through the Strait of Juan de Fuca, which drains a major a major urban area in Canada, not captured in our model. Again, as with the hypothesized pattern on the US west coast as noted in the NOAA and CSIRO data, this transport explanation could be a focal point for further investigation.

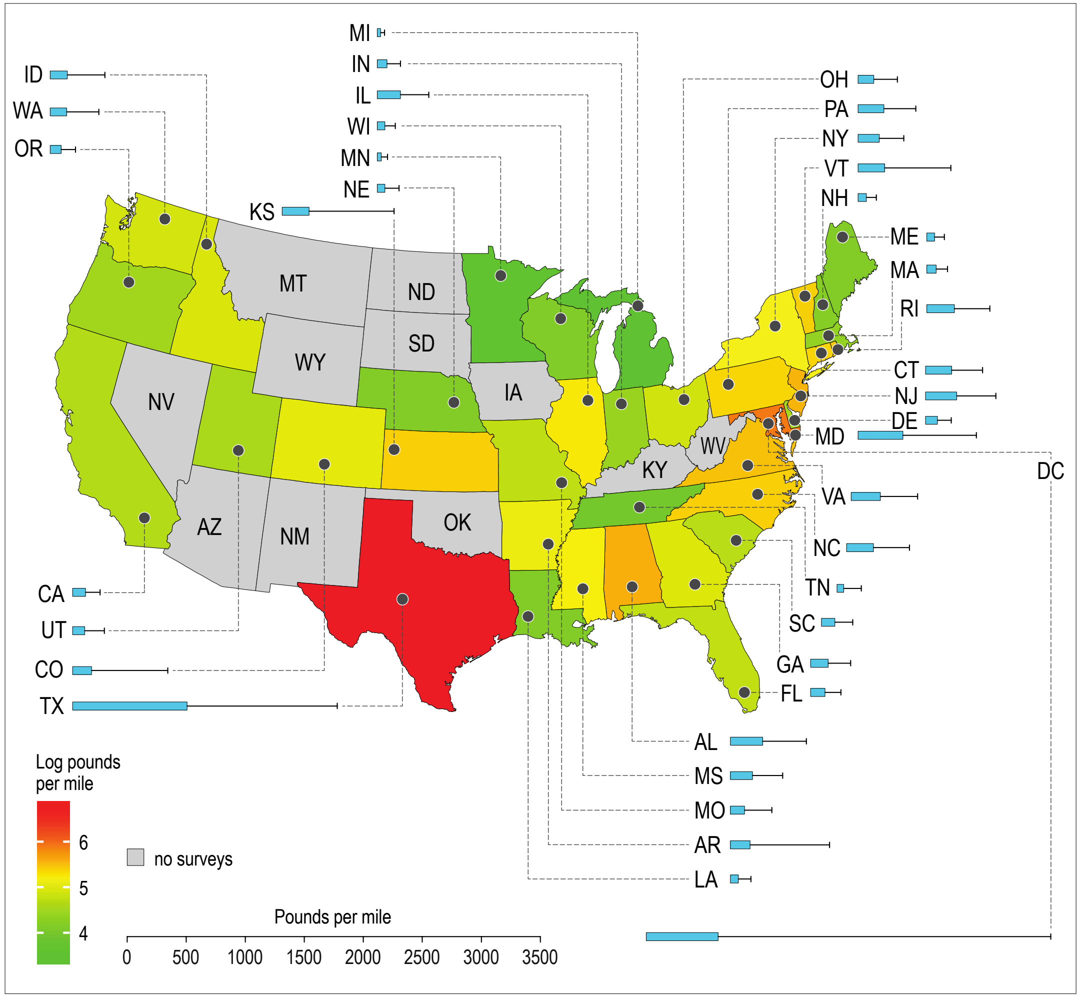


Figure 2.2.c Median debris density (pounds per mile) for raw data from ICC surveys, on a per-state basis.
Green values represent lower debris levels, red indicates higher debris levels. Note that the values are on a log scale. Grey states are those for which we had fewer than 5 survey locations. Bars to the side of each state show the median value (in blue) and the 95% confidence interval of the median (black lines). Bars are not on a logged scale.

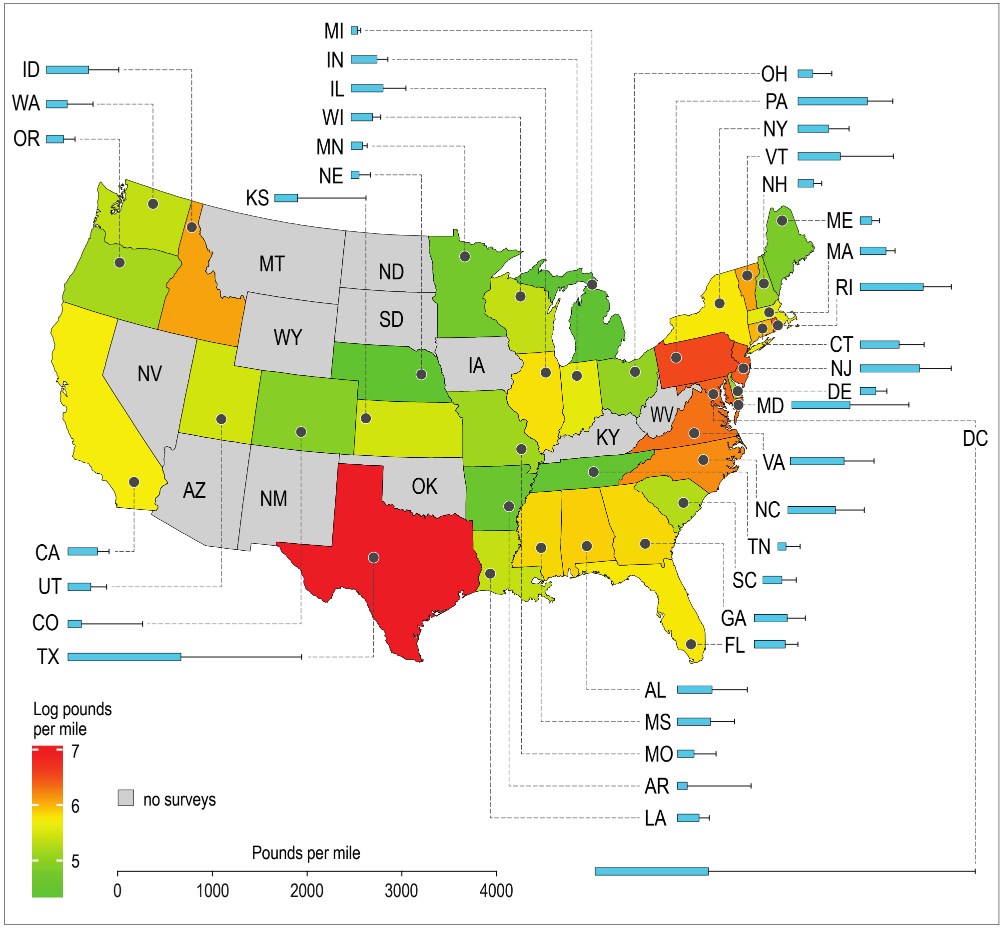


Figure 2.2.d Debris levels for ICC GAM site-level modelling, on a state-by state basis.
This model represents predicted debris levels after correction for sampling error using the statistical model described in Section 6.4, to account for sampling effort and other relevant variables. Green values represent lower debris levels, red indicates higher debris levels. Note that the values are on a log scale. Grey states are those for which we had fewer than 5 survey locations. Bars to the side of each state show the median value (in blue) and the 95% confidence interval of the median predicted values (black lines). Bars are not on a logged scale (black lines).

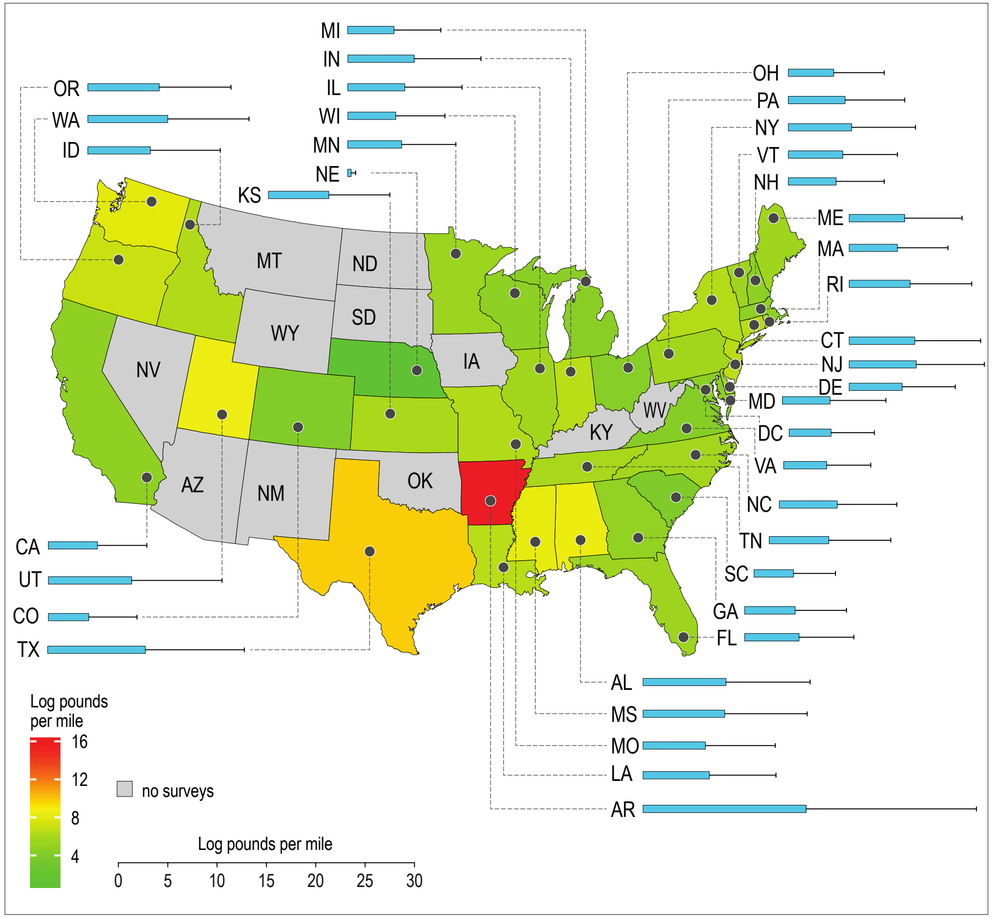
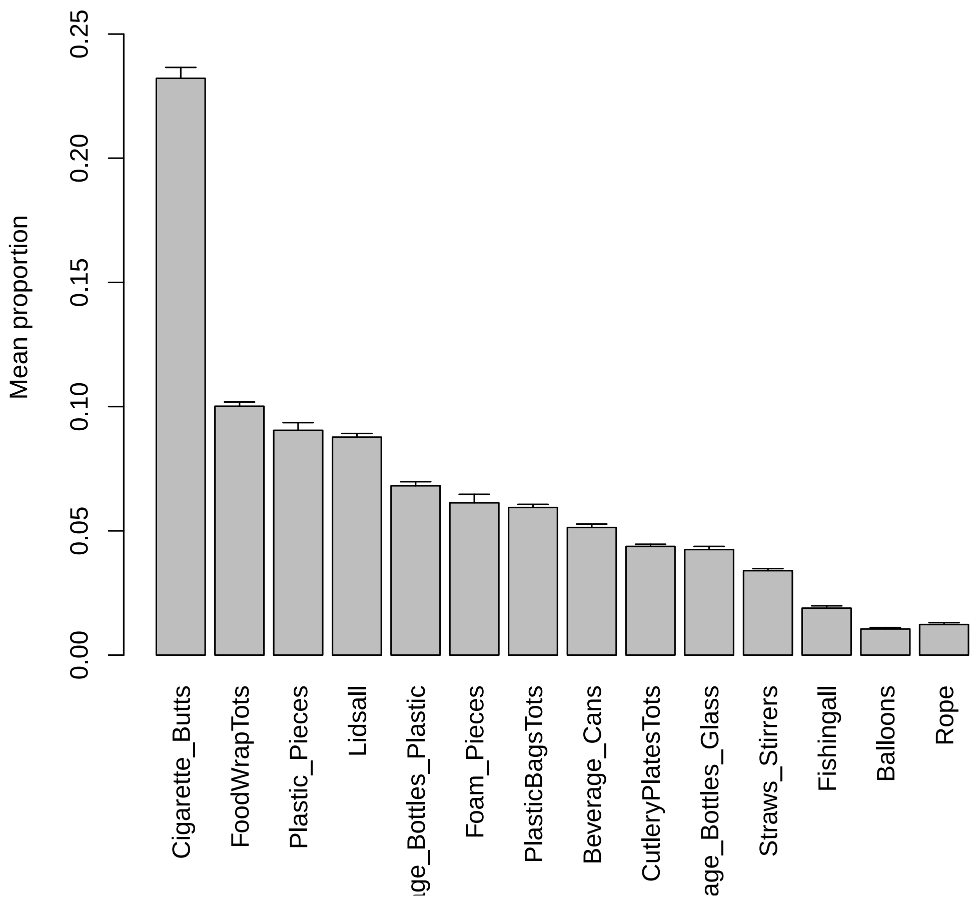


Figure 2.2.e Debris levels for ICC GAM spatial smooth, on a state-by state basis.

This model represents only the spatial pattern in the data, after accounting for local population density, access, input from nearby watersheds, and a variety of other factors that drive debris loads. This ribbon represents the spatial pattern in the unexplained variation in the best fitting statistical model, as described in Section 6.4. Green values represent lower debris levels, red indicates higher debris levels. Note that the values are on a log scale. Grey states had fewer than 5 survey locations. Bars to the side of each state show the median value (in blue) and the 95% confidence interval of the median (black lines).Bars are on logged scale.

2.3 Are there specific littered items that are most (and least) abundant? Do these change locally or regionally?

For both the NOAA and ICC data, we calculated the mean proportion of the top 10 most common items, plus items that pose a high threat to wildlife; balloons and fishing items (Figure 2.3.a). Cigarette butts, food wrap, plastic beverage bottles, and lids are all common items in both datasets. However, there are distinct differences in relative abundance, with cigarette butts reaching nearly 25% of all items in the ICC data, while they are only 6% in the NOAA data. This is due to the fact that the most abundant items in the NOAA data set are plastic fragments (hard plastic, filmed plastic, and foamed plastic), as well as plastic rope. Together these categories make up about 50% of the NOAA data. It is difficult to isolate a reason for this difference, as there are a number things that differ across the data sets including site, sampling protocol, and survey effort. However it is likely that ICC clean-up volunteers are more likely to pick up large items as opposed to smaller fragments which are collected during all NOAA surveys. Conversely, cigarette butts may be smaller than 2.5cm, which is the size limit for NOAA surveys. However, these items are well publicized in ICC literature as being the top item found in beach surveys, and they are very recognizable as litter in the United States due to litter education campaigns, so they may in fact be a target item for clean-up volunteers.



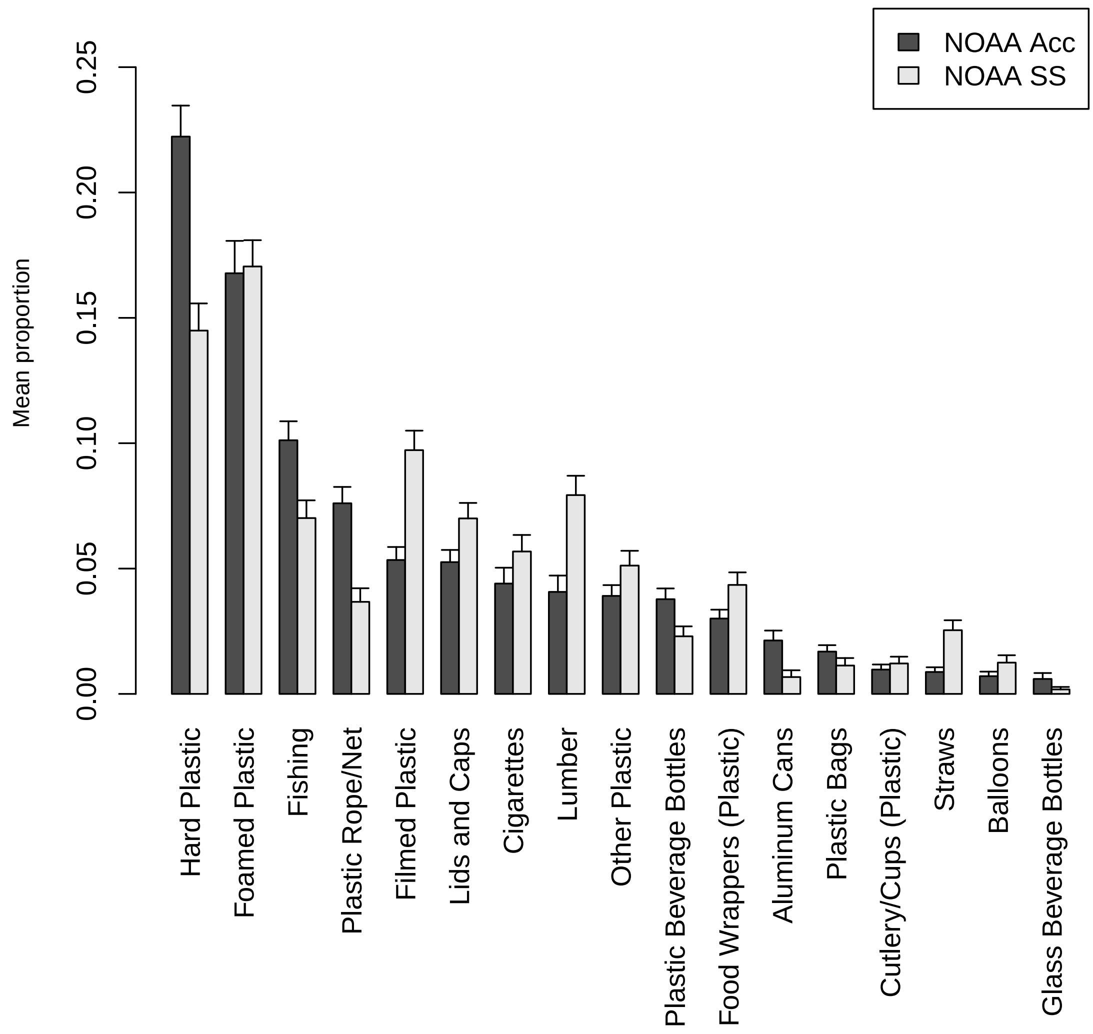


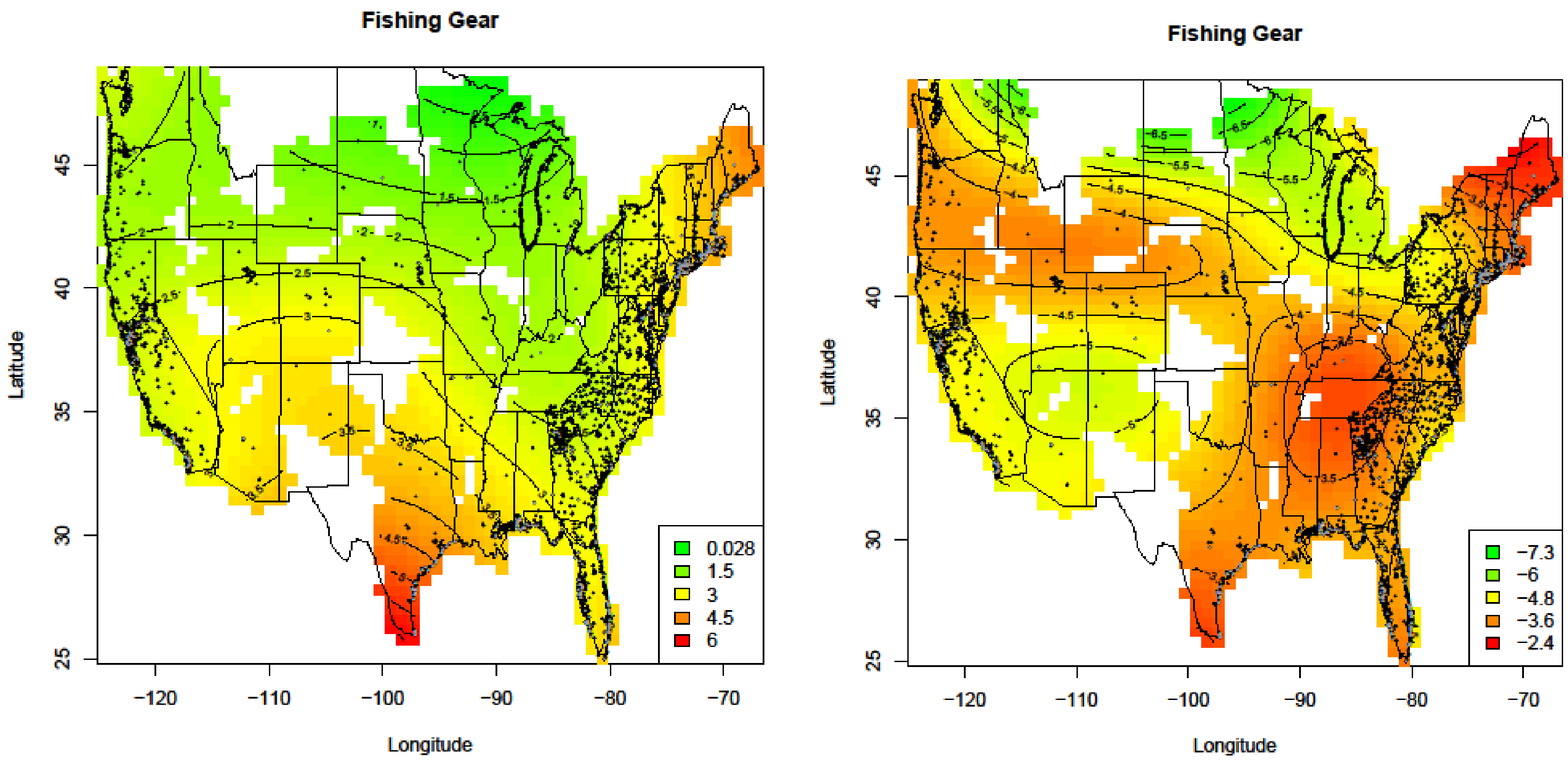
Figure 2.3.a The mean proportion of commonly found items from coastal surveys. Panel a) ICC coastal cleanup data. Error bars show the standard error of the estimates. Panel b) NOAA standing stock and accumulation survey data.

We investigated the distribution of a number of common items across sites at the nation scale, using the ICC data. In this analysis we standardized the data to control for sampling bias, but did not incorporate any additional variables as we wanted to allow the spatial surface in the models to retain all of the site level variation in the data (See Section 6.4 for detail). A key point to consider in evaluating these estimated surfaces is that they are not reliable in areas where there is little data. The underlying reason for this is that the models attempt to match the pattern of change in the data across the landscape. However, in areas with no data there is nothing to constrain their shape and they tend to carry over curvature imposed by nearby areas with larger amounts of data. Even in areas where there are data, but the points are sparse, nearby areas with large amounts of data will influence the shape of the spatial surface over the sparse areas. An example in the following plots is the effect of the very dense sampling in Texas and southern California, which has a strong effect on the estimates in the sparsely sampled interior area of the US in between these regions.

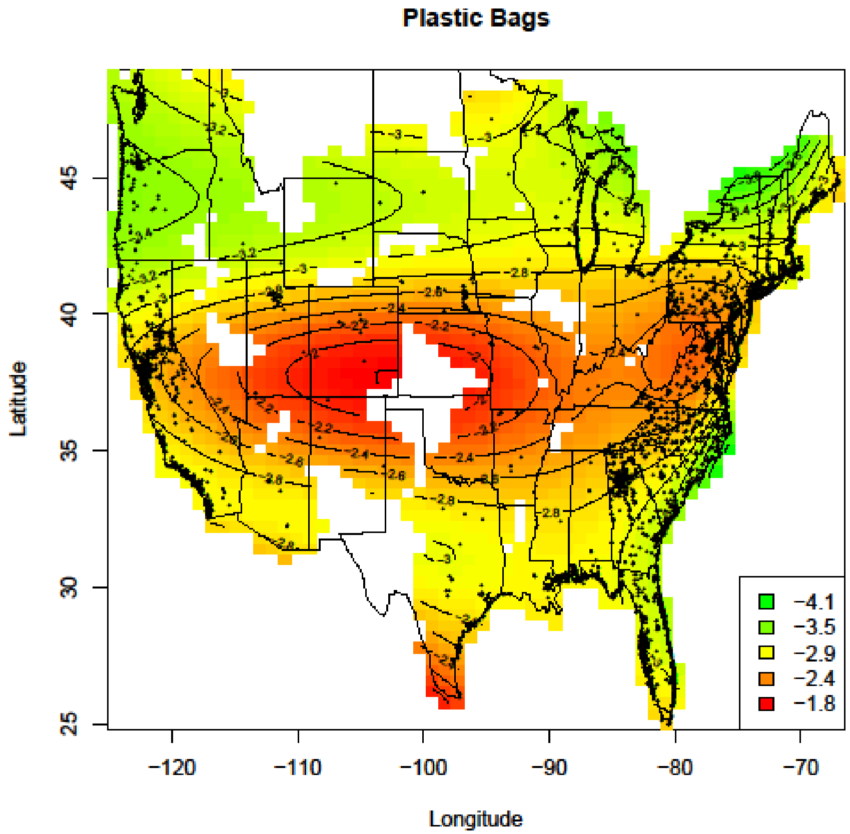
We modelled the abundance of 5 different types of items: fishing gear, plastic bags, balloons, plastic beverage bottles, and cigarette butts. These specific items were chosen as a subset of them (fishing gear, bags, and balloons) came out as posing the highest threat to wildlife in a recent analysis (Wilcox et al. 2016). Several of them are among the most common items found in clean-up efforts generally (bags, bottles, and cigarette butts). Two of the items have been the focus of significant industry engagement (bottles and bags), including in some states economic incentives to reduce loss into the environment (bottles).

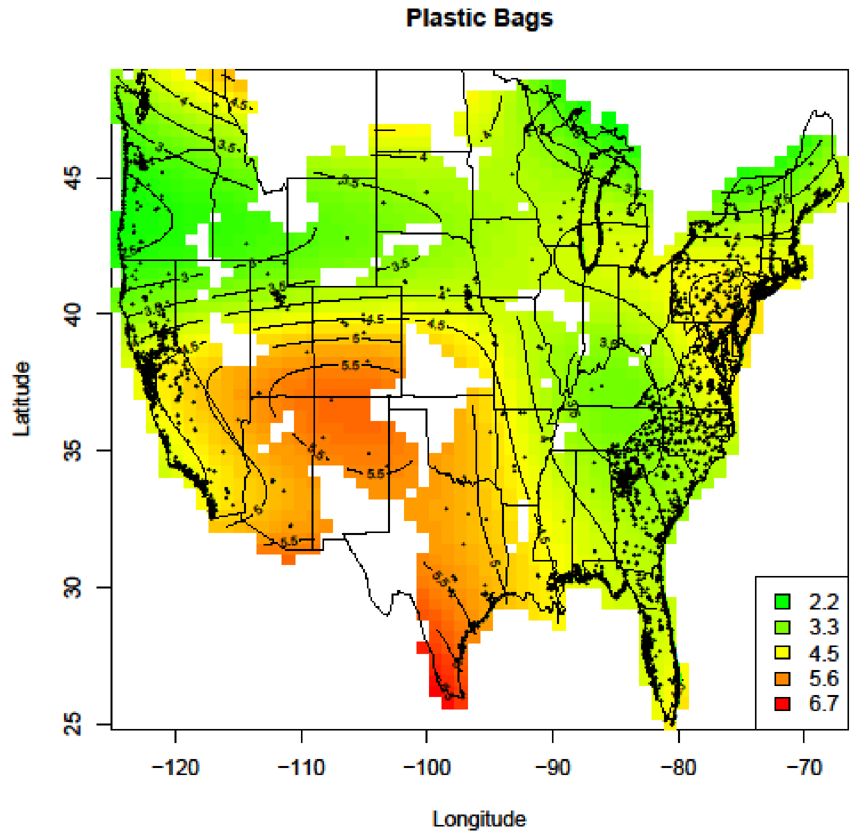
Starting with items with predicted high ecological impacts, we examined both the absolute load and the relative portion of the total items at each sampling site. Fishing gear was particularly common on the Texas coast, followed by the northern Atlantic coast, and subsequently the coastal margin of southern Florida. Interestingly, it makes a relatively high proportion of debris in these areas also. It also forms a high proportion of items along the northern part of the Pacific coast, where loads in general are much lower. Balloons follow quite a different pattern, with a relatively constant distribution across the US in both their absolute abundance and in their relative contribution to loads at a site. Plastic bags were relatively common in Texas and in southern California. However, this was not reflected in their relative frequency in debris loads – they composed a similar fraction of the total load across the US.

Plastic beverage bottles form a relatively large component to debris loads at inland sites in the eastern United States. This pattern is reflected to some extent in the absolute number of bottles removed at clean-up sites, although as with load in general, beverage bottles appear to be particularly high in Texas. The absolute load of cigarette butts was relatively high in the coastal eastern US, and the southern and northern ends of the US west coast. Their relative contribution to debris loads was also high on the US east coast, but did not show particularly strong spatial patterns elsewhere.



Fishing Gear
Latitude
Longitude
0.028
1.5
3
4.5
6
Fishing Gear
Latitude
Longitude
−7.3
−6
−4.8
−3.6
−2.4


Plastic Bags
Latitude
Longitude
-4.1
-3.5
-2.9
-2.4
-1.8


Plastic Bags
Latitude
Longitude
2.2
3.3
4.5
5.6
6.7

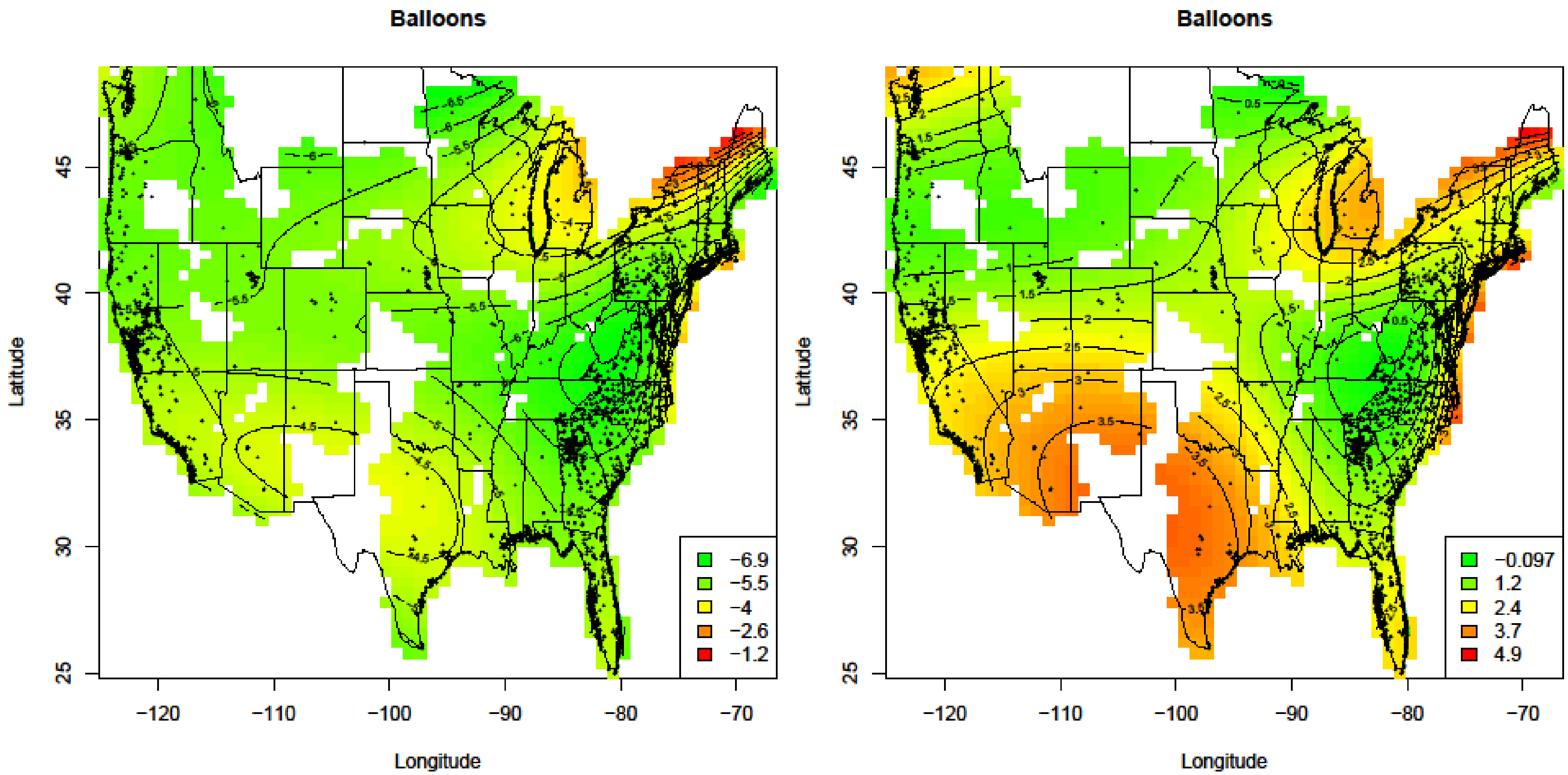
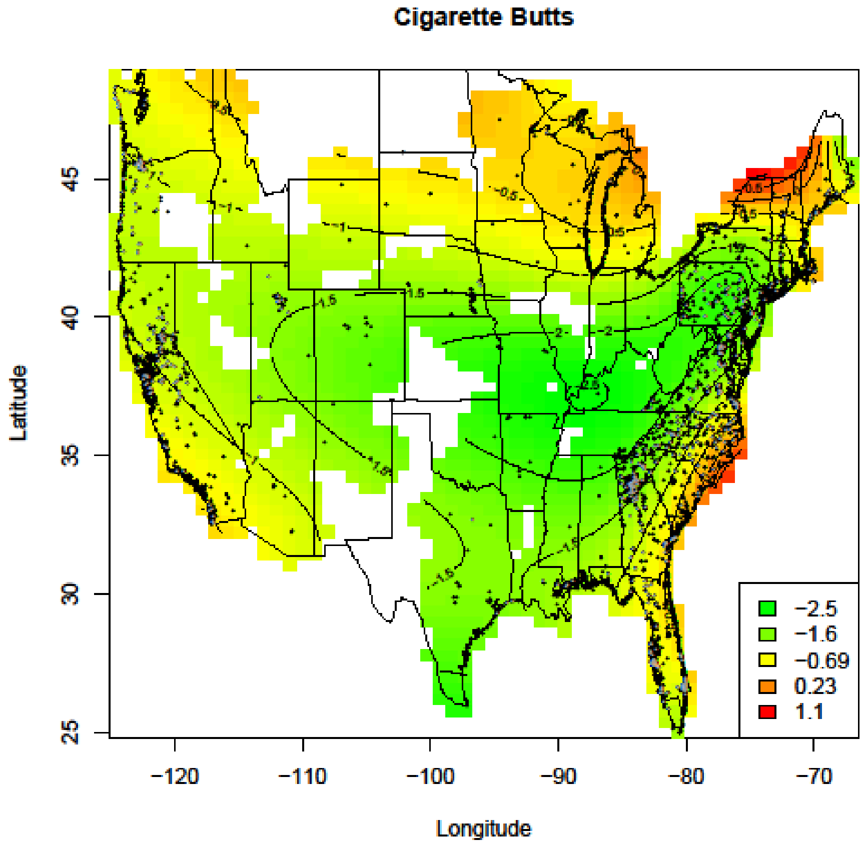
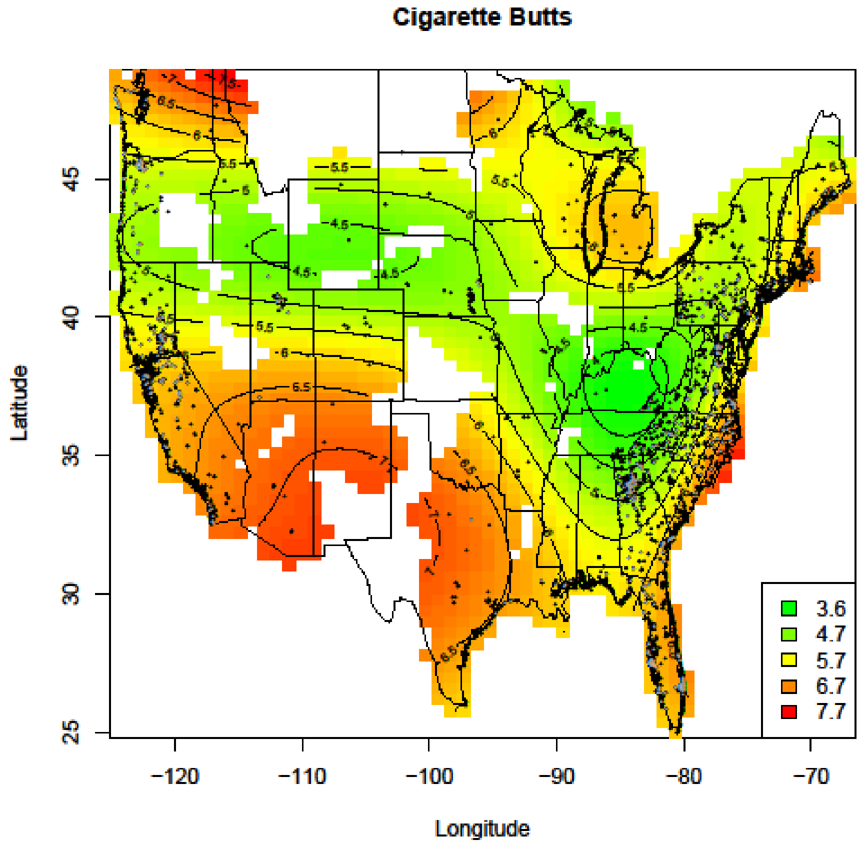


Figure 2.3.b Items with predicted high ecological impacts at ICC clean-up sites in the United States.
Colors represent the spatial density on a log scale, from lowest values in green to highest in red. The data are standardized for sampling bias (i.e. sampling area and number of people on the clean-up), but otherwise unadjusted. Black points in the plots illustrate the locations of sampling sites. Points with white centers are those sites with the top 10% of densities for the items in question. Plots on the left of each pair show the proportion the item forms of the total number of items at a survey site; plots on the right show the absolute count for the number of items. See Section 6.4 for model details.



Cigarette Butts
Latitude
Longitude
45
40
35
30
25
−120
−110
−100
−90
−80
−70
−2.5
−1.6
−0.69
0.23
1.1



Cigarette Butts
Latitude
Longitude
45
40
35
30
25
−120
−110
−100
−90
−80
−70
3.6
4.7
5.7
6.7
7.7

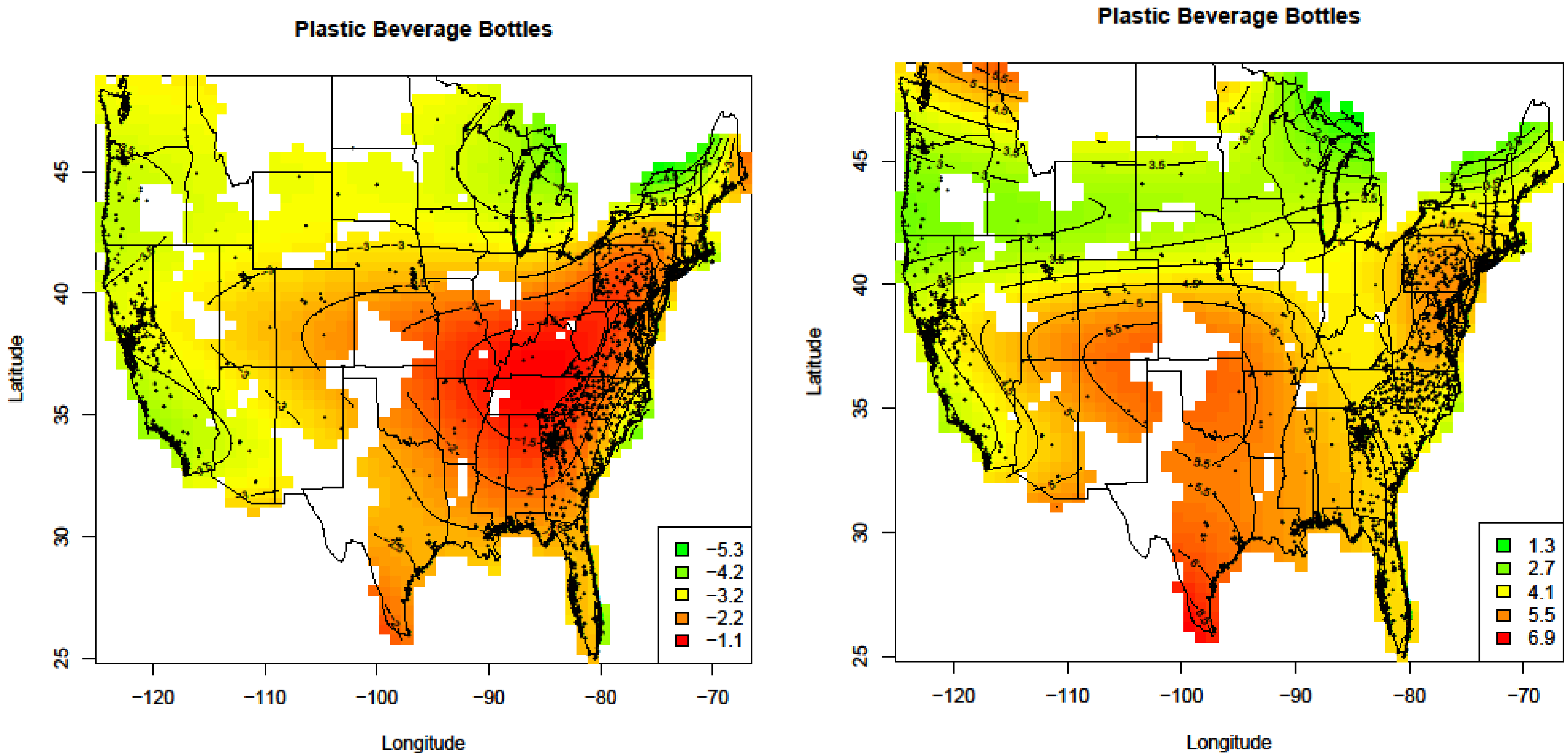


Figure 2.3.c Items commonly found at ICC clean-up sites in the United States.
Colors represent the spatial density on a log scale, from lowest values in green to highest in red. The data are standardized for sampling bias (i.e. sampling area and number of people on the clean-up), but otherwise unadjusted. Points in the plots illustrate the locations of sampling sites. Points with white centers are in the top 10% of densities for the items in question. Plots on the left of each pair show the proportion the item forms of the total number of items at a survey site; plots on the right show the absolute count for the number of items. See Section 6.4 for model details.

2.4 How effective are policies/legislation in reducing debris loads?

One of the key questions often asked is whether there is a relationship between legislation and debris. Given the increasing profile of debris on wildlife, tourism, economies, etc., there are a number of consumer items that have been under consideration for legislative action. These include everything from microbeads in facial scrubs to plastic bag bans and beverage container deposit legislation. Evaluating policy responses to particular consumer items is possible in the context of the survey methods, depending on the item of focus. For instance, microbeads are too small to be sampled effectively using the NOAA, ICC or CSIRO survey methods, but plastic bags and beverage containers fall under those items that are both commonly found on beaches (see section 2.3) and that are associated with significant impacts on wildlife (Wilcox et al. 2015).

We analysed the NOAA and ICC datasets to ask how effective is container deposit legislation (CDL)? To address this, we determined the total number of containers that would be eligible for the container deposit scheme for both the NOAA Accumulation and the ICC data sets. We determined the mean proportion of containers within each survey by state (Figure 2.4.a; Figure 2.4.c) for NOAA and ICC data respectively). The analysis includes plastic beverage containers, glass beverage containers and cans. containers.

Beverage containers compose a smaller proportion of the debris collected in states with CDL (California, Hawaii and Oregon) compared to states that do not provide a cash incentive for recovery of beverage containers (Alaska, Texas, Virginia and Washington) based on NOAA's data (Figure 2.4.a). The same pattern holds for the ICC data, which includes data from more states than those monitored by NOAA's accumulation surveys, though it is less clear-cut. Of the 43 states in which clean-ups took place during 2012-2015, six of the nine states with the lowest mean proportion of beverage containers have CDL. Additionally Connecticut and New York have CDL in place, and are in the bottom 45% of the states in terms of beverage container frequency (Figure 2.4.c).

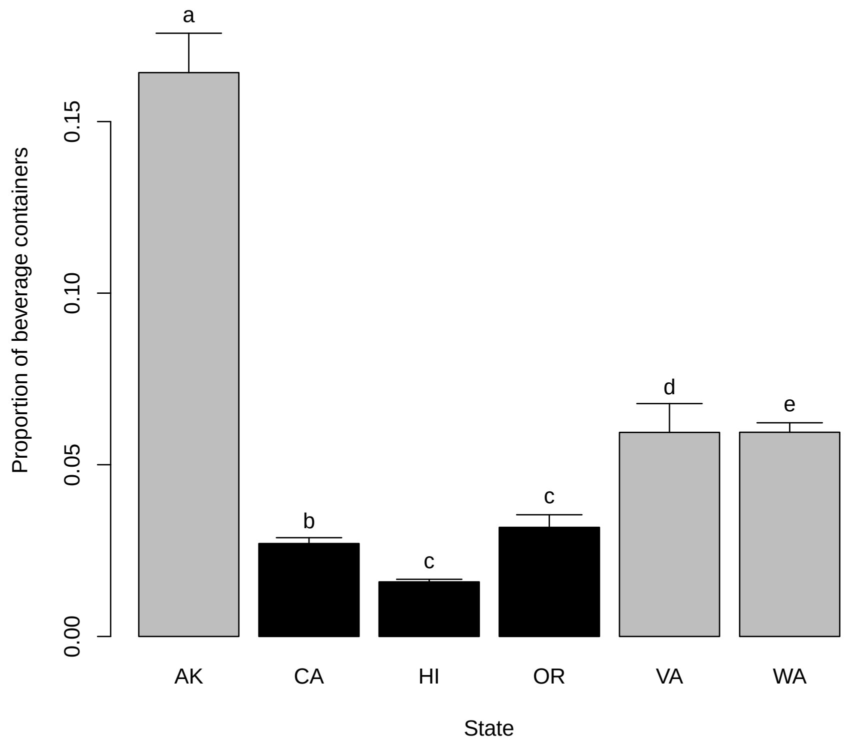


Figure 2.4.a. The mean proportion of beverage containers from NOAA's accumulation data (2012-2016).
Note that states in black are those which have CDL and those in grey do not provide cash for containers. Error bars are standard error of the mean. Letters above the bars indicate statistically significant differences.

Because it is conceivable that people in certain states simply consume higher proportions of beverages than other states, we also calculated the proportion of lids to containers. Under current container deposit legislation, containers have value; lids do not. We would generally expect that every plastic or glass bottle would have a corresponding lid, and it is likely that lids and containers would be discarded together. Aluminium cans are, of course, slightly different, with the pull tab typically remaining with the container. The container calculations incorporate all beverage containers, both aluminium cans as well as plastic and glass bottles. Therefore we would not expect to see a 1:1 ratio of lids to containers, but would still expect a difference in the ratio of lids to containers between CDL and non-CDL states, because a significant proportion of containers are either plastic or glass.

The three states with CDS captured within the NOAA Accumulation data tend to have higher ratio of lids to beverage bottles than states without CDS (Figure 2.4.b). The results are similar, though not quite as pronounced for the proportion of lids to beverage containers from the Ocean Conservancy data (Figure 2.4.d). All but one of the states with CDLs falls in the upper half of the states in terms of lids to bottles. Similarly, none of the 10 states with the lowest ratios had a CDL.

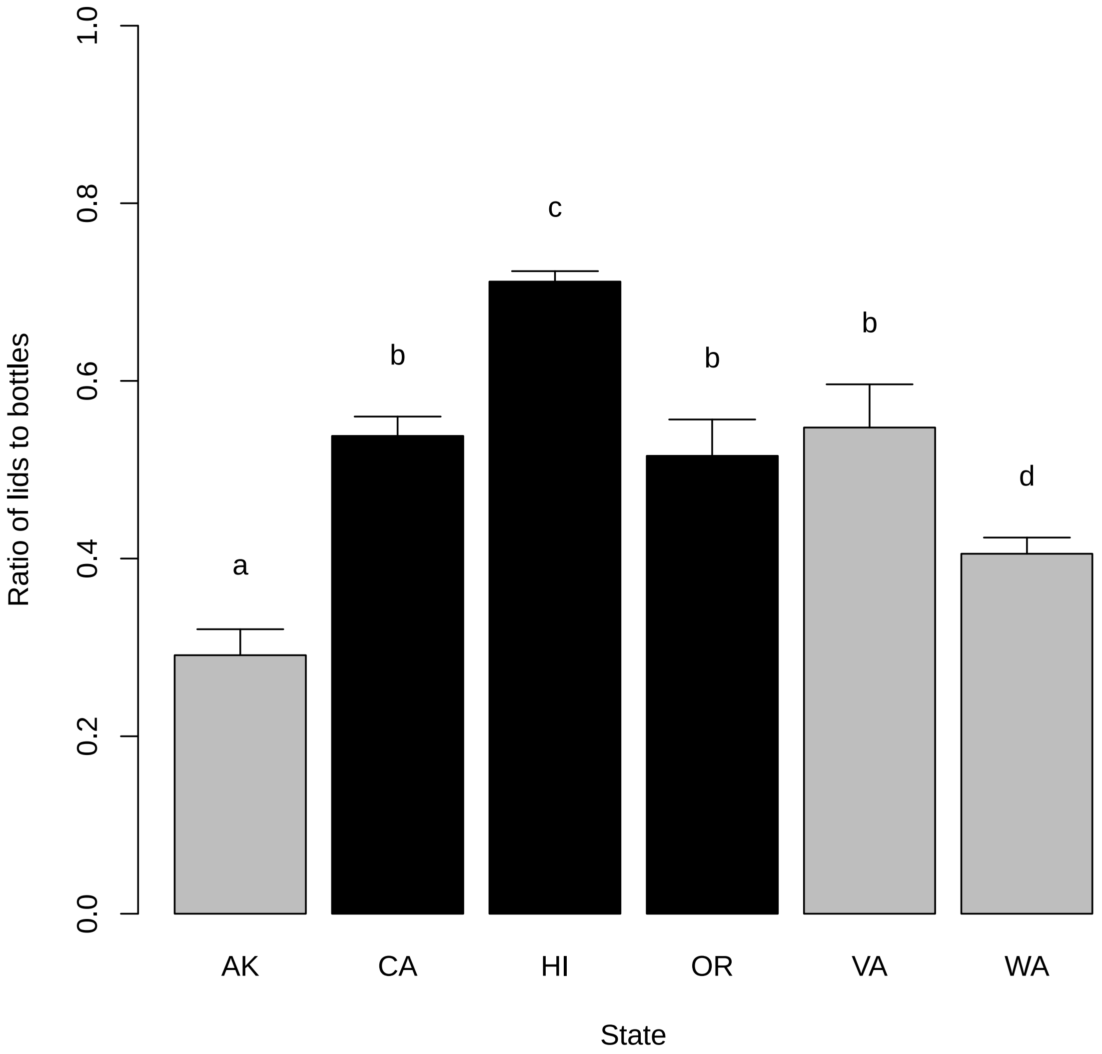


Figure 2.4.b Mean proportion of lids to containers in NOAA Accumulation surveys (2012-2016). Error bars are standard error of the mean. Letters above the bars indicate statistically significant differences. Black bars are states with container deposit schemes in place, grey bars do not have CDL.

Overall there appears to be strong evidence that CDLs affect the chance that bottles end up in the environment. It is unclear if this reduction is due to consumer behaviour, or subsequent removal by scavengers. However, in either event the outcome for the environment is similar. The pattern we identified in the data relative to CDLs would probably be yet stronger if we included economic factors driving recovery rates, such as local materials prices and socio-economic levels near the survey sites.

Currently, there are a number of states considering CDL, such as Maryland, Virginia, Massachusetts and Texas (and the District of Colombia), though this is a contentious issue. Nonetheless, this provides great opportunity for pre- and post-legislation information which can be used to evaluate the effectiveness of legislative changes on the proportion of beverage containers being mismanaged.

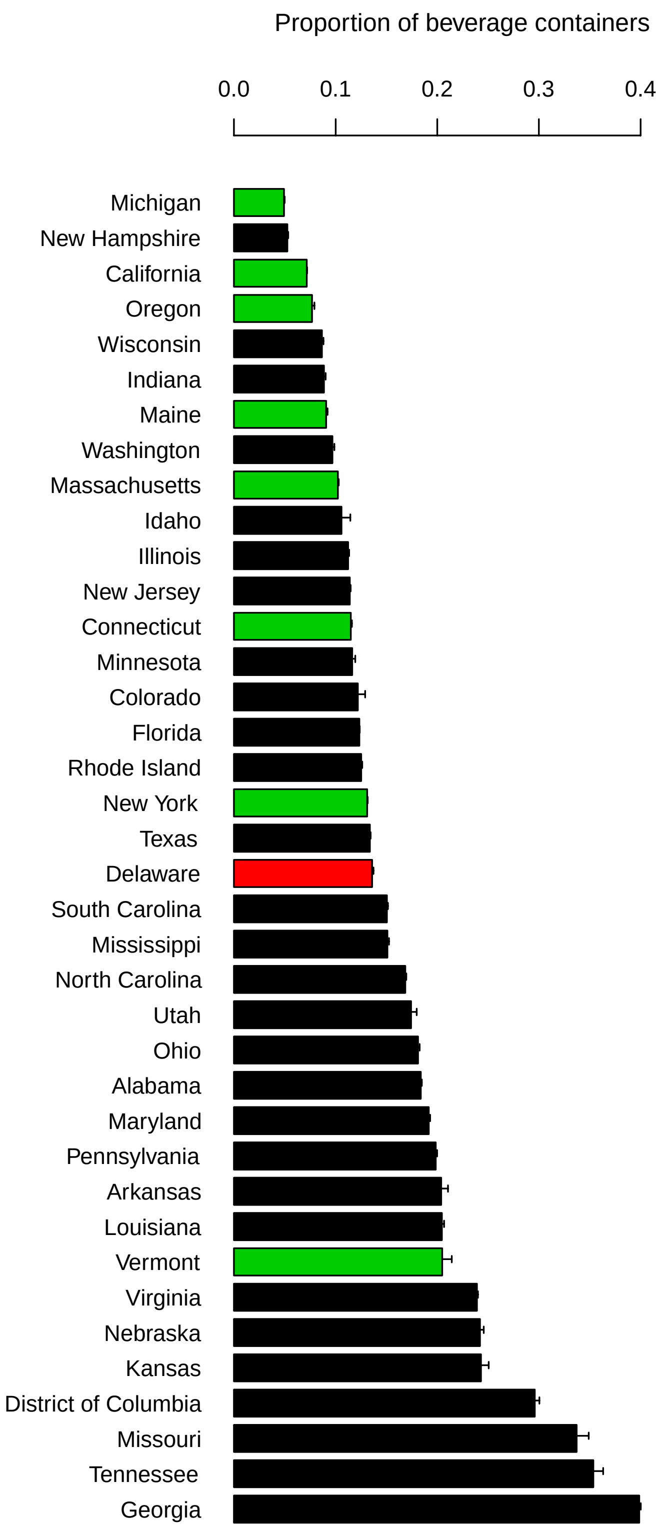


Figure 2.4.c Mean proportion of containers in ICC surveys.
Error bars are the standard error of the mean. Green bars are states with container deposit schemes in place, black bars do not have CDS. The red bar, Delaware, formerly had a CDS and now has a unified recycling scheme.

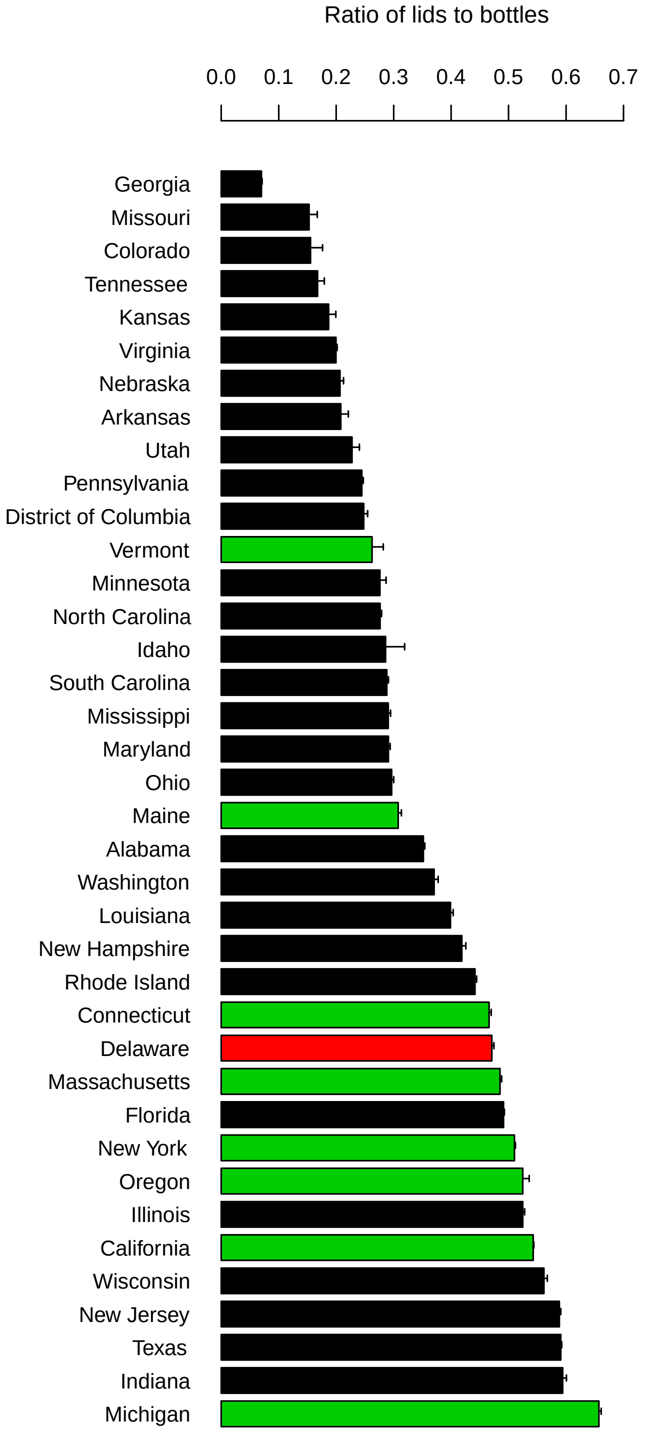


Figure 2.4.d Mean proportion of lids to containers in ICC surveys (2012-2015).

Error bars are the standard error of the mean. Green bars are states with container deposit schemes in place, black bars do not have CDS. The red bar, Delaware, formerly had a CDS and now has a unified recycling scheme.

2.5 Does the "diversity" (types and relative frequency) of marine debris vary spatially or temporally?

What seems like a straightforward question actually has a complex set of questions and factors embedded. If there are particular questions about patterns of diversity to drive analysis, they can be tested. For example, asking questions of specific items (as in 2.3) yields interpretable, useful results. Because of the high number of data categories, variety of survey types, uneven sampling and variety of sources at the site level, however, this is a difficult question to ask of all the data at once (even of a single survey or clean-up type).

In the absence of clear, explicit questions, data mining may not necessarily yield clear patterns (see Sections 6.5, 6.6). We used a number of data mining tools to look for patterns in the distribution of items across the categories recorded by NOAA and the ICC to evaluate whether we could identify different types of sites or identify drivers that created different debris profiles. This effort produced patterns in some cases, however, the analyses tended to group at two ends of the spectrum: either too many types of sites and possible drivers to make sensible interpretations or too few in order to identify meaningful differences among sites. While returning to this line of investigation might prove fruitful in further work, it will be essential that it is question driven rather than based on a data mining approach.

Working with individual loads of particular items, as in the 5 items we analysed specifically (Section 2.3) or asking specific questions will result in clearer patterns. For example, if the goal is to identify dumping sites, investigating the frequency of items that are unambiguously associated with the source might be fruitful. In this case, one might look for tires, concrete, car parts or something that would be linked to illegal dumping, and then ask specific questions of the data (Do we find these items in remote sites far from roads? Are they at sites with easy road access along urban margins?). In a case like this however, it is not clear that there is much to gain from a multivariate approach, considering all the types together. A simpler approach could be to make individual models for each candidate across the items that have been identified as being relevant. The outputs from these analyses can then be combined to infer locations or contexts that appear to be hotspots across all of the candidate items.

Part of the complexity of the issue is because items are not unambiguously connected to sources. So, although you might like to simply use frequency of different items to indicate a source, there is not a one to one match. Furthermore, because sites have multiple sources, the 'assemblage' of items is going to be well-mixed. This means there are not necessarily sites with specific characteristics or specific items that solely occur at them. For many purposes (such as looking at policy effectiveness, what are particular items to target for reduction campaigns, etc.), analysing the load of each item at a site separately (for example, at the individual level) can answer the question with greater clarity.

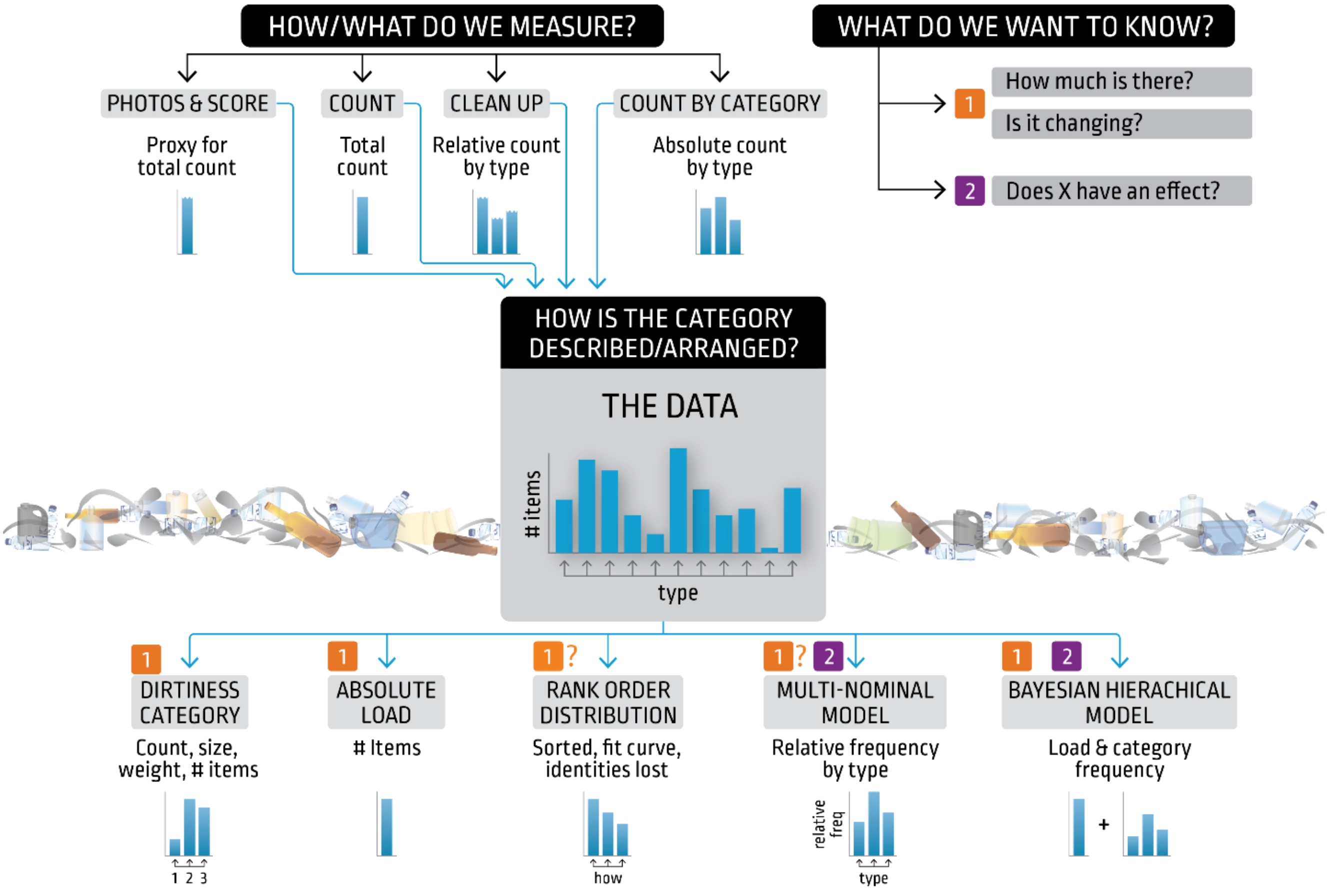


Figure 2.5.a How is the category described/arranged?
Description of what questions might be asked of data (1, 2 in orange and purple), how or what is measured, with how the data may be collected, described or arranged and what questions and analyses might be conducted to address particular questions and different types of data.

Fundamentally, there is a trade-off among the analytical approaches that can be taken to marine debris data arising from the various survey types that have been carried out (NOAA surveys, ICC clean-ups, CSIRO transects, Figure 2.5.a). At one extreme, the data can be aggregated up into a total count of items across all categories or a total weight (Figure 2.5.a, lower left). The advantage of this approach is that by using a single category of data modelling efforts can focus on the full complexity of space and time patterns, incorporating both driving variables, such as local population size, and nuisance variables, such as sampling effort. This is the approach detailed in Section 6.4 of this report.

At the other extreme, one might model the abundance of items in each category, across the tens of categories that are recorded in the NOAA or ICC data. The challenge in taking this approach is that models describing the abundances in each category may differ, leading to a very complex interpretation of the data (Figure 2.5.a, lower right). Furthermore, categories may be positively or negatively correlated so the direction of the link between items and abundance may be difficult to interpret (consider the case of bottles and caps presented in section 2.4; see Section 6.4 for details of results from modelling debris loads for NOAA and ICC data). Intermediate tools such as richness curves or rank order distributions, as typically used in fields like community ecology (Figure 2.5.a, central bottom) do not seem to be particularly linked to useful questions in this context, though this is something we considered and explored.

2.6 Do patterns of distribution and abundance of marine debris change over time (i.e., are there discernible temporal differences in characteristics of debris in regions with sufficient sampling)?

If there is a time trend in the debris loads at survey sites, the final statistical models we fitted should include a year term and that term should be statistically significant. These models are detailed in Section 6.4, but in this section we focus primarily on the time trend indicated by the analyses. The models for the NOAA accumulation data had a year term in the best-fit model, with a positive coefficient, suggesting that debris densities were increasing with time. This was by far the most important coefficient, based on effect sizes, indicating that the temporal trend is an important explanatory component in the model. The best fitting model for the ICC data also included a year term, indicating there was a temporal trend in the data. In contrast to the NOAA accumulation data, the ICC data had a negative time trend, with a coefficient of -0.049. The effect size for this term was the largest of the terms in the model, indicating that the temporal trend was important in explaining the volume of debris removed from ICC sites.

It is difficult to interpret the difference in the time trend in these two data sets. The samples do not occur at the same sites, there is some variation in sampling over time in both data sets, and sampling protocols differ in the two datasets. Given that the ICC data comes from clean-ups, one might expect sites that are repeatedly visited to be cleaner each time, generating a negative time trend driven primarily by the sampling method.

A possible explanation for the negative trend in the ICC data is site selection changes with time. If sites with the highest loads have been chosen for clean-ups early in the program, with lower load sites added later, this could produce a negative trend. In the absence of randomization in site selection, this is a difficult bias to identify. It could be possible to look at clean-up events by the individual organizations over time to investigate this site selection-driven effect.

The time trends could be complicated by the interval between surveys. When we investigated the effect of the time between clean-ups on the ICC data it was significant, but negative. This suggests that past clean-ups are affecting the amount found in future ones, however, in the opposite direction one would expect – the longer since the last cleanup, the less debris found. Again, this could be due to choices by organizations conducting cleanups. The organizations might not choose to revisit a site if they perceive it as still being clean from their last cleanup, thus generating a positive correlation between the debris load at a site and the time since it was last cleaned. In contrast, we found that the longer the time between surveys for NOAA accumulation surveys, the more debris reported, as one would expect. The NOAA protocol does control for some of these site choice biases, as established sites are cleaned on a regular interval.

However, the overall positive time trend in the NOAA data may also be affected by sampling bias. There is variation among the NOAA sites in the times at which they were sampled. Only 3 of the 11 sites sampled in the first sampling periods are also sampled in the last periods (see points at left, Figure 2.2.a). Similarly, only 3 of the 15 sites sampled in the last period were also sampled in the first period of the survey (see points at left, Figure 2.2.a). There is also variation in the relative load across these sites, which together with changes in sites sampled, could generate a temporal trend. In the first period of sampling, there were roughly 10 sites, of which six seem to have relatively low raw debris densities (see points at left, Figure 2.2.a). In the last period of sampling, several very high density sites had been added (sites 50 and 51, see Figure 2.2.a), creating a possible temporal bias in sampling. Overall, examining the pattern in the raw data there is not a general trend of increase or decrease with time in the NOAA accumulation data (see points on left, Figure 2.2.a), suggesting that the temporal trend is likely due in part to sampling variation and not an overall shift in load.

In summary, while we found support for a temporal trend in our analysis it appears likely that this is due to a sampling bias. Overall there does not appear to be a clear trend in the NOAA data, and we would expect a similar case in the in the ICC data. A fundamental problem in evaluating a time trend is the lack of balance in sampling over time across sites. If a time trend is a key interest for program design, it will be critical to address this sampling issue.

2.7 Do we find that coastal sites at river outlets have higher debris loads?

We investigated a few different ways of determining whether sites with river outlets had higher debris loads. First, we incorporated the distance to the nearest river into the statistical models to determine whether simple proximity explains any of the variability in the data. However, a simple distance measurement does not take into account the size of the river or how the amount of debris input into that river from upstream may vary. To determine the magnitude of these factors, we calculated the size and population of the watershed of the river nearest to the site, which we then evaluated in the statistical models of debris loads.

The USGS delineates watersheds at several different scales. We wanted to use a scale that would accurately reflect the terrestrial input into the rivers exiting close to the survey sites. However, we found that the appropriate scale varied slightly between survey sites. Some rivers were better represented by the scale 5 watersheds, others by scale 4. We therefore calculated the watershed area and population within that watershed for both 5 and 4 scales, to determine which was most reflective of the data (see detail in section 6.2.4).

We incorporated both the distance to the river and watershed factors into the GAM model selection for the CSIRO west coast data, new NOAA Standing Stock data from 2016, and both existing and new NOAA Accumulation data sets (see Tables in Section 6.4). We were unable to carry out watershed analyses for the ICC data, because with so many inland sites, it is much more complicated to determine proximity to river mouth.

In the CSIRO models, the distance to the nearest river was included in the best-fit model, and was statistically significant. The coefficient was positive, indicating that sites further from river mouths had more debris. Watershed area (5) and watershed population (5) both appeared in the model. Interestingly, the area was negatively correlated with debris, while population was positively correlated. This may be driven by watersheds comprising large, undeveloped tracts of land. All three effect sizes, however, were relatively small (Figure 6.4.4.a).

In the NOAA Standing Stock data collected in 2016 watershed population (4) was included in the best model, and was positively correlated with debris. Again, effect size was relatively small. The sample size, however, was relatively small in this model, given the number of surveys, so these results should be interpreted with caution.

In the Accumulation model based on pre-2016 data, watershed area (4) was a statistically significant factor in the model, though with a small effect size. It was negatively correlated with debris amounts, similar to CSIRO GAM results. River distance was not statistically significant, but negatively correlated, meaning that NOAA accumulation survey sites further from a river have lower debris levels.

By their appearance in the best-fit models, it is clear that distance to nearest river, as well as the watershed characteristics, influence the amount of debris at a site. However, the story is not necessarily straightforward for all survey methods and sites. Again sampling bias may play a role in the complexity of this story. Survey sites in the NOAA data do not appear to have been chosen at random, and thus there may be some underlying bias generating relationships between debris loads and river proximities. One option for addressing this question would be to set up a structured sampling system around several river systems. This could be done as a single investigation at one point in time, and analysed specifically to understand the effect of rivers on loading.

3 A comparison of data collection methods

As part of Phase II of the collaboration, we compared NOAA, ICC and CSIRO-developed protocols to estimate, quantify and compare debris amounts and types. Comparing among these methods provides an opportunity to evaluate the relative power of each method to uncover pattern and process in marine debris at local, regional and national scales. We first provide a discussion of the sampling design considerations relevant for marine debris surveys. From this discussion we illustrate some considerations for the various survey methods. We then turn to comparing the methods based on data collected in the field using the NOAA and CSIRO methods in July 2016. Due to logistical constraints, ICC clean-ups were not conducted, so we cannot directly compare all four methods based upon simultaneously collected data. However, we do discuss clean-ups and statistically designed surveys with respect to data quality, concerns and constraints and extent of data. Based on experience and analyses, we provide a suite of recommendations to improve statistical power, reduce data collection effort and associated costs, improve scientific inference, and maximize scientific and policy insights related to marine debris monitoring and clean-up efforts going forward.

3.1 Differences in survey design across the data sources

Survey design is a key component in developing a quality data set. It is useful to consider design at a number of levels, working down through a hierarchy. First, at an overall level, surveys should be balanced across any variable for which inference is to be made. Thus for temporal trends, surveys need to cover the period in question. Similarly, for spatial trends it is best if all locations are covered consistently. If effects of river outlets are of interest, sampling should be structured according to their locations and balanced across factors that could affect their effects, such as the population in the watershed. Deviations from balanced sampling, for instance variations in sampling over time or location, can create confounding in the data, making it difficult to interpret. Second, it is important to control bias in site sampling. This is particularly true in a situation like the one at hand, where there is correlation between the chance of choosing a site and the variables affecting the site. For instance, access to coastal sites might be part of the survey location choice, but is also likely to affect visitation rates by the public and thus deposition rates for debris. It is important to use tools like randomization to avoid these biases to the extent possible, and where not possible to collect data to allow estimation of their effects in the analysis. Third, due to variation at the sites it is important to have within-site replication. Coastal locations vary significant in their loads even at small spatial scales (compare sites 10 & 24, Figure 2.2.a). Replication at the site level, and stratification of these replicates across the conditions at each site can assist in reducing variability at each site and allowing estimation of the driving variables for the variation where it appears. Finally, at the finest level, controlling survey effort and observation error is a key consideration. Ideally, any item in a survey should have an equal probability of detection, irrespective of size, shape, location, and observer. This is clearly an impossible task, thus it is important to control observer effort and detection probability to the extent possible. This can be done through standardizing search area, search time, and search speed. Recording information on the size and color of items can help with standardizing observations for detection error, particularly when considered in the context of survey conditions like substrate type and color. Finally, if the study goal includes predicting outside the observed conditions, it is essential that the sampling hierarchy described above covers the range of conditions for which predictions will be made. Analysis of different data types requires a multitude of statistical tools. Clearly identifying the main questions or goals of the project at the outset allows for appropriate analysis and interpretation of data. For example, if one wants to identify the baseline level of litter on the coastline and the goal is to make projections outside of where litter was collected or reported at sites, it is important to stratify the sampling such that various coastal types are sampled in proportion to their occurrence. If survey sites only encompass one substrate type or are of one shape, aspect, or slope, it is difficult to make predictions about the amounts of debris that occur at other sites within the region. However, if that is not a goal of the monitoring, such factors need not be incorporated into the survey design. With these considerations in mind, we evaluate each of the methods in turn.

Table 3.1.a Design characteristics for the data collection efforts.

Issue	International Coastal Cleanup	NOAA Accumulation Surveys	NOAA Standing Stock Surveys	CSIRO Surveys
Stratification of sites	No	Partial	No	Yes
Randomization of site location	No	No	No	Yes
Replication within sites	No	No	Yes	Yes
Stratification within sites	No	No	No	Yes
Randomization within sites	No	No	Yes	Yes
Control of survey effort	No	Yes	Yes	Yes
Control of detection probability	No	Yes	Yes	Yes

The various data sources we analysed address the sampling design issues outlined above to varying extents, as illustrated Table 3.1.a. The data sources can basically be grouped into volunteer driven clean-up efforts (ICC and NOAA Accumulation) and designed surveys (NOAA Standing Stock and CSIRO). Volunteer efforts are to some extent driven by the availability and initiative of volunteers, and thus various aspects of the resulting sampling design reflect this process. At one extreme among our data sets are the ICC data, which are typical for data from litter removal and beach clean-up efforts elsewhere. Site choice is driven in part by volunteer initiative, replication and randomization typically are not a feature of the data collection effort, there is some level of information on sampling effort collected, but little prescription about how effort is expended in the field, and detection probability is generally not controlled. While it is possible to control for biases introduced by this protocol to some extent, where data is available, the survey methods introduce sampling variation and to some extent reduce the inferences that could be made from a similar sized dataset collected in a more structured manner. Clearly, clean-up data is a side benefit of an activity that is not designed generally as a data collection exercise. Thus, any changes to the protocols need to recognize this situation and embrace any limitations imposed by the primary goals of the clean-up activity. It is possible to address some of these issues with formal protocols, as in the case of the NOAA Accumulation data. In this context there is some stratification in site location, particularly in the northern portion of the US west coast, and some effort to control survey effort and detection probability by providing guidance on the search pattern and minimum item size for the surveys.

However, even with structured protocols as in the NOAA accumulation data case, there remain some issues with volunteer driven sampling designs. Variation in the spatial and temporal patterns of sampling lead to biases that may affect estimates of temporal and spatial trends, as noted in Sections 2.6 and 2.7. Moreover, for both the ICC and the NOAA Accumulation data there is a correlation between the number of people surveying a site and the amount of debris found at the site, even when we control for area and other variables driving debris loads (see Section 6.4). This suggests either an increase in sampling effort or an increase in detection probability with more people. This is true even for the NOAA accumulation data, where the protocols attempt to control sampling effort and detection probability. There are at least two possible mechanisms driving these biases. The first mechanism has to do with search effort. The overlap among the area people search likely increases as more people participate in a survey. This reduces the distance between an observer and any given item, increasing the chance it is spotted. Second, this same effect means that a single location is searched multiple times. Other factors may also contribute. For instance, although the NOAA method uses a constant search pattern, larger groups may make navigating this pattern more precise, reducing the chance that portions of the area are unsearched. There may also be some expansion of the search area, particularly where participants are focused on cleaning a site, as opposed to data collection, as the primary objective. The second mechanism has to do with detection probability. Adding additional people to a sampling team may result in a shift in the detection of items at the lower bound. This is particularly likely in contexts where participant's primary objective is a clean-up. As larger items are found and collected, a picture of the appropriate level of removal effort may emerge across the group. As larger items are removed, this may inspire participants to seek smaller items that would otherwise not be targeted. In the NOAA accumulation protocol this should be minimized due to the specification of a lower size limit on items collected. However, it could still increase detection of items at the lower end of the allowable size range. In the ICC protocol there is no lower size range, thus one would expect this effect to be even stronger. Exacerbating this phenomena is that smaller debris items are generally more abundant than larger ones. Thus as more participants leads to the targeting of smaller items, the total count increases because of better detection rates, but also due to sampling size classes of items that are more abundant. Based on anecdotal observations of staff involved in clean-ups, this social dynamic does appear to occur during clean-ups, as does expansion of the area boundaries with more participants.

The NOAA Standing Stock method is closer to a designed survey. In this context there is replication within sites, and those replicates are chosen in a randomized manner. As with the Accumulation method, there is a minimum size for items to be included, which should help control observation error. There is control on effort, although we found that there is significant variation in the amount of debris found, depending on how many people participate in the survey (Section 6.4). However, the pattern suggested that the number of people on a survey might be correlated with the organization conducting the survey, and that variation among organizations in survey execution could be driving the differences (Section 6.4).

The CSIRO method is a designed survey, including stratification and randomization at both levels in the hierarchy (site choice and survey location choice). Observation effort is tightly controlled, with area surveyed constant across surveys, and structured in a manner that precludes variation between surveys. Detection error is controlled by standardizing visual acuity among surveyors (i.e. controlling for the distance between the observer's eye and the area searched) and the rate at which items are encountered. The effect of these controls on observer effort and detection error are visible in the similarity of the colored dots at a site for the CSIRO data (see left side, Figure 2.2.b), despite the fact that the surveys are stratified across different substrates at a site (such as beach, rock slab, or cobbles) which would expect to result in different debris densities on transects within a site.

3.2 Comparison of survey protocols at shared survey locations

We wanted to determine whether the NOAA Accumulation, NOAA Standing Stock, and CSIRO survey methods yielded similar results based on a set of common survey sites. We also wanted to evaluate how accurately a model fit to data from each method can predict outside of the survey sites. For instance, since NOAA surveys are all conducted on sandy beaches, how well will they predict debris on other substrates? The analysis is based on data collected by project staff, in collaboration with NOAA and Ocean Conservancy staff, during July 2016 for the CSIRO and NOAA Standing Stock methods. These surveys were conducted at the same sites, typically at adjoining locations. We utilized data collected as part of NOAA's existing accumulation sampling program, collected between 15 June and 15 August 2016 in the same region for comparison. We limited comparisons to Accumulation surveys that were within 5 km of the CSIRO or NOAA Standing Stock survey to be compared. For the CSIRO and NOAA Standing Stock data we used the mean across the replicates at a site to compare with the single value reported in the NOAA Accumulation surveys. In total we were able to compare 9 sites across the three methods, with the CSIRO data overlapping an additional site for both NOAA datasets, allowing a 10th comparison.

The three datasets do not appear to be strongly correlated, indicating that densities measured by the datasets are relatively different (Figure 3.2.a). If the protocols yielded similar answers, one would expect the points in all three graphs to be lined up diagonally. Note that these are only for the 9 sites in common between all three data sets, and only for the survey closest in date to the CSIRO survey date at that location. The two NOAA protocols appear to be almost negatively related, with high standing stock surveys having low accumulation surveys and vice versa.

We tested for correlation among the data from each of the surveys. We used two measures from the datasets, the total number of items collected and the density of items. Because the data are non-normally distributed, we tested if the relative order of the total debris counts or debris densities among the surveys was correlated, using a Spearman's rank order test. The correlation tests showed that the CSIRO and Standing Stock methods were positively correlated on a rank order basis, although the correlations were not particularly strong (Table 3.2.a).

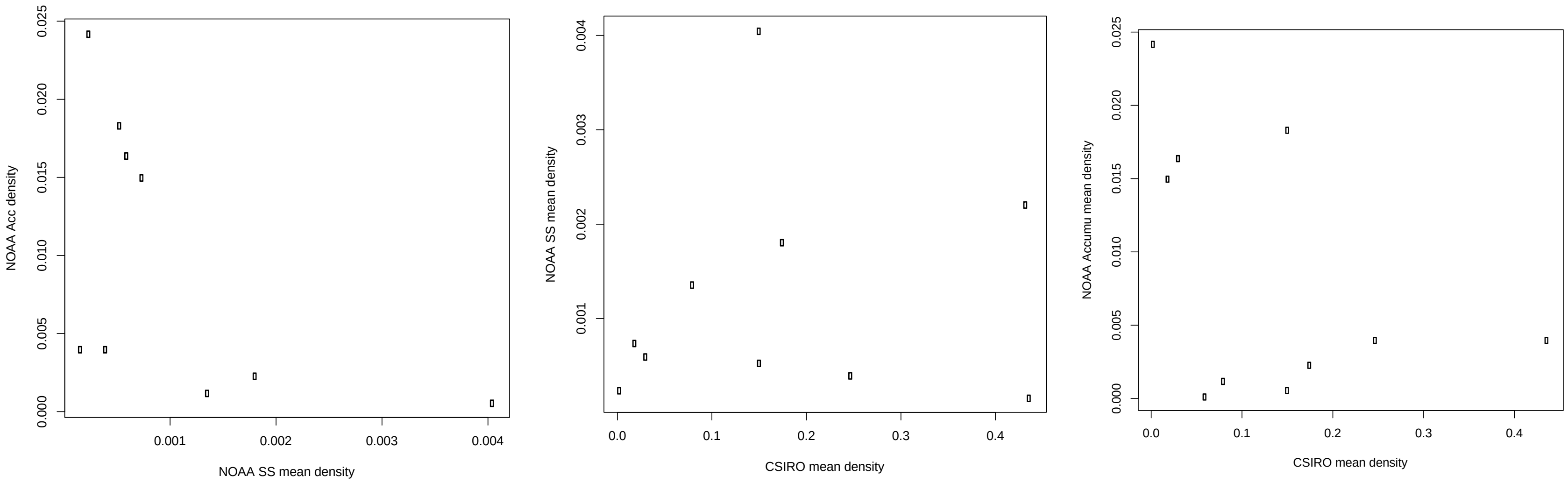


Figure 3.2.a The mean density for NOAA Standing stock sites (July 2016 data) and density of correlated NOAA Accumulation sites (Jan-Aug 2016).

The effect of the lower boundary on the size of items recorded can also be seen in the comparison among the data sets (Figure 3.2.a). The CSIRO method, which includes much smaller sized debris, consistently measures much higher densities of debris at a site, even when compared to the NOAA Standing Stock protocol, which was conducted by the same staff in the same location and time. This difference also likely leads to a difference among the sources that are being measured. Larger items tend to be more associated with litter, as can be seen in the comparison between frequencies of items recorded in the NOAA Accumulation and Standing Stock surveys, to those recorded by the ICC surveys (Figure 3.2.a). The frequency of large, identifiable objects is significantly lower in the NOAA surveys than the ICC surveys, while fragments are more common. Similarly, the CSIRO methodology will pick up even smaller fragments than those found in the NOAA surveys. Small plastic fragments dominate the CSIRO surveys. These small fragments are typical of materials found in coastal environments and offshore, and likely the result of marine transport to the survey sites, as opposed to direct deposition of litter.

This difference in the importance of marine and terrestrial sources of litter at various sites might explain the low level of correlations between the CSIRO and NOAA Standing Stock surveys. For instance, sites near urban areas appear to be strongly influenced by local, and likely terrestrial, sources of material. Remote sites, particularly ones that are difficult to access, appear to be more strongly influenced by marine sources. The urban sites typically have higher numbers of large consumer items, and thus higher densities recorded following the NOAA protocol. The CSIRO protocol would include these materials. However, at remote sites, where materials are generally smaller, likely marine sourced, fragments, the NOAA method would find much lower densities as items would frequently be smaller than the 2.5 cm diameter minimum size for inclusion in the data.

Table 3.2.a Results of correlation tests for raw data.
Because the numbers of transects varied between methods, values (mean or by plot) were either repeated to correlate with multiple transects from other methods, or mean values were compared. The statistic value represents the strength of the correlation. Higher absolute values mean that the two data sets are more strongly correlated. The sign of the statistic indicates whether the values are positively or negatively correlated. The p-value indicates whether the correlation is statistically significant. Spearman's index tests the rank-order correlation between values.

	Spearman's test		
Accumulation: CSIRO			
Test	Statistic	P-value	
Total items by transect	0.054	0.778	
Density by transect	-0.259	0.167	
Mean total items per transect	0.073	0.841	
Mean density	-0.292	0.413	
CSIRO:SS			
Test	Statistic	P-value	
Total items by transect	0.543	0.002	*
Density by transect	0.348	0.059	+
Mean total items per transect	0.555	0.104	
Mean density	-0.006	1.000	
SS:Acc			
Mean dens	-0.653	.-57	
Mean tots	-0.350	0.359	

There were no other significant correlations among the sampling methods, either in debris counts or density, based on their rank orders (Table 3.2.a).

To determine how well each survey methodology can predict outside its range, we selected all NOAA transects for both the Accumulation and Standing Stock methods on the west coast that were conducted between 15 June and 15 August 2016. We fitted a GAM model to each of the two datasets, incorporating all potential covariates, including the location along the west coast (see Appendix for details). We fit a similar model to the CSIRO data collected on the US west coast during July 2016 (see Appendix for details). Because the data vary, the best-fit GAM models contain different terms for each of the data sets. We then used the best-fit model for each dataset to predict the values at each of the transect sites for the other datasets, as well as for its own sites. In some instances, we were not able to use the full GAM model to predict outside of the range of the data, either because the site-level details were not collected in the same way, or because the levels of certain factors found within a data set were not within the reference data used to create the best-fit model. For example, CSIRO methods include collecting data on the composition of the land behind the transect site (backshore), as well as the shape of the coastline (concave, convex, etc.). These variables were included in the best-fit model, but because the NOAA data sets did not record such information, we were unable to include these variables in the full model comparing the three survey methods.

We worked around this in two ways. First, we fitted a simplified model to the full data set (Table 3.2.b). Next, we selected the common sites, determined the backshore and slope for the NOAA sites, and fitted the full model to a truncated data set (Table 3.2.c). Additionally, one of the covariates we collected from GIS layers was land use category. CSIRO surveys covered a much larger range of land use than did the NOAA surveys, so NOAA models did not have all of the reference levels within their data set. We therefore had to amalgamate land use into three categories which were common to all data sets (Barren, vegetated, and developed).

We then compared the predictions to the actual data gathered using the relevant methodology (Table 3.2.b, Table 3.2.c). We compared them based on the correlations between the predicted and observed values for each comparison, at the transect level. We used the correlation between the rank ordered values for each transect. In other words, we asked how well the models predicted the density of debris, and how well they could rank them in the order of their debris density.

Table 3.2.b Results of correlation tests for full data sets using truncated models to predict data. Density values for the 2016 data set for each of the three methods (CSIRO, NOAA Standing Stock, and NOAA Accumulation) are predicted using the best-fit model from each of the survey methodologies, and compared to one another. These results use the entire data set, so each data set has some sites that are not common to either of the other methods, and therefore in some cases the models are truncated to suit the available data. The sign of the statistic indicates whether the values are positively or negatively correlated. The p-value indicates whether the correlation is statistically significant Spearman's index tests the rank-order correlation.

	Spearman's rank test		
CSIRO sites			
Prediction Model	Statistic	p-value	
CSIRO	0.53	0	*
NOAA SS	-0.09	0.42	
NOAA A	0.27	0.01	*
Standing Stock sites			
Prediction Model	Statistic	p-value	
CSIRO	0	1	
NOAA SS	0.91	0	*
NOAA A	0.3	0.02	*
Accumulation Sites			
Prediction Model	Statistic	p-value	
CSIRO	-0.01	0.97	
NOAA SS	-0.4	0.01	*
NOAA A	0.92	0	*

Overall, models were reasonably good at predicting within their own data set, but not particularly good at predicting outside of the data. We then repeated the same analysis, but allowing the models to make use of all the covariates collected according to each survey protocol.

Table 3.2.c Results of correlation tests using the full models, but only on sites in common between all methods.
Density values for the 2016 data set for each of the three methods (CSIRO, NOAA Standing Stock, and NOAA Accumulation) are predicted using the best-fit model from each of the survey methodologies, and compared to one another. These results use the full model, but only on the common sites. The sign of the statistic indicates whether the values are positively or negatively correlated. The p-value indicates whether the correlation is statistically significant. Spearman's index tests the rank-order correlation.

Spearman's rank test

CSIRO sites			
Prediction Model	Statistic	p-value	
CSIRO	0.7	0	*
NOAA SS	0.12	0.51	
NOAA A	0.21	0.23	
Standing Stock sites			
Prediction Model	Statistic	p-value	
CSIRO	0.11	0.46	
NOAA SS	0.88	0	*
NOAA A	0.05	0.77	
Accumulation Sites			
Prediction Model	Statistic	p-value	
CSIRO	-0.01	0.97	
NOAA SS	0.19	0.6	
NOAA A	0.54	0.11	

With full models predicting on common sites, correlation was even less often statistically significant, likely because there were many fewer sites.

There are a couple of potential reasons for failing to find significant correlation between survey methodologies. First, although NOAA Standing stock surveys were generally collected on the same dates as the CSIRO surveys, there was as long as 5 months between the CSIRO survey and the closest

Accumulation survey. Hence, seasonal differences could account for the lack of correlation. Second, there are some significant differences in methodology between the Standing Stock and CSIRO data collections. In Figure 3.2.a, the mean CSIRO density is two orders of magnitude higher than the Standing stock measurements. This may be partially a result of the differences in survey area covered by the two methods. CSIRO surveys cover a much smaller area, and observers may therefore be able to more comprehensively collect debris in that area before observer fatigue sets in. More importantly, however, the NOAA methods are limited to items over 25 mm (~2 ½ inches) in diameter, while the CSIRO methods collect items down to 1 mm in diameter. It is possible that the densities of the smaller items included in the CSIRO method differ substantially across sites from the larger particles which are included in all three sampling methods. This is a likely explanation, given that smaller items are more likely to be from marine sources, whereas larger ones are more likely to be littered by beach users.

Considering the person-effort in evaluating survey methodology is also useful. As an indication of labor intensity, we determined the mean person hours required for carrying out the various survey methods (Table 3.2.d). Statistical analyses were carried out using data provided after October 2015 through August 2016. Although we were unable to conduct surveys using the ICC protocol during the 2016 sampling period to compare methodology, we do report mean statistics for the full ICC data set as a comparison. Note that the times reported in standing stock data should be taken with caution, as it is not recorded whether transects are conducted simultaneously or sequentially.

Table 3.2.d. The average number of person hours, transects per site, the mean and median area surveyed during 2016 surveys using NOAA standing stock and accumulation methods and CSIRO transects. ICC data also reported as a comparison.

	Mean person-hrs per site	# Transects per site	Mean area surveyed (m^2)	Median area surveyed (m^2)
NOAA Standing stock	2.04	1-4	3,927	3,130
NOAA Accumulation	3.14	1	18,721	6,000
CSIRO transect	1.45	3	119.12	84
ICC data (full data set)	N/A	N/A	4924 lineal meters of shoreline	1609 lineal meters of shoreline

There are always trade-offs in survey design. By reporting or collecting only larger items, observers can cover a larger area, and the variability in the sampling is therefore reduced. However, selecting for larger items potentially biases the sample more towards littered items as opposed to ocean-borne debris, which is usually smaller in size (e.g. fragments). This difference may account for the lack of correlation both between the raw data, as well as among the predicted data sets. Certain areas are likely to be more prone to littering, while others may have a higher abundance of ocean debris. If the methods differ such that ocean debris is less represented, the correlation even between rank order of sites will be less likely.

3.3 Power analysis

We were able to distinguish effects of a number of variables in the current sampling design for the NOAA surveys, CSIRO surveys, and the ICC clean-ups (see tables Table 6.4.1.c, Table 6.4.2.d, Table 6.4.3.d, Table 6.4.4.a). Given this pattern, it appears that the current data collection protocols have adequate power to distinguish important features in the data. For the NOAA and ICC datasets we were able to identify time trends based on the current sampling strategy.

Despite the conclusion that the current protocols do have adequate power, we chose to investigate the future power of the NOAA Accumulation Survey to distinguish annual changes in debris levels. In order to do this we used the model given in Table 6.4.1.c to simulate data from two possible years. In order to do the simulation we removed the year term from the model, refit it to the whole dataset, and then used it to predict the values in the 2015 data without a year term. This data served as our base year. We then simulated a second year after the base year, using the fitted model as before, but adding a small proportional change to the accumulation data. We then added a random deviate to both the baseline and the second year data, based on the variance in the residuals of the model fitted to the full Accumulation Data. We then fitted a new model, now including a term for the change between the baseline and the second year and evaluated whether the model was able to detect the change or not. We repeated the data simulation and change detection 100 times for each level of proportional change. We explored the range of possible changes from $1x10^{-13}$ to $1x10^{-2}$. This range spans the values from well below that detected in the existing NOAA Accumulation Data (see Table 6.4.1.c, Year term) to well above it. Detections of change were considered successful when the term for the change in the statistical model was significantly different from 0 with a p value of less than 0.05.

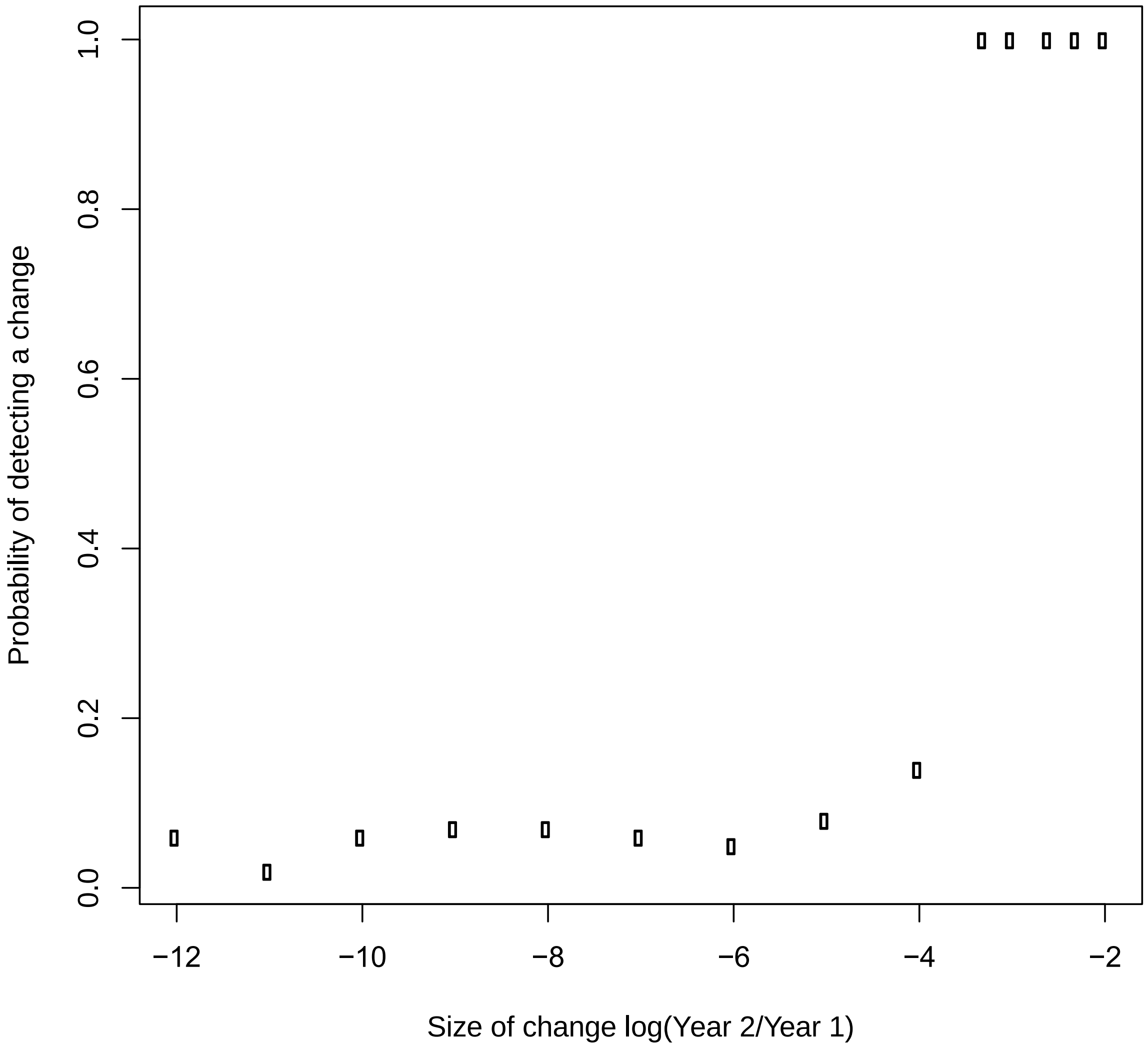


Figure 3.3.a Probability of detecting a change between years and the estimated value of the change for the NOAA Accumulation Survey in Washington, Oregon, and California.
Change detections are based on a Monte Carlo simulation using 100 samples. Detection is considered successful when a term representing time is significantly different from 0 with a p value of less than 0.05. The log of the ratio of debris densities between years is the proportional change, e.g. 1.01 is equivalent to a 1% increase in debris density values between years. In log terms, -2 in the plot.

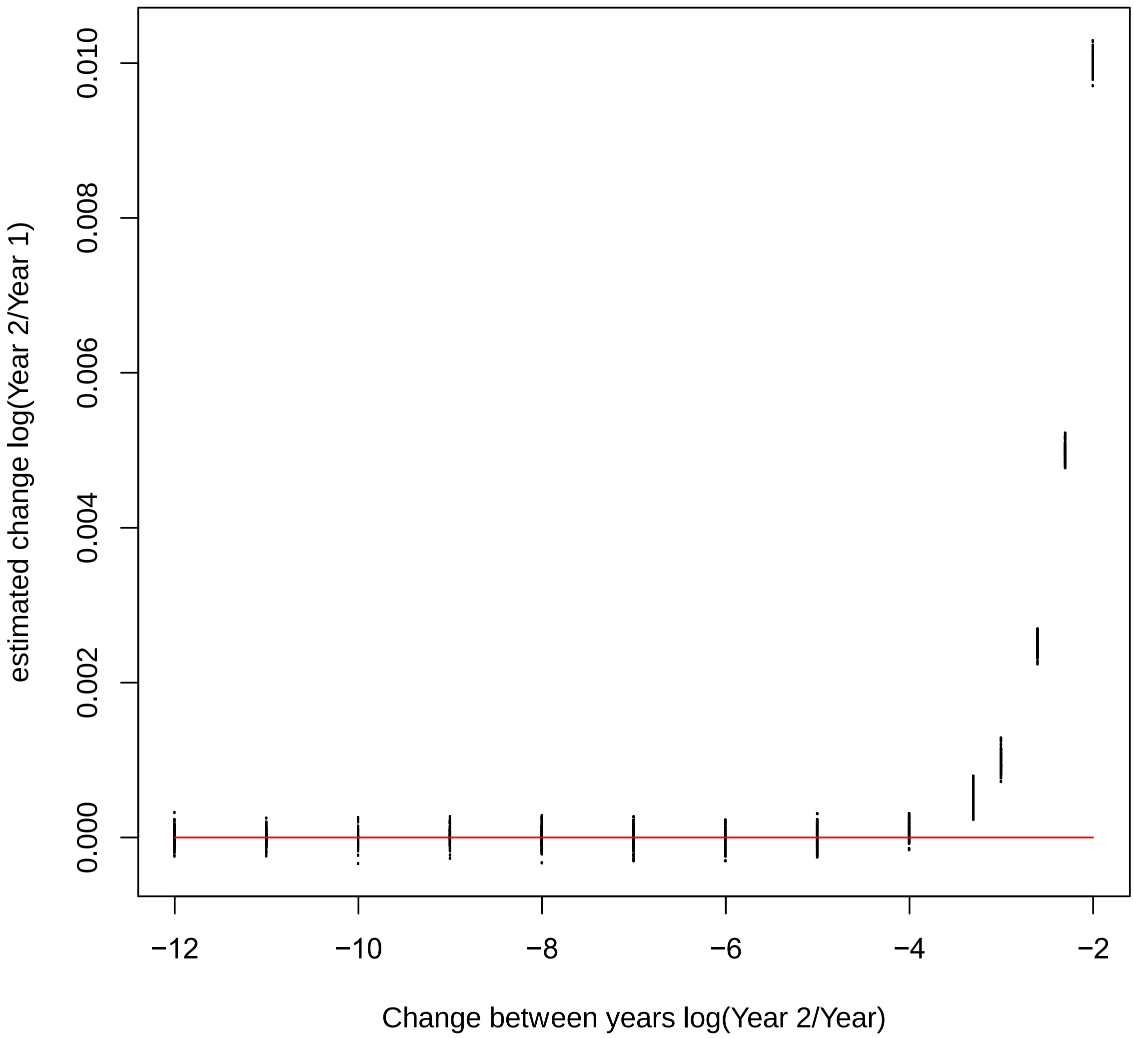


Figure 3.3.b Estimated change between years given a known change for the NOAA Accumulation Survey in Washington, Oregon, and California.
Change detections are based on a Monte Carlo simulation using 100 samples. Each point in the plot represents the value of the regression term for the difference between the baseline and the next year, using the model for the NOAA accumulation data given in Table 6.4.1.c. The log of the ratio of debris densities between years is the proportional change, e.g. 1.01 is equivalent to a 1% increase in debris density values between years. In log terms, -2 in the plot.

Assuming the sampling design and effort for the NOAA Accumulation Data in 2015, the current methodology can detect annual changes in debris densities of 0.0005 or more with high confidence. Values below 0.0001 are not possible to detect. Thus, a twentieth of a percent or greater change overall in the density of debris between years, assuming it is evenly distributed across all sites, should be detectable using the current approach. This should also hold true for a decrease in densities, as well as the increase analyzed here.

Two important caveats to keep in mind are the effect of the lower detection threshold on the survey approach and the effect of asymmetry across survey locations. Since the NOAA Accumulation Survey uses a fixed area search, as debris densities decrease, there will be fewer items in the survey area. If densities decrease to a level where surveys return counts of zero items, it will no longer be possible to estimate changes as the densities will be below the detection limit of the survey design. One possible solution is to adapt the survey method to increase search effort in locations where no debris has been found. The CSIRO method uses this adaptive sampling approach, increasing sampling effort until an item is encountered or the maximum number of surveys per site is reached. Spatial variation in the changes over time will also be harder to detect than the scenario examined here. The capacity to detect the trend will depend on the number of sites it affects, and how strong the effect is. One would anticipate a trend of the magnitude examined here, but occurring over a subset of the sites, would be harder to detect. The exact effect will depend on the aggregation of the sites affected, the size of the effect, and the variation in the remaining sites.

Further power analysis would be possible using a similar approach, either examining the power gained/lost by adding samples, or the capacity to detect other types of changes such as spatial effects or impacts of new variables not explored in this project. The key steps are to fit a model to the data not including the term of interest, project the data, add relevant noise into the data, and fit a new model incorporating the terms used to predict the new data, along with the new term of interest.

4 Conclusions

The International Coastal Cleanup data, together with NOAA's Accumulation and Standing Stock datasets represent a rich and informative source of information for understanding debris along the US coastline, and in portions of the interior waterways. With the addition of data collected during 2016 using a stratified random design, developed by CSIRO, we were able to investigate a number of major questions concerning the distribution, sources, and dynamics of marine debris in the continental US. We estimate there are somewhere between 20 million and 1.8 billion pieces of plastic on the coastline of the continental US. Given sampling biases and detection limits of the various sampling methods, the number is likely at the upper end of this range. In terms of hotspots, areas near cities, near rivers, and in some cases near international borders appear to have particularly high loads of debris. There is also some variation across political units, with states such as Texas, Alabama, Mississippi, and Washington having higher debris loads than one would otherwise expect. Some of this effect appears to be due to cross border oceanographic currents, which show clear effects on the US west coast and potentially in Texas. Some of these geographic patterns are also likely due to policy differences. We found a very strong and clear effect of container deposit legislation on the composition of waste at the state level, suggesting that other policies could also make major differences across communities and states.

The ICC and NOAA protocols result in data that is relatively aligned in terms of the types of items and frequencies of those items. Evaluating the ICC data, which has by far the widest spatial and temporal coverage, it is clear that individual items differ substantially in their frequency in the environment. For instance, fishing gear is particularly prevalent along various parts of the east coast. It also forms a high proportion of the relatively small amount of debris on the west coast. By contrast, plastic bags and balloons form a roughly constant fraction of the debris across the US, but appear particularly abundant in the southwest. These analyses point to the critical value of the item-by-item data collected by the ICC and NOAA surveys. While we found it very challenging to produce meaningful information using data mining tools on the profile of all items collected at clean-up sites, the data was very useful for asking questions about effects of policies, patterns for items where there has been significant industry engagement, or for items that are expected to be particularly ecological harmful. While this data across item types is somewhat more onerous to collect, it has tremendous value when examined with specific questions in mind. It would be impossible to answer these same questions with data on aggregate loads, in which items were not separated out.

In the second phase of this project, we evaluated the strengths and weaknesses of the various sampling methods, ranging from the ICC's volunteer clean-up data to the CSIRO's stratified random survey method. Much of the available data on marine debris is sourced from volunteer clean-up activities, either through the International Coastal Cleanup or through NOAA's two monitoring programs. These programs necessarily strike a balance between volunteer engagement and data collection. This balance is key for engaging participants, but implies some compromises from a survey design and data quality perspective. This is not to say that the data from these efforts is not useful, clearly it is. The technical appendices present analyses of this data, identifying a range of important factors affecting debris loads, including socio-economic levels, urbanization, access, inputs from rivers and estuaries, and a variety of other driving variables. However, as noted in Section 2 of the report, there are a number of biases in the clean-up based data which make interpretation of patterns challenging.

Considering survey design principals such as stratification, randomization, and replication, we suggest the protocols examined in this study increase in design rigor from the ICC, NOAA accumulation survey, NOAA standing stock survey, and CSIRO survey. There are a range of options that could help improve the data quality resulting from these sources, including developing a better understanding of the site selection process and search dynamics of volunteer participants to attempting to develop better spatial and temporal coverage in the survey designs. Some of these improvements can likely be done through interviews or detailed evaluation of the existing data, others will require modification of methods or survey designs, some improvements may require field experiments to understand human dynamics during clean-ups.

A key finding from our field evaluation of the two NOAA and the CSIRO method at common sites is that the distribution and dynamics of coastal debris are highly variable. Estimates of debris densities from the CSIRO and NOAA Standing stock methods at the same location were correlated at a rank order level, but gave quite different estimates of actual densities. Neither was correlated with nearby NOAA accumulation surveys, even on a rank order basis. This difference in the patterns across sites and methods suggests that not only is good survey design important in order to reduce variability in the surveys, but also that the system itself is just highly variable even at within a coastal site. Notably, the CSIRO surveys also identified much higher densities, due to the smaller minimum size of items included in the sampling (25 mm for NOAA, 1 mm for CSIRO). Comparing the ability of models fitted to each of the survey datasets to predict densities at sites outside their survey, models fitted to either the CSIRO or the NOAA accumulation datasets were able to make reasonable predictions. However, the NOAA standing stock surveys did not support accurate predictions of either the CSIRO or NOAA accumulation sites. This is a curious result, particularly given that the standing stock surveys and CSIRO surveys were conducted at the same sites in adjoining locations. This may be driven by differences in the minimum size of items included in the two surveys, 25 mm for the standing stock versus roughly 1 mm for CSIRO. This implies that the NOAA method may be biased more toward litter from terrestrial sources, while the CSIRO method includes these items, but also smaller and more abundant fragments transported by ocean currents. Anecdotally, based on staff observations during fieldwork this difference in source influence seemed to be born out, with CSIRO estimates generally higher than NOAA, but much more so at remote sites with little influence from nearby urban areas. We compared the cost, in person hours, of the 3 survey methods. The CSIRO method is much more expensive on a cost per unit area basis, covering 57 square meters per person hour as compared to 1,534 square meters for the NOAA standing stock method. However, the CSIRO method may be more cost effective overall, requiring only 70% of the time per site in comparison with the NOAA standing stock protocol, and 40% of the time in comparison with the accumulation protocol. Thus, from the perspective of implementing a more formal large-scale survey, there may be some efficiencies and improvement in accuracy in adopting some aspects of the CSIRO method.

The technical appendices to the report present results from an in-depth examination of a wide range of factors affecting debris loads at coastal and inland sites. These analyses were a necessary part of developing the supporting information to evaluate the key questions covered in the body of the report. However, as these variables were outside the scope of the key questions we have not provided an extensive discussion of the results or their implications. In brief, we did identify a number of important drivers of debris loads, which are similar across all four datasets we examined in this study. Debris loads increase with poverty, population and access, and are strongly affected by land use and vegetation type. Further investigation of these drivers of debris loads could be useful, as they could be used to identify key opportunities for intervention. The ICC, NOAA standing stock, and NOAA accumulation data were all also affected by the number of participants on a survey, with more people resulting in higher counts of debris per area surveyed. This result suggests that there are some aspects of survey effort that are not well controlled in these protocols. Again, the statistical analyses, particularly with some addition follow-up, could be useful in guiding improvements in design and analysis of the available data.

4.1 Specific Recommendations and next steps

4.1.1 Develop a national baseline

There is currently no nation-wide structured survey available as a baseline. While the ICC data does provide national coverage, it is to some extent hampered by compromises inherent in volunteer data collection. A nationwide survey designed using the principals presented here would require relatively little time and cost, and could provide a useful reference against which to compare trends, interventions, and geographic patterns. We suggest that the west coast surveys done as part of this study could be replicated on the Gulf and Atlantic coasts of the US at relatively minimal cost. The US west coast required roughly three weeks by two staff members to complete. This survey would be most usefully done on a periodic basis, perhaps every 5 to 10 years, to track major changes in debris with time.

4.1.2 Continuation of volunteer based data collection

While we have extensively covered the issues arising with data collected from volunteers during clean-up events, this data has also provided a rich source of information in our analysis. We suggest that a small number of investigations of the existing data and organizations, such as understanding how sites are chosen or how volunteers search a location, could significantly improve the value of this data. However, even in its current form without any modification the data from these volunteer efforts provides a rich picture of the debris load in the US.

4.1.3 Investigate the effect of drivers and responses to marine debris

We did not expound on the in-depth models we used for our analysis, and only covered the effect of local government and state policies at a cursory level. We suggest that a further expansion of the analysis of these drivers could provide useful information for understanding both the drivers of debris loads on the US coast and inland waterways and the data, and for targeting interventions through clean-ups, outreach, incentives, and regulation. The ICC data in particular could support a national analysis of bag ban, container deposit and other interventions, evaluating both their effectiveness and cost. Based on experience from a similar analysis in Australia, it is likely that effects of these interventions are discernible in the coastal debris data. Moreover, combined with data on cost, it would be possible to provide feedback to decision-makers on the return on investment for the various interventions available from adding garbage cans at beaches to prosecuting illegal dumping.

4.1.4 Understand the linkages between land-based activities and inputs to the marine system

Clearly much of the load of marine debris is a result of losses from land. Currently there is not a clear picture of the processes driving these losses and their relative importance. A key opportunity would be linking the available coastal data to data on loads and transport processes on land, building a better picture of the important processes and possible intervention points. In this project we investigated the role of outflows from rivers and estuaries in driving local debris densities. There is clearly a relationship, however, from a preliminary analysis the relationship appears to have some complexities. Understanding the role of riparian transport could assist in evaluating existing infrastructure for solid waste control, and identifying opportunities for effective investments. Furthermore, because survey sites in the NOAA data do not appear to have been chosen at random, there may be some underlying bias generating relationships between debris loads and river proximities. One option for addressing this question would be to set up a structured sampling system around river systems. This could be done as a single investigation at one point in time, and analysed specifically to understand the effect of rivers on debris loads.

4.1.5 Establish design parameters and a sampling system for a national monitoring program

Establishing the goals of a national monitoring system and creating a design that could achieve those goals could help increase the effectiveness of NOAA's investments in data collection and analysis. Currently the NOAA data collection efforts appear to be driven by a mix of local opportunities and historic programs. Designing a national monitoring system that could provide high quality data at a range of investment levels would mean NOAA could periodically put delivery of a national dataset out to tender, but with a clear picture of the likely person-hours required, the expected data structure and sampling design, and with a pre-existing analytical design and data management system. This would allow NOAA to establish a national baseline and implement periodic monitoring in a very cost-effective manner, and allow NOAA to deliver these outcomes in an inter-operable manner over time using the most cost-effective public or private sector providers. A fundamental issue to consider if one goal is to evaluate changes through time, is the currently lack of balance in sampling over time across sites. If a time trend is a key interest for program design, it will be critical to address this sampling issue.

4.1.6 Issues with current data and benefits of a national survey

A key outcome of using a structured survey would be the ability to control site selection and survey effort. Section 3 in the report provides a discussion of the benefits of randomization and controls on survey effort. One key benefit is the capacity to resolve a number of the factors that cause bias in the current NOAA and ICC data. For instance, in the analysis of the NOAA standing stock data there does appear to be some effect of the quality of the local group on the debris density recorded (see Figure 6.4.2.c and related text). Using a trained survey team following a clear protocol could resolve this issue. Similarly, in the NOAA Accumulation and the ICC data there is an effect of the number of people on the survey, with higher debris densities recorded when there are more people. This suggests that sampling is more thorough with more people, independent of the density at the site. While more thorough sampling is good in principal, it suggests that sites with fewer people are being under sampled. Ultimately the issue is that there is not tight control on sampling effort, so effort itself is driving the debris estimates, in addition to the dirtiness of a site.

In section 6 of our report, at the end of each analysis subsection we presented the output of a thorough analysis of the drivers of debris densities at each site. Improving the quality of the underlying data, through use of a designed survey, could allow these analyses to be improved, yielding more clear understanding of what drives the debris load. In addition, most of the US is not covered by the existing data. Only the US west coast has adequate coverage in the existing datasets, and only for the accumulation data. The NOAA standing stock data is limited to concentrated efforts, but in relatively few places. Having a comprehensive data set covering the continental US coast would allow the analysis to be extended to cover other regions outside the US west coast. This could be useful in understanding other emerging patterns, such as the very high loads suggested by the ICC data on the southern Texas coastline. Finally, there is currently no comprehensive national baseline for debris densities on the coast. This hampers monitoring change as policies or other interventions are implemented, in addition to more general tracking of the state of debris in the environment.

4.1.7 Timing, labor, and cost of a national survey

Roughly speaking, both NOAA's standing stock and the CSIRO transect method are relatively similar in their labor requirement. We provided a table in the report in section 3 giving the labor required per unit area. Assuming a sampling intensity on the order of what we used for the US west coast of approximately 1 survey site per 100 km, we estimate that a team could cover approximately 3 sites per day. On the west coast, this equated to roughly 2 weeks for the whole west coast using the CSIRO transect method. Assuming an 8 hour day, the CSIRO method with 2 people requires 45 minutes per site, which yields 2 hours and 15 minutes of survey work per day. Thus roughly 5 hours and 45 minutes are transit time and other activities. The NOAA standing stock method is slightly slower, requiring 0.6 more person hours per site, and thus to achieve the same efficiency one would need a 3rd person.

The coasts of Washington, Oregon, and California, (totaling around 14,160 km) were surveyed in roughly 10-12 days of fieldwork using the CSIRO method during the second phase of the project. Using this as a guide, the cost for a national survey can be estimated. The Gulf coast is roughly 26,000 km, and thus should require approximately 1.8 times the time to cover. The Atlantic coast is roughly 46,000 km, and thus should require 3.3 times as long. Based on these estimates, about 104-110 person-days would yield a dataset with a sampling site with three replicate samples every 100 km of coastline. Given a team of two, a national scale survey at 100km resolution would require 52 days.

Based on this estimate, one could make decisions about the frequency of sampling, and coverage, given a known budget. For example, it should be possible to get a scaled down dataset with a survey site every 300km (as driving costs are 5/8 of the time requirement doing 3 surveys per day, a reduction does not translate to an equivalent reduction in surveys). Using a pre-determined survey design and analysis methods, it would be possible to establish a national monitoring system that could be contracted out to providers, but which would provide structured and comparable data over time and locations, with a reasonably predictable cost. This would mean that NOAA could determine its monitoring goals (identify change over time, identify hotspots, etc.) and structure the survey tendering process to achieve those goals at reasonable cost.

5 References

Hardesty BD, TJ Lawson, T van der Velde, M Lansdell, G Perkins and C Wilcox. 2016. Estimating quantities and sources of marine debris at a continental scale. *In press,* Frontiers in Ecology and the Environment.

Hardesty BD, C Wilcox, TJ Lawson, M Lansdell and T van der Velde. 2014. Understanding the effects of marine debris on wildlife. A Final report for Earthwatch Australia. http://www.csiro.au/Organisation-Structure/Flagships/Wealth-from-Oceans-Flagship/marine-debris.aspx.

Lippiatt, S., Opfer, S., and Arthur, C. 2013. Marine Debris Monitoring and Assessment. NOAA Technical Memorandum NOS-OR&R-46.

Opfer et al, 2012. Marine Debris Shoreline Survey Field Guide, U.S. Department of Commerce National Oceanic and Atmospheric Administration, National Ocean Service, Office of Response and Restoration, Marine Debris Program, https://marinedebris.noaa.gov/sites/default/files/ShorelineFieldGuide2012.pdf.

Smith, S. D. A. and A. Markic 2013. Estimates of marine debris accumulation on beaches Are strongly affected by the temporal scale of sampling." PLOS ONE **8**(12): e83694.

USGS, 2014. Small-scale Dataset – Railroad Stations of the United States, https://catalog.data.gov/dataset/usgs-small-scale-dataset-global-map-railroad-stations-of-the-united-states-201403-shapefile.

USGS, 2015. Small-scale Dataset - Major Roads of the United States 199911 Shapefile, https://catalog.data.gov/dataset/usgs-small-scale-dataset-major-roads-of-the-united-states-199911-shapefile.

Wilcox, C. N Mallos, GH Leonard, A Rodriguez and BD Hardesty. 2015. Estimating the consequences of marine litter on seabirds, turtles and marine mammals using expert elicitation. Marine Policy.

6 Appendices

This appendix section is divided into a multitude of sections in which the details that inform or were used for analyses presented in Sections 2 and 3 are contained.

6.1 Data

A number of different data sets and data types were used in various components of the analysis to answer each of the two objectives. In addition to survey data from NOAA, OC and CSIRO (see details in Section 6.1.1 below), we also collated information that characterises the actual survey locations. This includes features such as roads, rivers, railways, watersheds, land use/land cover, and human population. We included these data in modelling debris or order to assess some of the potentially important factors or drivers that may help to explain any variation in amounts of debris reported at different survey sites (details in Sections 6.2 and 6.3 below).

6.1.1 Survey Data

Site debris survey and clean-up data was sourced from Ocean Conservancy (OC) and the National Oceanic and Atmospheric Administration (NOAA). Due to the distinct nature of the data, it was split into three sets of data to be analysed individually (Table 6.1.1.a). More details on the survey procedure of NOAA may be found in NOAA's field guide (Lippiatt et al., 2013). ICC data collection protocols can be found at http://act.oceanconservancy.org/site/DocServer/ICC_Eng_DataCardFINAL.pdf?docID=4221.

Table 6.1.1.a Sources and types of survey data used in analysis

Data source	**Description**
Ocean Conservancy	International Coastal Clean-ups (ICC)
National Oceanic and Atmospheric Administration	Accumulation surveys
National Oceanic and Atmospheric Administration	Standing stock surveys

The entire survey data consists of 16,486 records across all three data sources. This includes 6,379 site locations with some sites being surveyed at multiple times spanning 2009 to 2015. Sites were each given a unique identifier which was carried throughout the analysis and used to identify sites. The data was projected into World Mercator and all distances and areas were calculated in kilometres unless otherwise noted. Various covariates were collected to answer questions about what important factors correlate with or influence the volume and mix of debris identified in the surveys.

Table 6.1.1.b Details of survey data used in analyses from NOAA and Ocean Conservancy

Site Type	# of unique locations/sites	# of survey dates	Date range
Accumulation	284 (unique)	894	Jan 2012 – Aug 2016
	1443 surveys over multiple dates		
Standing Stock	66 (unique)	372	July 2009 – Aug 2015
	826 surveys over multiple dates		
ICC	6223 (unique)	517	June 2010– Oct 2015
	12822 (over multiple dates)		

6.1.2 International Coastal Clean-ups

The Ocean Conservancy coordinates an annual voluntary debris clean-up known as the International Coastal Cleanup (ICC). Debris is collected, categorized, counted and weighed on a voluntary basis each year at a variable number of sites around the globe. For this study only data from the United States were assessed. With the ICC data, sites are visited and cleaned. These sites are of varying sizes with a varying number of participants. Some sites may be revisited each year, but not all sites are revisited annually, nor are the exact same sites necessarily cleaned up each year, though there is some consistency between years. More information on the ICC may be found at http://www.oceanconservancy.org/our-work/international-coastal-cleanup/do-it-yourself-cleanup-tool.html.

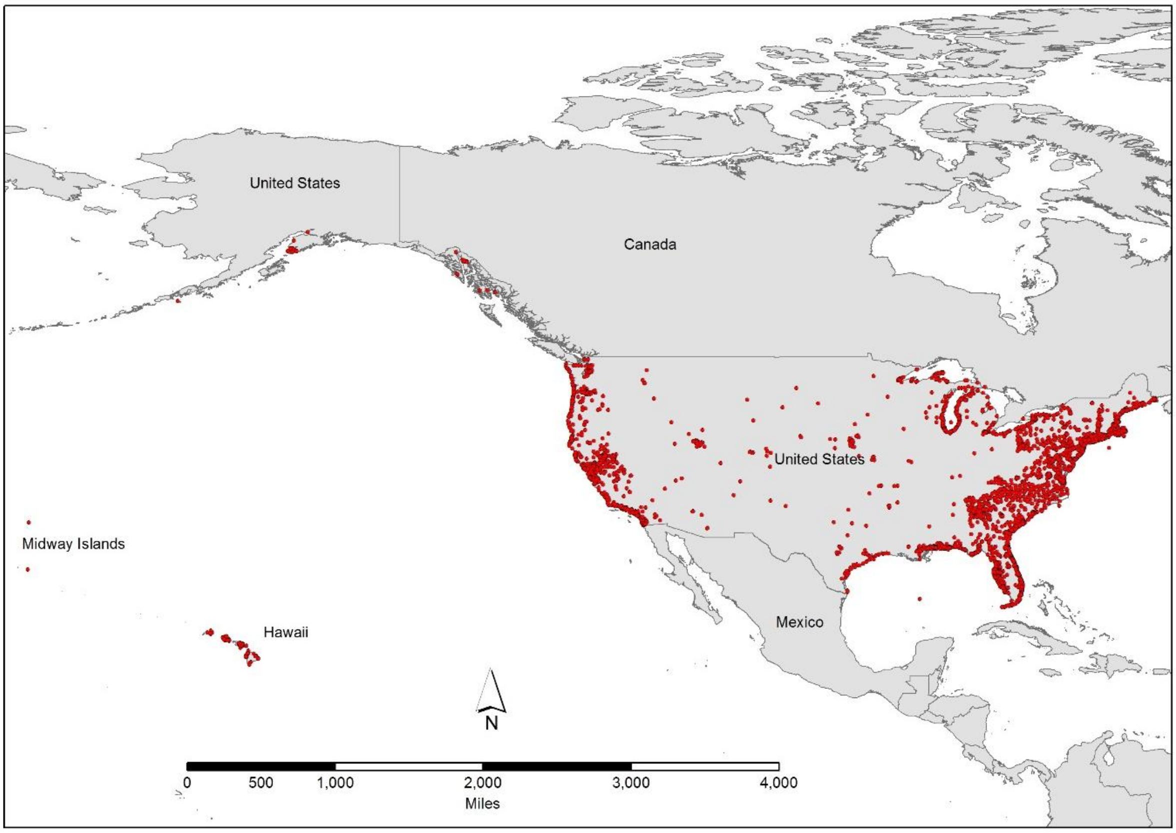


Figure 6.1.2.a Location of ICC survey sites

Table 6.1.2.a Number of ICC surveys by year and state

State	***2010***	***2011***	***2012***	***2013***	***2014***	***2015***
Alabama	25	23	23	30	31	0
Alaska	21	11	9	1	4	0
Arizona	5	1	1	4	3	0
Arkansas	5	3	2	2	1	0
California	628	598	401	482	535	2
Colorado	2	2	2	4	0	0
Connecticut	58	54	40	37	53	0
Delaware	36	38	38	42	46	0
District of Columbia	4	2	3	0	3	0
Florida	417	351	481	502	392	1
Georgia	120	141	158	128	160	1
Hawaii	69	55	62	54	66	0
Idaho	2	1	2	3	1	0
Illinois	62	48	57	46	53	0
Indiana	23	14	19	13	17	0
Iowa	3	0	0	0	0	0
Kansas	2	2	2	2	2	0
Kentucky	0	0	1	1	0	0
Louisiana	24	24	23	4	30	0
Maine	6	12	41	46	51	0
Maryland	23	30	27	36	35	0
Massachusetts	66	53	80	83	78	0
Michigan	102	93	111	111	107	0
Minnesota	13	7	19	28	6	0
Mississippi	55	76	69	52	0	0
Missouri	2	2	4	2	2	0
Nebraska	17	15	10	20	18	0
Nevada	2	1	0	1	1	0
New Hampshire	32	35	27	29	32	0
New Jersey	52	91	13	55	52	0
New Mexico	2	0	0	1	0	0
New York	207	174	151	159	203	1
North Carolina	79	92	81	85	82	0
Ohio	13	13	21	48	47	0
Oklahoma	2	0	0	0	0	0
Oregon	44	85	14	47	93	2
Pennsylvania	54	78	97	78	109	0
Rhode Island	79	78	71	47	78	0
South Carolina	99	100	43	67	51	1
South Dakota	1	1	1	2	1	0
Tennessee	0	0	8	5	0	0
Texas	33	35	31	30	35	1
Utah	11	17	25	14	1	0
Vermont	0	2	4	1	10	0
Virginia	150	123	147	126	127	1
Washington	22	14	5	31	40	0
West Virginia	1	1	1	0	0	0
Wisconsin	43	38	38	31	44	0
Wyoming	1	2	1	0	0	0

6.1.3 Accumulation surveys, NOAA

Accumulation surveys collected over time provide an estimate of the flux of debris onto the shoreline (in units of #items/m2 /time). This is achieved by recording and removing all debris from a site at a particular frequency. Because of the repeated survey approach, these surveys are resource intensive. NOAA recommends that some consistency exists in the timing of accumulation surveys at a site. The methodology calls for a single transect per site, with all debris to be removed at each survey, and for sites to be visited repeatedly. Their guidelines suggest that sites are surveyed daily, over 12 days, or monthly over the year. The timing is influenced by the study lifetime and objectives for the data collected and resources available. Accumulation surveys have been used to look for a spike in debris deposition from major debris-generating events or variations due to climactic events.

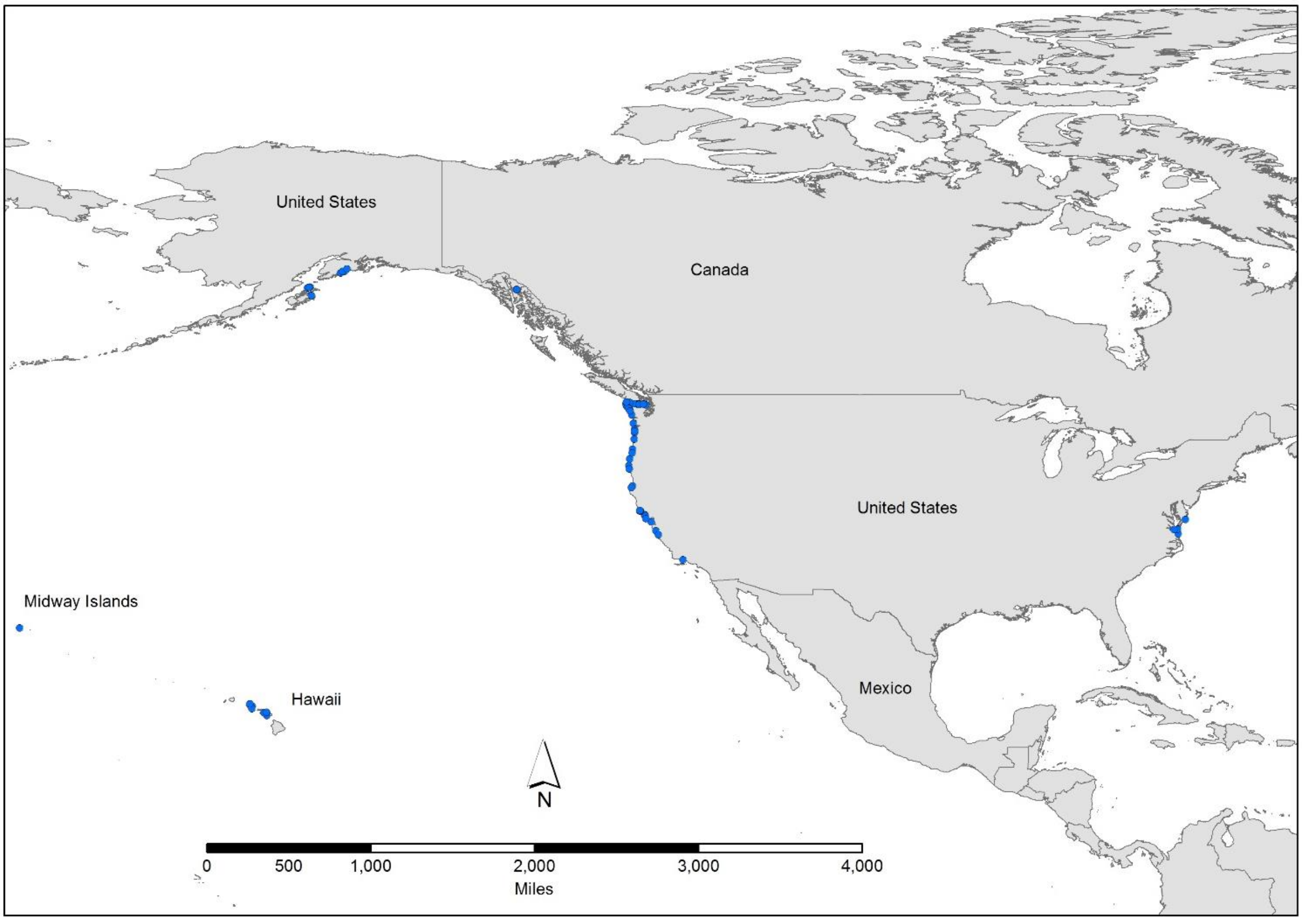


Figure 6.1.3.a Location of accumulation survey sites (NOAA)

Table 6.1.3.a Number of Accumulation surveys by year and state

State	***2012***	***2013***	***2014***	***2015***	***2016***
California	27	139	95	78	53
Oregon	11	21	41	48	33
Washington	45	182	207	130	88
Alaska	0	5	6	18	8
Virginia	0	1	36	25	0
Texas	0	0	0	7	1
Hawaii	37	65	17	6	1

6.1.4 Standing stock surveys, NOAA

Standing stock surveys collect information about the amount and type of debris within discrete transects along a shoreline. Survey data integrity is maintained by not removing debris from the site. Replicate transects of 5m in length occur within sites. Typically, four of twenty transects are randomly selected and surveyed each sampling period. These data are used to determine the density (# of items per unit area) of debris present. These surveys are less intrusive and resource intensive than accumulation surveys, as observers report or record information about debris items observed but do not remove an litter or debris found. Debris density reflects the long-term balance between debris inputs and removal and is important to understanding the overall impact of debris. This data is used to characterise drivers for debris deposition and attrition.

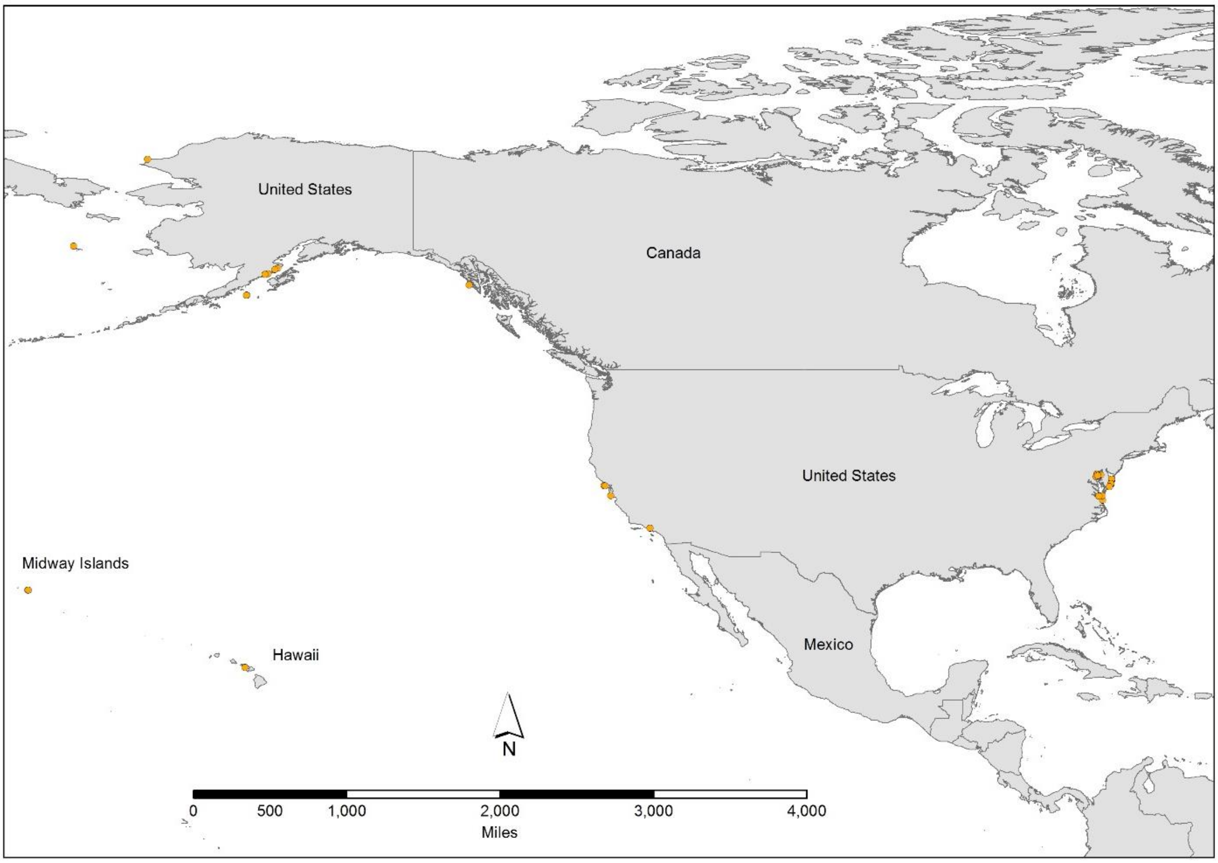


Figure 6.1.4.a Location of standing stock survey sites

Table 6.1.4.a Number of Standing-stock surveys by year and state

State	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Alaska	0	0	0	101	60	12	12
California	0	0	0	108	230	223	140
Delaware	0	0	432	0	0	0	0
Hawaii	0	0	0	0	16	57	11
Maryland	12	36	504	0	2	0	0
Virginia	0	0	0	0	0	148	80

6.1.5 CSIRO coastal debris surveys

To address each of the two objectives for this project, we also carried out a stratified random coastal debris survey along the west coast of the United States in July 2016 (Table 6.1.5.a). We selected the initial survey site to correspond with a NOAA site to the west of Seattle Washington and then selected sites approximately every 100 km down the west coast of the US to San Diego (See Figure 6.1.5.a). At each site we recorded the GPS location where we accessed the site, date, observer, weather conditions, wind speed and direction, human visitors visible on the beach and time of day. For each transect we recorded the time it took to carry out each transect, the transect start and end location and the transect length. To account for factors that may affect debris deposition and retention, we recorded the coastline shape and aspect, substrate type and color, gradient and backshore type at each transect (Supplementary Information). We surveyed 2 one-meter belts (one observer per belt), along each transect running perpendicularly from the waterline to the endpoint 2 meters into the terrestrial vegetation above the coastal zone. Only items detectable on the surface from head height were recorded. We recorded all items observed, along with the material type and color. For further detail see Hardesty et al. 2014; 2016 in press).

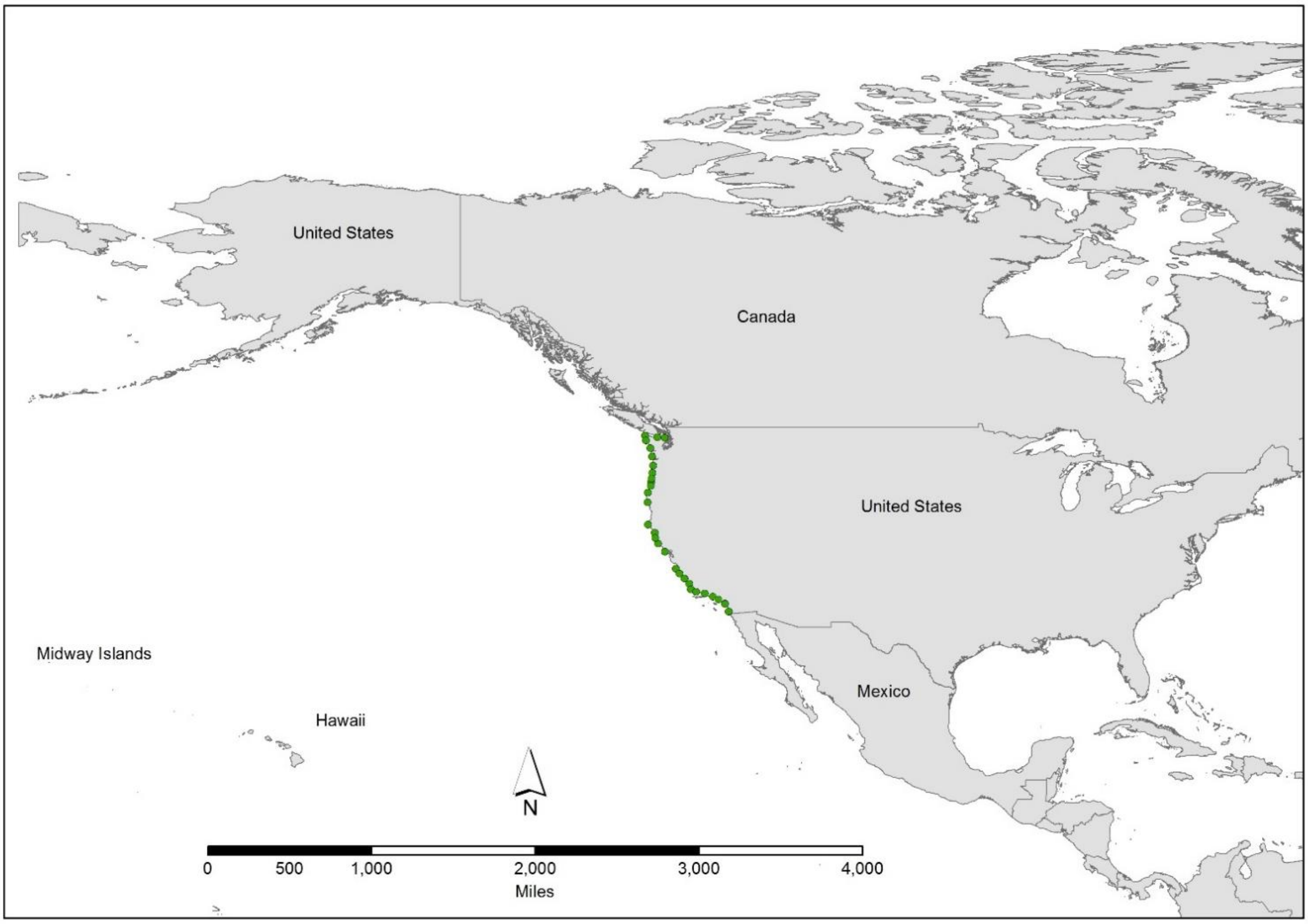


Figure 6.1.5.a Location of CSIRO coastal debris survey sites (2016)

Table 6.1.5.a Number of CSIRO sites (total number of surveys) by year and state, with total count of items.
Note that the total count of items in California is driven by a high number of items from a single site (near the river mouth at Long Beach).

State	**2016**	**Total item count**
California	19 (62)	5850 (4629 items from one site)
Oregon	7 (21)	482
Washington	2 (6)	370

6.2 Site characterisation

Data was sampled to characterise each site and to quantify if they have an influence on debris load and type. A single distance to features was calculated or the density of attributes was collected using the covariate sample design presented in 6.2a. Covariate segments are concentric circles with a radius of a given distance (1km, 5km, 10km, 25km, and 50km) from the survey sites. This allows for an analysis of the distance at which site characteristics have an influence on debris type and volume. Where data can be safely be assumed to be constant over the time that the surveys occurred, the most reliable data source was used. For data that changes through time, data was derived from a linear model of multiple samples across time for the day each survey occurred.

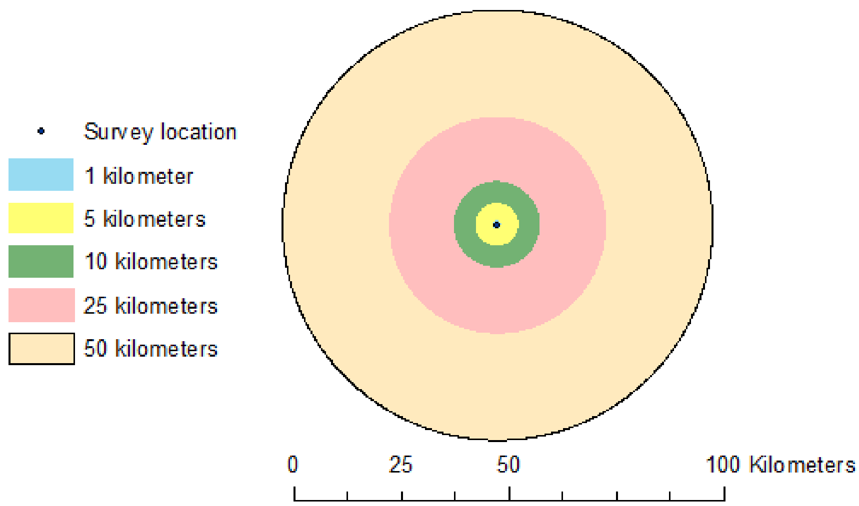


Figure 6.2.a Covariate sampling design depicting the concentric distance around survey points for which information was collated or summaried for inclusion in modelling and statistical analyses.

6.2.1 Roads

We used the distance to the nearest road as a proxy for the potential number of people accessing sites. To do this, we determined the distance in kilometres to the nearest road. Sites were further characterised by the amount of road within the covariate sample segments (within 1km, 5km, 10km, 25km and 50km). The data used in this analysis is the USGS Major Roads dataset (USGS, 2015) and is produced by joining the individual state roads layers. The dataset provides detail on road type (Table 6.2.1.a). This allowed for the amount of each type of road within each covariate segment to be calculated.

Table 6.2.1.a USGS Major Roads dataset functional road classes

Functional Road Class (FRC) ID	Description
0	Freeway or Other Major Road
1	Major Road Less Important than a Freeway
2	Other Major Road
3	Secondary Road
4	Local Connecting Road
5	Important Local Road

6.2.2 Railways

As another proxy for determining the accessibility and presence of people near each site, we looked at the proximity to railway stations. For this, we collated information on the distance in kilometres to the closest railway station from each survey site. The rail data used for this analysis was the Railroad Stations dataset (USGS, 2014) collected as part of the United States Geological Society (USGS) Small-scale Dataset. This map layer includes Global Map data showing Amtrak intercity railroad terminals in the United States. The data are a modified version of the National Atlas of the United States data set of railroad and bus passenger stations of the United States. There are no Amtrak stations in Alaska or Hawaii.

6.2.3 Rivers

To examine if the outflow of a river has any influence on the amount of debris at a particular site we looked at the proximity of survey sites to river mouths or river outlets nearby. To do this, we calculated the distance in kilometres to the nearest river mouth from each survey location. The river layer used for this analysis was the USGSs (2010) USA Rivers and Streams. To determine where the river mouth was we intersected the rivers layer with the coast and where the 2 layers met within 200m (to account for spatial variance of the two layers) of each other we called a river mouth. This point was then used as the feature to which the closest distance from the survey sites was calculated.

6.2.4 Watersheds

The watersheds of the US are subdivided into 6 nationally consist levels of detail. For this analysis we looked at 2 levels. The 8 digit Watershed Boundary Dataset (Natural Resource Conservation Service (NRCS, 2013) version (4^{th} level) and the 10 digit Watershed Boundary Dataset (NRCS, 2013 version) (5^{th} level). We used these watershed levels to look at the population within each watershed boundary. The purpose of including this watershed factor in models was to determine the potential influence of terrestrial (land-based) input into the rivers and hence potential for the flow-on effects on coastal sites.

6.2.5 Land cover

To determine what the land use at each survey site we accessed the land cover as described by the National Land Cover Database 2011 (NLCD, 2011). The National Land Cover Database 2011 (NLCD) is the most recent national land cover product created by the Multi-Resolution Land Characteristics (MRLC) Consortium. NLCD contains 16 land cover classifications (Table 6.2.1.a), applied consistently across the United States at a spatial resolution of 30 meters. NLCD is based primarily on a decision-tree classification of Landsat satellite data (circa 2011).

Conterminous United States

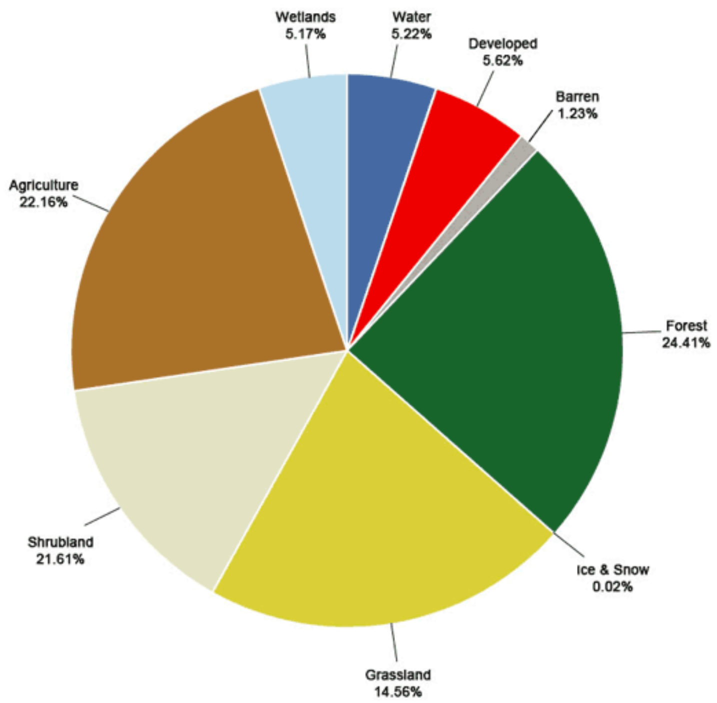


Figure 6.2.5.a A generalized summary of the main NLCD and cover classes for the conterminous United States.
(Note: Some NLCD land cover classes have been grouped for display purposes. A detailed conterminous United States class proportions are described in Table 6.2.5.a.

Table 6.2.5.a Conterminous National Land Cover Database 2011 classifications and distribution

NLCD Land Cover Class	**Percentage**
11. Water	5.22
12. Perennial Ice Snow	0.02
21. Developed, Open Space	3.26
22. Developed, Low Intensity	1.46
23. Developed, Medium Intensity	0.68
24. Developed High Intensity	0.22
31. Bare Rock/Sand/Clay	1.23
41. Deciduous Forest	10.84
42. Evergreen Forest	11.56
43. Mixed Forest	2.00
52. Shrub/Scrub	21.61
71. Grasslands/Herbaceous	14.56
81. Pasture/Hay	6.65
82. Cultivated Crops	15.51
90. Woody Wetlands	3.87
95. Emergent Herbaceous Wetlands	1.30
Total	100.00%

Note: This table is for illustrative purposes only, we encourage you to download the most recent NLCD data file to complete detailed land cover analysis.

Alaska

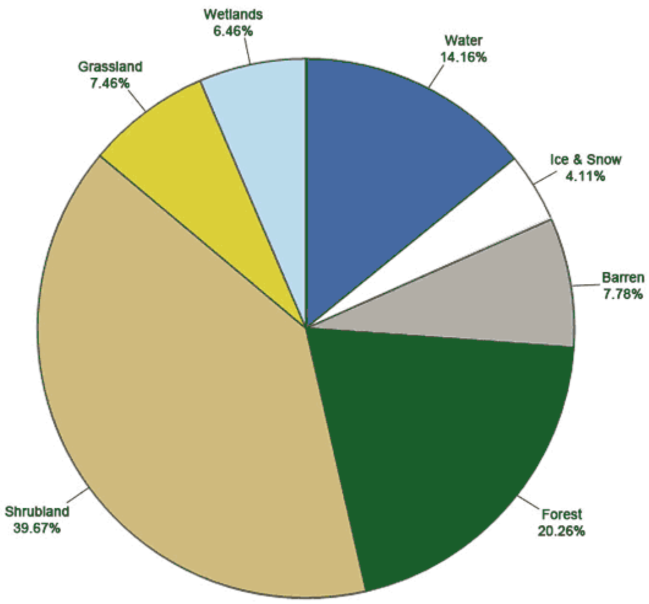


Figure 6.2.5.b. A generalized summary of the main NLCD2011 land cover classes for Alaska. (Note: Some NLCD2011 land cover classes have been grouped for display purposes. A more detailed description of Alaska class proportions is described in Table 6.2.5.b).

Table 6.2.5.b Alaskan National Land Cover Database 2011 classifications and distribution

NLCD2011 Land Cover Class for Alaska	Percentage
11. Open Water	14.16
12. Perennial Ice Snow	4.11
21. Developed, Open Space	0.02
22. Developed, Low Intensity	0.06
23. Developed, Medium Intensity	0.01
24. Developed High Intensity	0.00
31. Bare Rock/Sand/Clay	7.78
41. Deciduous Forest	3.24
42. Evergreen Forest	13.75
43. Mixed Forest	3.27
51. Dwarf Shrub	17.10
52. Shrub/Scrub	22.57
71. Grasslands/Herbaceous	1.72
72. Sedge/Herbaceous	5.71
74. Moss	0.03
81. Pasture/Hay	0.00
82. Cultivated Crops	0.02
90. Woody Wetlands	3.48
95. Emergent Herbaceous Wetlands	2.98
Total	100.00%

Note: This table is for illustrative purposes only, we encourage you to download the most recent NLCD data file to complete a more detailed land cover analysis.

6.2.6 Population and building density

Population, building density and socioeconomic factors were evaluated for each of the covariate segments for the date of each survey at each site. The highest spatial and temporal resolution data available for this task is the American Community Survey operated by the United States Census Bureau (USCB). The American Community Survey (ACS) is a mandatory, ongoing statistical survey that samples a small percentage of the population every year which is extrapolated across the country. The annual national data uses a continuous measurement methods. In this survey, a series of monthly samples produce annual estimates for the same small areas (block groups) formerly surveyed via the decennial census long-form sample. Boundaries are defined by the USCB's Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) database. The geodatabases include information for the fifty states, the District of Columbia, Puerto Rico, and the Island areas (American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, and the United States Virgin Islands). ACS is available as 1, 3 and 5 year estimates, each of these datasets using a different basis and extent to determine values.

Values from the 5-year estimate ACS dataset were used in this study for several reasons:

1. The longest period of data is used (60 months) as a basis for value,
2. There is data for all areas (block groups),
3. It contains the largest sample size, and
4. The values are the most reliable, however are also the least current

For this study the following data were gathered:

Table 6.2.6.a Variables, codes and descriptions used for the 5 year estimated American Community Survey (ACS) in this study.

Variable	**ACS Code**	**Definition**
Population	B00001e1	Total count of population
Housing	B00002e1	A housing unit may be a house, an apartment, a mobile home, a group of rooms or a single room that is occupied (or, if vacant, intended for occupancy) as separate living quarters. Boats, recreational vehicles (RVs), vans, tents, railroad cars, and the like are included only if they are occupied as someone's current place of residence.
Median age	B01002e1	The median age is the age that divides the population into two equal-size groups. Median age is based on a standard distribution of the population by single years of age and is shown to the nearest tenth of a year.
Median earnings	B20002e1	Earnings represent the amount of income received regularly for people 16 years old and over before deductions for personal income taxes, Social Security, bond purchases, union dues, Medicare deductions, etc. The median earnings represent the value that divided the standard distribution of earnings in two. Values are in units of United States dollars ($US)
Workforce	B23025e2	This category includes all civilians 16 years old and over who either (1) were “at work,” that is, those who did any work at all during the reference week as paid employees, worked in their own business or profession, worked on their own farm, or worked 15 hours or more as unpaid workers on a family farm or in a family business; or (2) were “with a job but not at work,” that is, those who did not work during the reference week but had jobs or businesses from which they were temporarily absent due to illness, bad weather, industrial dispute, vacation, or other personal reasons. Excluded from the employed are people whose only activity consisted of work around the house or unpaid volunteer work for religious, charitable, and similar organizations; also excluded are all institutionalized people and people on active duty in the United States Armed Forces.
Unemployed	B23025e7	All civilians 16 years old and over are classified as unemployed if they (1) were neither “at work” nor “with a job but not at work” during the reference week, and (2) were actively looking for work during the last 4 weeks, and (3) were available to start a job. Also

		included as unemployed are civilians who did not work at all during the reference week, were waiting to be called back to a job from which they had been laid off, and were available for work except for temporary illness. Examples of job seeking activities are: • Registering at a public or private employment office • Meeting with prospective employers • Investigating possibilities for starting a professional practice or opening a business • Placing or answering advertisements • Writing letters of application • Being on a union or professional register
Population poverty	B17021e1	The data on poverty status of households were derived from answers to the income questions. Since poverty is defined at the family level and not the household level, the poverty status of the household is determined by the poverty status of the householder. Households are classified as poor when the total income of the householder's family is below the appropriate poverty threshold. (For nonfamily householders, their own income is compared with the appropriate threshold.) The income of people living in the household who are unrelated to the householder is not considered when determining the poverty status of a household, nor does their presence affect the family size in determining the appropriate threshold. The poverty thresholds vary depending on three criteria: size of family, number of related children, and, for 1- and 2-person families, age of householder
Population male	B01001e2	Total count of male population
Population female	B01001e26	Total count of female population
Median age male	B01002e2	Median age of males
Median age female	B01002e3	Median age of females

6.3 Analytical approach

Our mathematical approach to analysing the NOAA and ICC data incorporated the various categories of items and their relative frequency of occurrence at replicate beaches. We used a number of different models in analysing the data, depending on the particular question which was being addressed. Generally speaking, from the data we want to know the following:

- What is the baseline quantity of litter at a site?
- Is the amount of debris at a site changing with time?
- Are there different types of sites?
- Can we make predictions about why or how sites may change?

We investigated patterns in the total load of debris in three datasets, NOAA's Accumulation data (ACC), NOAA's Standing Stock data (SS), and the International Coastal Cleanup data (ICC). In each case we focused on three questions: 1) what is the average load at a site, i.e. what would a baseline look like, 2) how does this load differ by location, time, and other explanatory variables, 3) which regions or sites appear to be the best or worst in terms of how much debris is observed (by count)? We considered a number of issues in our analysis. First, the data are composed of counts of items, weights of items, or volumes of items. Therefore, any model of the data needs to accommodate the statistical nature of the data, that is, it is bounded between zero and infinity, and in some cases it may be composed of integers. Second, sampling effort will likely be strongly related to load, and thus our models need to take sampling effort into account. Third, there may be historical factors that are important, such as the time since the last clean up or survey. This is particularly true for the ACC and ICC datasets, as both involve removal of items.

Models incorporated date (year and month), distance to nearest city, and a number of beach characteristics (e.g., beach shape, beach slope, direction of dominant winds, and beach substrate type). We independently applied these models to NOAA and Ocean Conservancy data sets to uncover unique insights from each approach.

In Phase 2 of the project we conducted fieldwork on the west coast of the US, in cooperation with staff from NOAA and Ocean Conservancy. During this fieldwork we collected additional data using a structured survey design developed by CSIRO, along with sampling paired sites using the NOAA standing stock protocol. With this additional data in hand, we then revisited the load modelling described in Section 6.4 a second time. We used similar models to those described above to analyse the new CSIRO protocol data, and then revisited the NOAA and ICC data analysis with an expanded set of variables as described in Section 6.2. These analyses are presented below, as an addendum to the preliminary models developed during Phase 1 for the NOAA and ICC data sets.

6.4 Modelling Debris Loads

This section presents the key results from the analysis of the NOAA and ICC data sets. Each of the subsections presents the preliminary analysis of the data done during Phase 1 of the project, with an addendum containing the final analysis outputs based on revisiting the datasets during Phase 2 of the project. The CSIRO data analysis differs, in that it was only available during Phase 2, and is thus condensed into a shorter section. As modelling the drivers of debris loads was not a key objective of the project we do not provide a detailed discussion of the final models and their implications. However, development of these models was necessary in order to control for the variation in the data to allow us to estimate the various quantities required to meet the key objectives. For instance, in estimating the effect of rivers on coastal debris loads, it is necessary to control for sampling bias, local land use, access, and a variety of other factors that could affect debris loads. Understanding the other driving variables allowed us to develop a more sensitive test for rivers for instance.

Table 6.4.a provides a framework for understanding which analyses were used for which data products.

Table 6.4.a GAM analyses and their associated results tables and data products.

Data set	Date range	Analyses completed	Data products from this analysis (page)	Results Table (page)
NOAA Accumulation	2012-2015	Preliminary analyses	Figure *6.4.1.b* (77)	Table 6.4.1.b (76)
		Full covariates	Figure 2.2.a (8)	Table 6.4.1.c (79)
	Summer 2016	Correlation analysis		
NOAA Standing stock	2009-2016	Preliminary analysis	Figure 6.4.2.b (83), Figure 6.4.2.c (84)	Table 6.4.2.b (82), Table 6.4.2.c (85)
		Full covariates		Table 6.4.2.d (86) Table 6.4.2.d (86)
	Summer 2016			
ICC	2010-2016	Preliminary analysis	Figure 2.3.b (23), Figure 2.3.c (25)	Table 6.4.3.b (89), Table 6.4.3.c (90)
		Full covariates	Figure 2.2.c (15) Figure 2.2.d (16), Figure 2.2.e (17)	Table 6.4.3.d (92)
CSIRO	2016	Full covariates	Figure 2.2.b (11)	Table 6.4.4.a (96)

For each of the final models developed for the datasets we present a statistical table giving the important variables included in the model (see Table 6.4.a).

These tables are all structured in a similar manner to allow evaluation of the results and comparison across driving variables and between datasets. Moving from left to right, the tables are labelled as follows.

<table>
<tr><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(>|t|)</th><th></th><th>Median value</th><th>Effect size</th></tr>
</table>

The estimate column gives the coefficient that was estimated for each variable. For instance, for a continuous variable such as “distance to road”, the Estimate column provides the coefficient that the distance is multiplied by in the best fitting model. For continuous variables, such as distances or population densities, the coefficients give the slope of linear relationship between a variable and the debris density. A negative coefficient means a negative slope to the relationship. For variables that are categories, the sign gives the effect of having that category present. For instance, if State is Oregon, then the coefficient presented is the effect on the density of debris of being in Oregon. The direction of the relationship is as one would expect, positive implies debris densities increase with an increase in the variable, negative implies a decrease. The term labeled intercept is the value the model would take when all continuous variables are 0, and the categorical variables are all at their reference levels. In order to estimate the effect of a categorical variable, one of the categories must be designated as the reference level, the coefficients of the other levels thus represent the deviation one would expect from the value at the reference level. The reference level for each categorical variable is specified in the table captions. The Std. Error column gives an estimate of the uncertainty in the value of the coefficient in the Estimate column. The larger the standard error, the more uncertain the value. The t value and Pr(>|t|) columns give the statistic and significance test for whether the parameter is significantly different from 0. By convention, if the significance test has a value less than 0.05 the parameter estimate is considered statistically significant (i.e. different from 0, and so important). The Median value column gives the median value of the variable that is relevant for the coefficient. This value, multiplied by the parameter in the Estimate column is displayed in the last column, labelled Effect size. This is a measure of the effect of the variable (at its median value) on the outcome. The Effect size can be used to measure the relative impact of the different variables included in the statistical model. The importance of the variables can be directly judged by the absolute value (i.e. excluding the sign) of the effect size. The sign gives a measure of the direction – i.e. given an increase in the variable, does the debris density increase (positive effect size term) or decrease (negative effect size term).

The generalized additive models used the in analysis also include what are known as smooth terms. These are functions used to approximate relationships with variables that are not easily represented using parametric terms as discussed above. These terms are described at the foot of each of the statistical tables, with the following column headings.

edf	Ref.df	F	p-value

In this case the terms generally include a two dimensional function of latitude and longitude, and sometimes other functions, such as a cyclic term for time of day or month of the year. These functions are constrained to have an average value of 0, and allow the model to incorporate deviations upward or downward from the model prediction in certain parts of the parameter space. For instance, the surface plots used to illustrate the hotspot analysis in Section 2.2 are visualizations of the two dimensional latitude and longitude smooth term in the relevant models. The first two columns, edf and Ref.df, give a measure of the actual and possible flexibility of the function used in the model. The F column gives a test statistic for the term, and the p-value column the measure of the statistical significance of the term. Again, by convention if the p-value column is less than 0.05 the term is considered to be significant.

6.4.1 Modelling total load of debris in the NOAA Accumulation data

NOAA's accumulation data is based on complete counts of items removed from a fixed site repeatedly over time. Survey sites include the west coast of the continental US, Alaska, three sites on the east coast of the continental US, and several sites in Hawaii (Table 6.4.1.a and Figure 6.1.3.a).

Table 6.4.1.a Accumulation survey total item count by year and state

State	***2012***	***2013***	***2014***	***2015***	***2016***
California	1997	15513	9280	11453	21055
Oregon	976	1099	3330	3342	225
Washington	3303	8998	9677	8259	8380
Alaska	NA	149	1368	3124	2218
Virginia	NA	29	1378	2444	3549
Texas	NA	NA	NA	4267	0
Hawaii	24833	50547	27226	2467	150

The accumulation data varies widely, with loads distributed between 0 and 375 items/m^2/day. Most observations are near zero, with a median value of 0.0002 items/m^2/day. There are a few high outlying values, resulting in a mean of 0.3503 items/m^2/day, much higher than the median. Rescaled using a log transform, the large number of small observations is clearly visible, with most observations below -5 on the log scale.

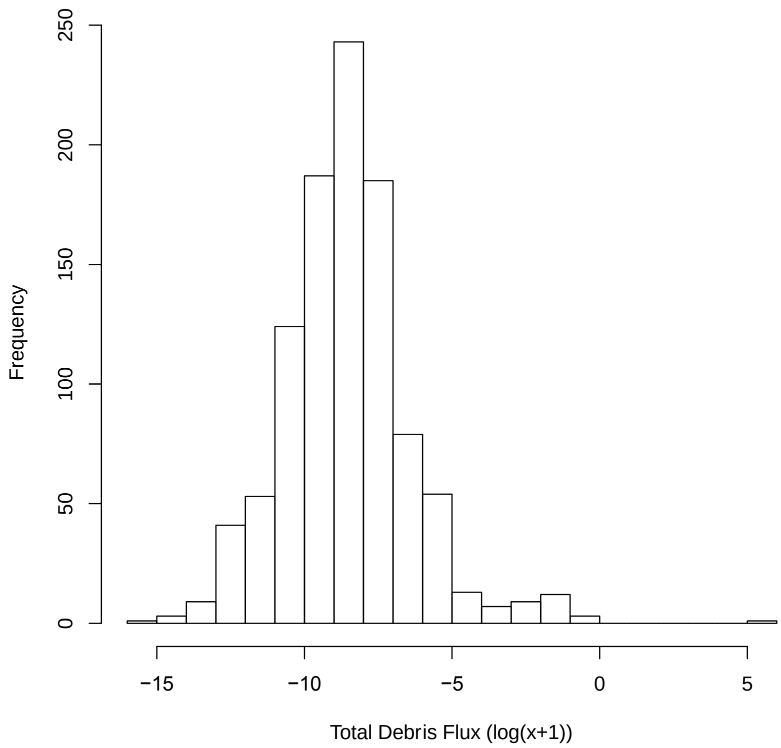


Figure 6.4.1.a Histogram of the total debris flux observations from the NOAA Accumulation data.

The total flux data is typical of count data, with most of the values at or near zero, and a few large values. This type of data typically requires a specialized distribution for analysis. In this case we use a negative binomial distribution, which is often used to model count data with long tails. The negative binomial distribution includes a rate parameter, typically the focus of modelling in regression, and a scale parameter that adjusts how variable the distribution is. In this case we first fitted a simpler model, incorporating a terms for state and county, which allowed us to estimate the scale parameter in the distribution. The optimal scale parameter was 0.162. We then used this scale parameter to build a more complex mode that incorporated state, and county.

To evaluate the contribution of a spatial component to the model, we limited the analysis to the west coast of the continental United States. In this region the sample intensity is relatively high, in comparison with the complexity of the geography. We represented the spatial structure in the data using the distance along the coastline, starting from the first sampling site in the Puget Sound area and moving west and then southward along the coast, ending at the sampling sites in the California Bight (Figure *6.4.1.b*, Figure 6.4.1.c, Other factors subplot).

We found that the three states along the west coast of the US differed significantly, with California having the highest debris densities, followed by Oregon, and subsequently Washington (Table 6.4.1.b, Table 6.4.1.c, Figure 2.2.a left ribbon). The spatial smooth for distance southward in the sampling region was highly significant, indicating that there is a significant spatial pattern in the data (Figure *6.4.1.b*, Figure 2.2.a right ribbon). Sampling sites in the Puget Sound and in the Strait of Juan de Fuca had elevated flux values, relative to other locations in Washington. Flux values declined moving southward along the coast to a minimum just north of San Francisco, California, and then increased southward toward the California Bight in southern California. It is important to note in interpreting these patterns that the spatial component models the residual variation across locations in the context of the state coefficients in the model. Thus, although the spatial surface shows high values in the Puget Sound area, overall California has higher flux values than Washington. The year term in the model was not significant, indicating there is no temporal trend in the flux data.

The two terms that we incorporated in to the model to address sampling effort, the period between samples and the number of people assisting in the sampling, were both significant. The amount of debris found in surveys decreases with the time since the preceding survey. This suggests that sites with small amounts of debris may be getting sampled less frequently than sites with larger amounts of debris. An alternative interpretation of this relationship could be that there is periodic resuspension or burial of debris that has been deposited on the coast. This mechanism would also be expected to lead to a negative relationship between flux and time between surveys. Indeed, Smith and Markic found that the amount of debris collected during monthly accumulation surveys was an order of magnitude less than the amount of debris collected in daily surveys (2013). However, the resuspension/burial mechanism would require that resuspension is episodic, potentially due to storms, while deposition is continuous. It might be possible to separate these hypotheses through a mixture of considering the ocean and atmospheric conditions between surveys, and discussion with survey staff as to scheduling of surveys at sites.

Table 6.4.1.b Parameter estimates for a statistical model of total debris flux in NOAA Accumulation samples from the west coast of the continental US. Flux values are log(x+1) transformed. Note that the term for State: California is incorporated into the intercept in the model. The median is the median value of the relevant covariate, multiplying it times the coefficient gives a measure of the effect size of each term. Factors can be taken to have a value of 1 using treatment contrasts, as in this case. Smooth terms in the model are constrained to have mean values of zero, and thus are best interpreted as deviations around the parametric components.

A. Parametric Covariates	Coefficient	Std. Error	p Value	Median	Median Effect
Intercept	-0.17	0.091	0.059	NA	-0.17
State:Oregon	-0.004	0.0012	0.00071	NA	-0.004
State:Washington	-0.0049	0.0013	0.00029	NA	-0.0049
Year	8.70E-05	4.50E-05	0.054	2014	0.18
Days since last survey	-1.70E-06	8.40E-07	0.046	29	-4.9E-05
(Number of Persons Assisting)^2	2.70E-06	4.50E-07	1.90E-09	2	5.4E-06

B. Smooth Components	edf	Ref.df	F	p-value
Month	4.7	8	1.3	0.045
Distance	8.9	9	15	3.0E-23

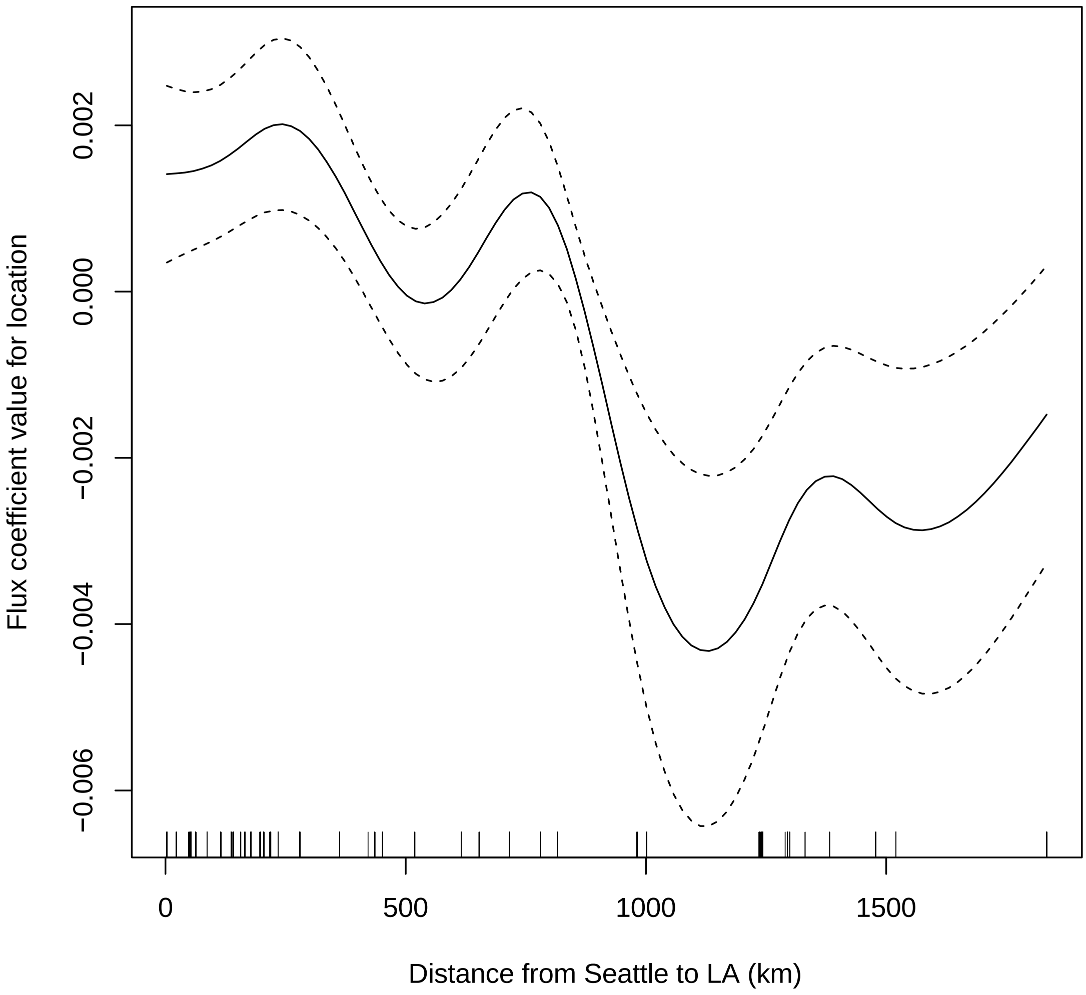


Figure 6.4.1.b *Spatial variation in the flux of debris along the sample sites on the west coast of the Continental US.*
The solid line is the coefficient for the effect of location on flux, while the dotted lines are the 95% confidence interval on the value of the coefficient at each location. The spatial component of the model is constrained to have a mean value of 0, thus this coefficient can reasonably be viewed as describing the deviation across the locations from the parametric components of the model. See Figure 6.1.4.a for a map of survey locations.

The second term we incorporated into the model to address sampling effort was the number of people participating in the survey. There was a small but significant positive relationship between the (second power of the) number of people and the density of debris recorded. Examination of the residuals from an initial model fit with a first order linear term suggested that the relationship between the number of people and the debris observed might be more than linear. We explored a more complex relationship using a smooth term, and concluded that a second order linear model might provide an adequate fit. Based on this exploration we compared a models including the number of people, the number of people squared, both terms, and a smooth of the number of people. Using AIC we found the best model to include the second power of the number of people participating (AIC N^2: -9913.9, N: -9907.4, N + N^2: -9912.5, s(N): -9912.6).

This result suggests there is a sampling effect, in which more observers at a site are able to identify more material within the same area. Moreover, additional observers at a site increase the count of debris at the site in a more than linear fashion. One possible mechanism for this phenomena may be saturation of observers. If survey sites have much more debris than an observer is able to find, within the constraints of the operation such as time, observer fatigue, observer selectivity, then adding additional observers would be expected to yield higher density measurements. This could be particularly true if larger debris items are less frequent than smaller ones. By definition of the lower cut off in sizes of items reported in NOAA data, this is certainly true. This is also likely true for ICC data as well (GL, NM, pers. obs.). In this case additional observers might be focusing on smaller items, which are more frequent than larger ones. A similar effect has been noted in other debris surveys, particularly where volunteers are involved (Hardesty et al. unpublished data), and may be related to a general pattern of shifting baselines. In this sense, the size distribution of items collected at a site sets a baseline for the search image that survey staff use – adding more staff reduces the baseline size as smaller things are identified, resulting in larger counts. One possible avenue for investigating this mechanism would be to look for a relationship between the size distribution of items at a site and the number of observers participating in the sampling. It would be necessary to control for the total amount of debris found at the sites in this case, as one would be trying to separate true rarity from saturation of the observers.

After fitting a model incorporating the full range of covariates, both state and year (from the original model) had extremely high effect sizes, but other factors, in particular certain land use categories, as well as rail distance and the fraction of people living in poverty within 50km of the site, also explained significantly amounts of the variability within the data. Notably, unlike in the simpler model, time since last survey was statistically significant, and positively correlated with debris, indicating that there is some measurable accumulation of debris between surveys. Thus the model can be used to estimate flux, as intended when the data collection protocol was designed. In addition, we found that the squared term for the number of observers was not included in the best model incorporating the full range of covariates, with a first order term resulting in a better fit between the model and data. As demonstrated here, adding explanatory variables can allow the model to better fit the data, simplifying some relationships and exposing underlying patterns in the data that are masked in simpler models (Table 6.4.1.c, Figure 6.4.1.c).

Table 6.4.1.c Table showing values for model of NOAA Accumulation Data (2012-2015). Distance to river and railway is included, state (OR, WA) are included, year, distance to roads, land use category and socio-economic factors. Factors included in the intercept are State: CA, and LandUse: Barren.

A. Parametric Covariates	Coefficient	Std.Error	p Value		Median	Median Effect
(Intercept)	-189.099	87.282	0.031	*	1.000	-189.099
River Distance KM	-0.005	0.005	0.321		7.496	-0.037
Rail Distance KM	0.063	0.007	0.000	*	104.955	6.583
State						
OR	-47.315	4.212	0.000	*	1.000	-47.315
WA	-45.448	3.865	0.000	*	1.000	-45.448
Year	0.111	0.043	0.011	*	2014.000	223.708
Roads 50km	-0.002	0.001	0.004	*	466.485	-1.030
Distance to nearest road	-0.146	0.027	0.000	*	0.959	-0.140
Land Use						
Developed, High Intensity	0.273	0.538	0.612		1.000	0.273
Developed, Low Intensity	2.863	0.423	0.000	*	1.000	2.863
Developed, Medium Intensity	7.920	1.722	0.000	*	1.000	7.920
Developed, Open Space	1.206	0.647	0.063	*	1.000	1.206
Emergent Herbaceous Wetlands	-0.508	0.250	0.042	*	1.000	-0.508
Evergreen Forest	2.118	0.294	0.000	*	1.000	2.118
Grassland/Herbaceous	2.052	0.427	0.000	*	1.000	2.052
Open Water	0.286	0.154	0.064		1.000	0.286
Shrub/Scrub	0.346	0.416	0.405		1.000	0.346
Woody Wetlands	3.406	0.569	0.000	*	1.000	3.406
Number of Persons Assisting	0.041	0.012	0.000	*	2.000	0.081
Population (50km)	0.000	0.000	0.000	*	36325.910	0.395
Watershed area (04)	0.000	0.000	0.000	*	7704.280	-1.307
Poverty fraction (50km)	-15.745	4.256	0.000	*	0.976	-15.363
Housing (50km)	0.000	0.000	0.019	*	2115.344	-0.522
Days since last survey	0.002	0.001	0.011	*	29.000	0.057
B. Smooth Covariates	edf	Ref.df	F		p-value	
s(InterpointDistance)	8.906	8.996	39.917		0.000	*
s(Month)	5.697	8.000	8.212		0.000	*

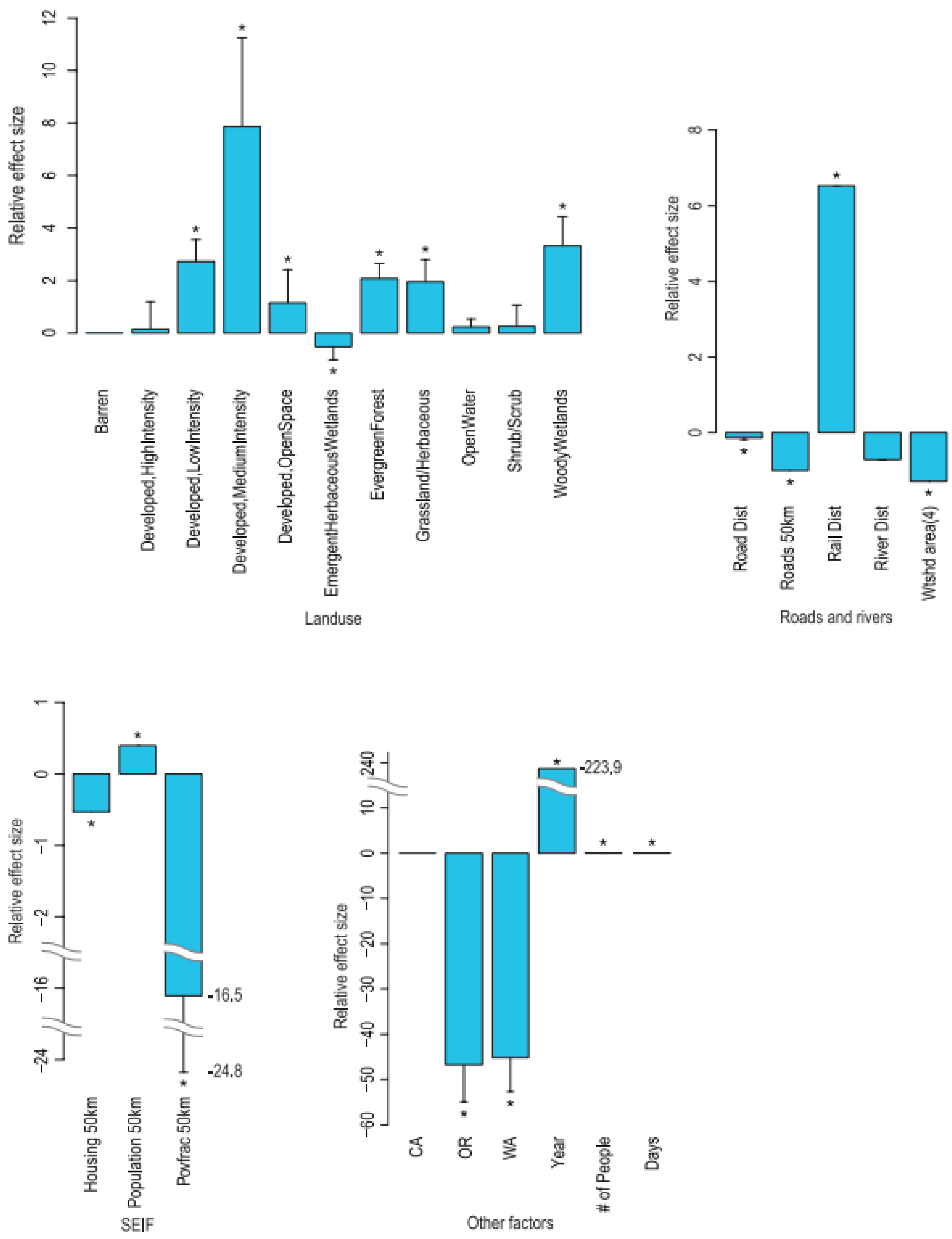


Figure 6.4.1.c Effect sizes for GAM modelling for NOAA Accumulation data (2012-2015)

Effect sizes give an indication of the relative importance of each model coefficient to the overall results of the model.

6.4.2 Modelling total load in the NOAA Standing Stock data

The NOAA standing stock data is collected following a different protocol from the NOAA Accumulation data. However, it displays similar characteristics in terms of its distribution, most observations are concentrated at the lower end of the values near 0, with a few observations at very high values.

Table 6.4.2.a NOAA Standing-stock survey total item count by year and state

State	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Alaska	0	0	0	1249	400	16	25
California	0	0	0	760	1750	2292	1112
Delaware	0	0	3468	0	0	0	0
Hawaii	0	0	0	0	614	1285	328
Maryland	744	1914	6366	0	293	0	0
Virginia	0	0	0	0	0	593	537

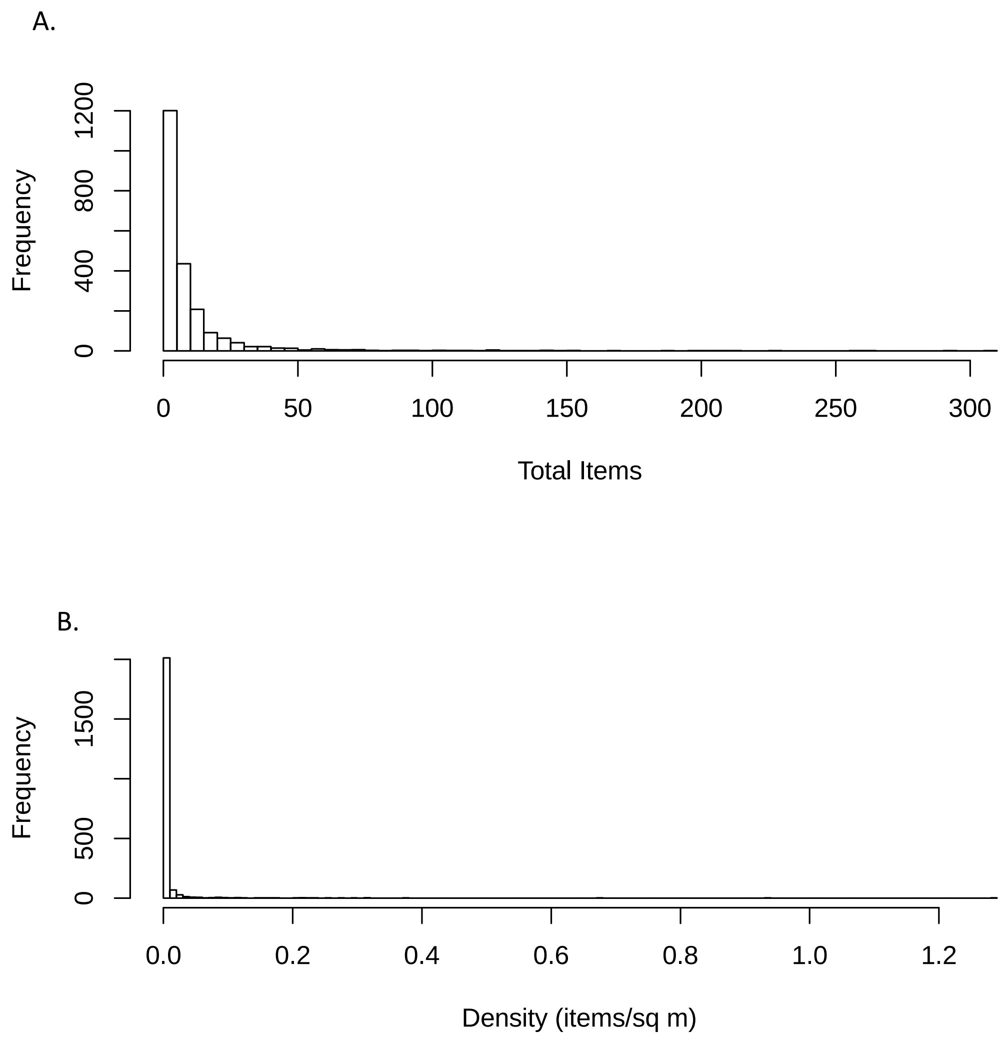


Figure 6.4.2.a Frequency distribution of observations in the NOAA Standing Stock data. Panel a) the frequency of counts of items in the standing stock surveys. Panel b) the frequency of densities of items, based on correcting the counts for the area surveyed.

Spatial coverage of the standing stock surveys is similar to the NOAA Accumulation surveys, although there is less comprehensive coverage of the west coast of the continental US (see Figure 6.1.4.a).

We fit a base model to the data, incorporating explanatory variables for state, month, year, and number of people assisting in the survey. The base model only explained 6% of the variation in the data, based on a deviance comparison against a null model (Table 6.4.2.b).

Table 6.4.2.b Statistical model of the density of debris recorded in NOAA Standing Stock surveys.

A. Parametric Covariates	Coefficient	Std. Error	p Value	Median	Median Effect
Intercept	1400	99	8.70E-44		1400
State: CA	0.36	0.14	0.0086		0.36
State: DE	-0.9	0.15	5.10E-09		-0.9
State: HI	1.5	0.18	6.80E-17		1.5
State: MD	0.56	0.15	0.00018		0.56
State: VA	-0.43	0.17	0.014		-0.43
Year	-0.7	0.049	3.50E-44	2012	-1408.4
Number of Persons Assisting	0.76	0.041	7.10E-73	1	0.76

B. Smooth Components	edf	Ref.df	F	p-value
Month	7.6	8	39	8.10E-62

However, there appears to be a pattern in the residuals with respect to the number of people assisting (Figure 6.4.2.c).

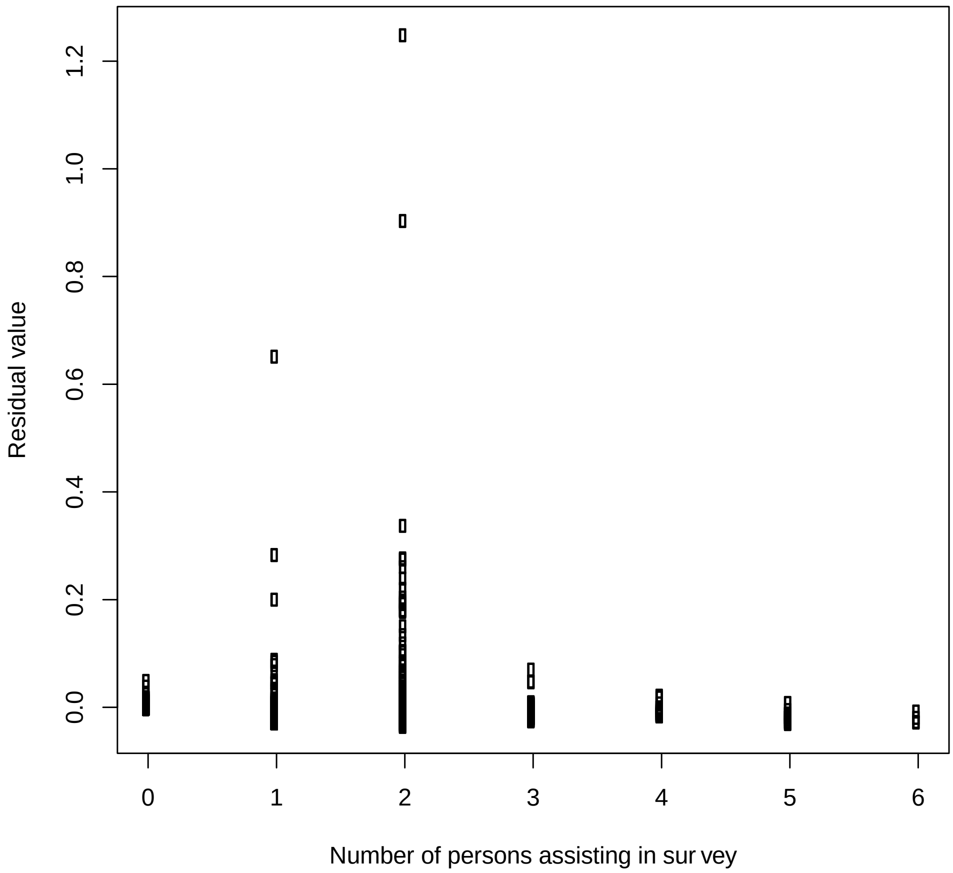


Figure 6.4.2.b Residual variation from the fitted model for the NOAA Standing Stock surveys. Note the variability when 1-2 observers are present, but the consistency when 0 or 3-6 observers are present. Presumably 0 observers means no additional observers aside from the person recording the data.

We refit the model using a smooth term to allow for variation in the results across the number of people assisting in a survey, and obtained a related result. There appears to be significantly low densities at sites where one person is assisting, and significantly high values where more than one person is assisting (Figure 6.4.2.c). It is likely that the pattern in this smooth term is resulting from the model using the number of people as a proxy for variation in survey quality among sites or some other variation in density across sites. One clue to this is that when this smooth component is included, the model explains a much higher proportion of the total variation in the data (48% based on a deviance comparison) and all of the parametric terms become highly significant (Table 6.4.2.c).

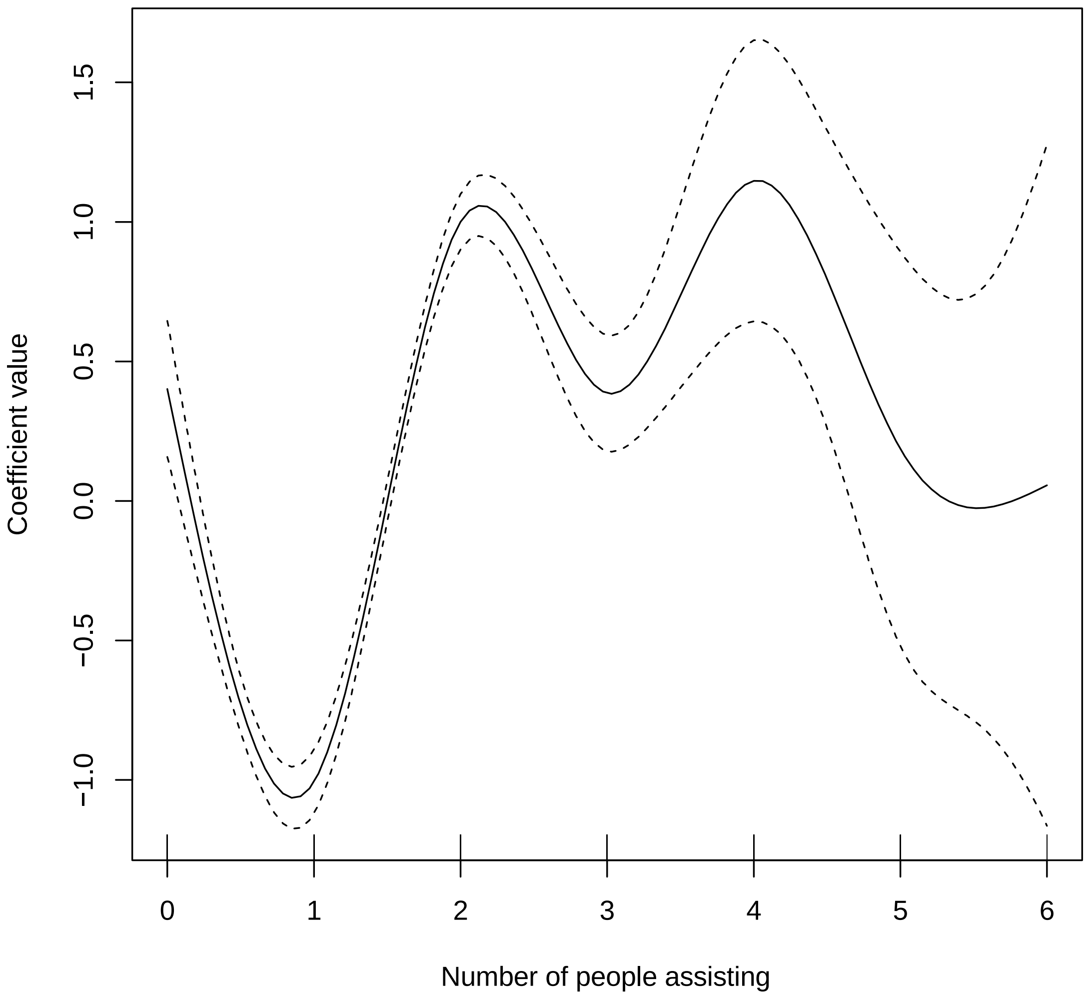


Figure 6.4.2.c Smooth component for number of people assisting in surveys from the base model for debris densities in the NOAA Standing Stock surveys.
The solid line is the value of the coefficient for the effect of the number of people on the debris density, the dotted lines are the 95% confidence intervals on the coefficient. The strong significant difference between 0 and 3 observers suggests that the smooth is acting as a proxy for some other driving variable. Sites differ significantly in the number of people assisting, with consistent numbers at many sites over time. Thus, the number of people assisting could be acting as a proxy for individual site identities.

We also constructed a unique identification for each site using its latitude and longitude. When this term was included in a model it was highly significant for 23 of the sites, and nonsignificant for the remaining 21. In this model, the smooth term for the number of people assisting was no longer significant, suggesting that the site level variable had captured the variation that had previously been represented by term for the number of people. One possible route for exploring this mechanism would be to obtain data on the organizations and staff members conducting the surveys. It is possible if there is enough overlap among surveys between organizations and individuals that it might be possible to estimate the quality of the surveyors, and thus remove its effect. It is also possible that the model could be expanded to include a component to model the variance introduced through different samplers, and thus control for it in attempting to estimate the effect of other variables.

Table 6.4.2.c Statistical model of the density of debris recorded in NOAA Standing Stock surveys, with the number of people assisting included as a smooth term instead of a parametric term.

A. Parametric Covariates	Coefficient	Std. Error	p Value	Median	Median Effect
Intercept	1100	110	5.50E-22		1100
State: CA	0.056	0.14	0.69		0.056
State: DE	0.12	0.19	0.5		0.12
State: HI	1.8	0.19	6.10E-22		1.8
State: MD	1.1	0.17	6.00E-10		1.1
State: VA	-0.54	0.17	0.002		-0.54
Year	-0.54	0.055	3.20E-22	2012	-1086.48

B. Smooth Components	edf	Ref.df	F	p-value
s(Month)	8.4	9	32	2.40E-58
s(No. of Persons Assisting)	5.7	6	83	2.90E-97

Although there are issues that remain with aliasing in the smooth for the number of people participating in the surveys, we moved on to do some preliminary exploration of the effect of environmental variables on the density of debris. We incorporated the total length of roads within 5km of each site as a proxy for the density of urbanization in the area around the site. We also incorporated a measure of the urbanization in a larger area around the site by using the residuals of the relationship between the length of road within 5km of the survey location and 50km of the survey location. By using the residuals we are able to remove any issues with co-linearity between the 5km and 50km values, allowing us to distinguish between areas where there is urbanization both locally and regionally, locally only, regionally only, and neither locally nor regionally. We also moved to using standardized versions of the continuous variables (Year, roads within 5km, roads within 50km, and number of persons assisting) to improve the model estimation. This model captured 68% of the variation in the data, based on a deviance comparison. Based on a goodness of fit test the model also has an adequate fit to the data overall, with no major issues due to over-dispersion. However, model fit to extreme high values could be improved, if we can identify appropriate covariates.

In terms of interpreting the estimated effects, there do appear to be some differences among the states in terms of debris densities. All of the states had lower values than Alaska, with the exception of Hawaii which was not significantly different from Alaska. Debris densities at coastal sites increases with the density of roads within 5km of the site, suggesting that debris scales with both access and population density at the local scale. Interpreting the effect of residuals between 5km and 50km, it is important to understand what the residuals mean. High residuals suggest that road densities at 50km are greater than 5km, so that urbanisation increases away from the site. Low residuals suggest the density at the local scale is higher than the regional scale; for example, small coastal communities. In this model, the coefficient for the residuals from the model is negative. Therefore, at high residual values, where road densities are greater away from the site than close by, we find lower debris levels. Conversely, sites that have more urbanization at the site itself than regionally will have higher debris levels. In effect, outlying coastal communities appear to have more debris on their coastlines. Of the substrate types, only the medium density one had a significant effect. These sites had less debris than the remaining categories, which did not differ from the overall average. Overall this model provided a reasonable fit to the data, explaining 69.5% of the variation in the data, based on a deviance comparison.

Table 6.4.2.d Statistical model of the density of debris in the NOAA Standing Stock surveys, including environmental covariates.
The substrate covariates are drawn from NOAA's survey data. Median effects are only shown for significant variables. Median effects for factors are reported for the case where each factor is in effect. The reference levels, which are incorporated into the intercept term are State: AK, Substrate: NA.

A. Parametric Covariates	Coefficient	Std. Error	p Value	Median	Median Effect
(Intercept)	-3.8	0.25	8.00E-51		-3.8
State: CA	-1.5	0.14	2.80E-24		-1.5
State: DE	-1.1	0.22	2.70E-07		-1.1
State: HI	0.37	0.26	0.15		0.37
State: MD	-0.82	0.17	8.70E-07		-0.82
State: VA	-2.1	0.18	3.20E-30		-2.1
Year	-0.27	0.07	1.00E-04	-0.26	0.07
5 km roads	2	0.1	2.50E-79	-0.43	-0.86
5-50 km road residuals	-1.8	0.086	1.50E-87	-0.48	0.87
Substrate: Barren	-0.9	0.2	3.80E-06		-0.9
Substrate: Developed, High	-1.4	0	0		-1.4
Substrate: Developed, Med	-2.5	0.29	9.80E-19		-2.5
Substrate: Wetlands	0.07	0.21	0.74		0.07
Substrate: Grassland	0	0.7	1		0
Substrate: Open Water	-1.1	0.22	1.30E-06		-1.1
Substrate: Scrub	-0.64	0.22	0.0037		-0.64

B. Smooth Components	edf	Ref.df	F	p-value
s(Month)	7.5	8	30	1.20E-48
s(No. of Persons Assisting)	5.7	5.9	17	2.20E-19

6.4.3 Modelling total load of debris in the ICC Data

The ICC data records the number of people participating, along with the length of the site. There was no area calculation possible, as the site shape or dimensions beyond length were not available. We found some other minor issues with the data. For instance, for sites with less than 10 total items the counts for each category were listed as NA. Given this, sites with low debris or litter levels are likely underrepresented in the data we were able to analyse.

Table 6.4.3.a ICC survey total item count by year and state

State	2010	2011	2012	2013	2014	2015	2016
Alabama	68585	85613	80176	134523	204885	NA	NA
Alaska	17388	13234	7123	13508	11386	NA	NA
Arizona	5153	162	2254	2905	2948	NA	NA
Arkansas	5410	4149	2697	4229	2971	NA	NA
California	886147	854496	616425	757482	1261970	2123	NA
Colorado	5703	61	5880	2214	NA	NA	NA
Connecticut	93432	68516	61838	58617	83963	NA	NA
Delaware	28451	64596	41022	65904	69872	NA	NA
District of Columbia	11023	11183	9578	NA	10471	NA	NA
Florida	606766	555859	583130	809352	715107	0	0
Georgia	0	97952	65087	135260	110010	6	NA
Hawaii	165254	179263	210536	189090	231764	NA	NA
Idaho	573	323	1984	2744	134	NA	NA
Illinois	99800	75679	61814	79854	94984	NA	NA
Indiana	28885	13381	18808	27327	34067	NA	NA
Iowa	757	NA	NA	NA	NA	NA	NA
Kansas	3977	2506	4483	2974	2292	NA	NA
Kentucky	NA	NA	910	1368	NA	NA	NA
Louisiana	21751	36791	24954	10169	51006	NA	NA
Maine	49800	49700	33347	47986	33144	NA	NA
Maryland	55532	60030	65817	103881	132785	NA	NA
Massachusetts	103358	88607	107418	150176	128284	NA	NA
Michigan	73403	80945	72131	86756	100249	NA	NA
Minnesota	12077	4396	12349	11930	4387	NA	NA
Mississippi	47746	60338	63762	58041	NA	NA	NA
Missouri	484	349	1740	1323	2648	NA	NA
Nebraska	9613	6744	4705	20849	14569	NA	NA
Nevada	858	2335	NA	4347	3472	NA	NA
New Hampshire	46726	44768	41401	43539	45503	NA	NA
New Jersey	101583	130378	18085	195947	103367	NA	NA
New Mexico	557	NA	NA	234	NA	NA	NA
New York	347654	280997	206408	277830	317720	162	NA
North Carolina	86844	70861	85622	153310	102850	NA	NA
Ohio	53028	34256	53663	69206	75400	NA	NA
Oklahoma	1783	NA	NA	NA	NA	NA	NA
Oregon	9433	8911	5800	7545	15020	8002	NA
Pennsylvania	95921	71338	102791	48722	111304	NA	NA
Rhode Island	131598	114294	124913	87477	157579	NA	NA
South Carolina	120110	116249	48121	155027	103005	185	NA
South Dakota	565	141	476	846	240	NA	NA
Tennessee	NA	NA	6652	3945	NA	NA	NA

Texas	188364	197953	185041	354776	313507	13409	NA
Utah	7741	7474	7318	3757	868	NA	NA
Vermont	NA	4195	2558	659	2890	NA	NA
Virginia	131871	105808	154890	152026	169559	0	NA
Washington	28173	15689	3253	39661	42371	NA	NA
West Virginia	66	94	170	NA	NA	NA	NA
Wisconsin	25138	27583	29639	24712	45737	NA	NA
Wyoming	789	1836	927	NA	NA	NA	NA

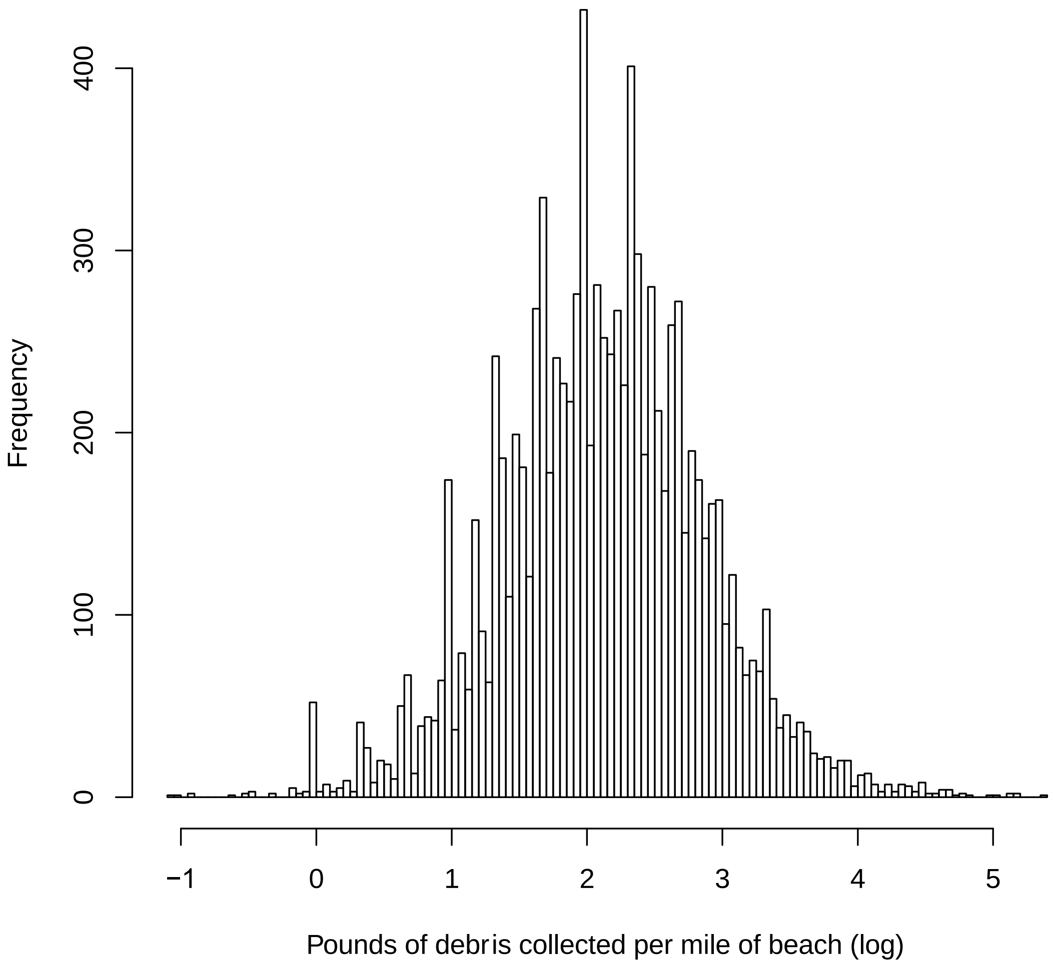


Figure 6.4.3.a Pounds of material removed by International Coastal Cleanup staff and volunteers at US sites between 2010 and 2015.

We were able to analyse 9,647 clean-up records, across 4,995 sites from the total 13,119 clean-up records. We excluded records that were missing distance records, total weight of debris, or were outside the window 2010 to 2015. The average site in the ICC data has a load of approximately 100 lb of material per mile. The values are nearly normally distributed on the log scale (Figure 6.4.3.a). In order to focus on the areas with significant sampling effort, we further focused our analysis down to sites within the continental USA, reducing the sample size to 9,207 surveys.

We modelled the ICC data based on the number of pounds of material per mile surveyed, in order to control for variation in the size of different clean-up sites. Our base model included a term for the total number of people participating in the clean-up, as a control on sampling effort. We explored terms for year and month in order to capture temporal variation in loads. We also explored a fixed effect for State, as we wanted to investigate whether different states had significantly different debris loads. We included County as a random effect, as it is possible that county level variables, such as waste collection rates, could add variation to the data. We also explored the inclusion of an underlying spatial surface for the data.

The ICC model explained 39.3% of the variation in the data, based on a deviance comparison. However, the model suffers from over-dispersion in the data, in particular there are a number of very high values that we have not been able to capture. Thus, this model should be taken as indicative, but parameter estimates are likely to change to some extent if we are able to adjust the model to better fit the data, either through incorporation of additional covariates or a change in the underlying distribution we are using to model the variation in the data. The model identified Massachusetts and Texas has having significantly high debris levels, while the remaining states with significant coefficients had lower levels than the overall mean (Table 6.4.3.b). States not listed in the table did not differ from the overall mean. There was a negative relationship between the amount of debris per mile and year, with a small notable decline over time. This equated to a 5% decline between years, based on the change between 2014 and 2015.

Table 6.4.3.b Statistical model of debris per mile surveyed from the ICC data from the United States. The full model included all of the states in the ICC data. The reference level for the State term, which is included in the intercept, is Alabama.

A. Parametric Covariates	Coefficient	Std. Error	p Value	Median	Median Effect
Intercept	110	21	3.30E-08	NA	110
State: Maine	-1.8	0.54	0.00077	NA	-1.8
State: Massachusetts	1.4	0.52	0.0056	NA	1.4
State: Michigan	-1.6	0.52	0.002	NA	-1.6
State: Minnesota	-2.5	0.78	0.0017	NA	-2.5
State: Mississippi	-2.5	0.59	1.60E-05	NA	-2.5
State: Missouri	-1.4	0.68	0.035	NA	-1.4
State: New Hampshire	-1.9	0.65	0.0039	NA	-1.9
State: Oregon	-1.4	0.6	0.017	NA	-1.4
State: Texas	1.5	0.58	0.011	NA	1.5
State: Wisconsin	-1.4	0.59	0.017	NA	-1.4
Year	-0.054	0.01	1.70E-07	2010	-108.54

B. Smooth Components	edf	Ref.df	F	p-value
s(Month)	7.4	8	13	7.1E-20
s(No. of Persons Assisting)	5.7	6	17	3.2E-19

Note that for brevity, we only include states that have significant coefficients The remaining states are: Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Nebraska, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia, Wyoming. For full dataset, please see Figure 6.1.2.a.

We extended the model to include a smooth spatial surface for location within the continental United States. We also included a term for the time since the preceding clean-up, as there is a possibility that past clean-ups affect the debris found at a site in subsequent clean-ups. This model was able to capture 46.5% of the variation in the data based on a deviance comparison (Table 6.4.3.c). It improved significantly on the simpler model (Table 6.4.3.b), with an AIC of 125305 versus 132291 for the simpler model. However, despite its improvement over the simpler model in terms of fit, there are still issues with poor fit to the overall data. These may be addressed as before with extensions of the model to allow for more complex effects from covariates or shifts in the distribution used to model the error. Thus, as before the covariates need to be interpreted with some caution, as there is a possibility some will shift as the model is improved.

The states with significant differences from the overall mean in terms of debris loads shifted slightly with the inclusion of additional terms in the model. The time interval between clean-ups (in days) had a significant and negative effect on the amount of debris sampled in a survey. This could be driven by sites with larger amounts of debris being cleaned more often, although other explanations are also possible. The negative trend with year remained at nearly the same value as in the simpler model.

Examining the spatial component from the expanded model, we can see that the model finds a residual pattern of relatively higher debris at clean-up sites in the central and south-eastern portions of the continental United States (Figure 6.4.3.b).

Table 6.4.3.c Statistical model of debris per mile surveyed from the ICC data from the United States. The full model included all of the states in the ICC data. The reference level for the State term, which is included in the intercept, is Alabama. See Table 6.1.2a for list of all states included in the model.

A. Parametric Covariates[1]	Coefficient	Std. Error	p Value	Median	Median Effect
Intercept	76	23	0.0011	NA	76
State: Connecticut	3	1.3	0.017	NA	3
State: District of Columbia	3.3	1.6	0.032	NA	3.3
State: Maryland	2.2	1.1	0.043	NA	2.2
State: Massachusetts	3.9	1.3	0.0027	NA	3.9
State: Pennsylvania	2.6	1.2	0.024	NA	2.6
StateRhode Island	3.2	1.3	0.012	NA	3.2
StateVirginia	2.8	1.1	0.0072	NA	2.8
TimeInterval	-0.00025	5.7E-05	9.1E-06	364	-0.091
Year	-0.054	0.01	1.7E-07	2010	-108.54

B. Smooth Components	edf	Ref.df	F	p-value[1]
s(County)	320	540	1600	0
te(Longitude,Latitude)	200	240	920	0
s(Month)	3.2	8	9.2	0.059

[1] *probabilities far below 0.05, e.g. 1E-20 were rounded to 0 for brevity.*

We also explored the inclusion of an underlying spatial surface for the data.

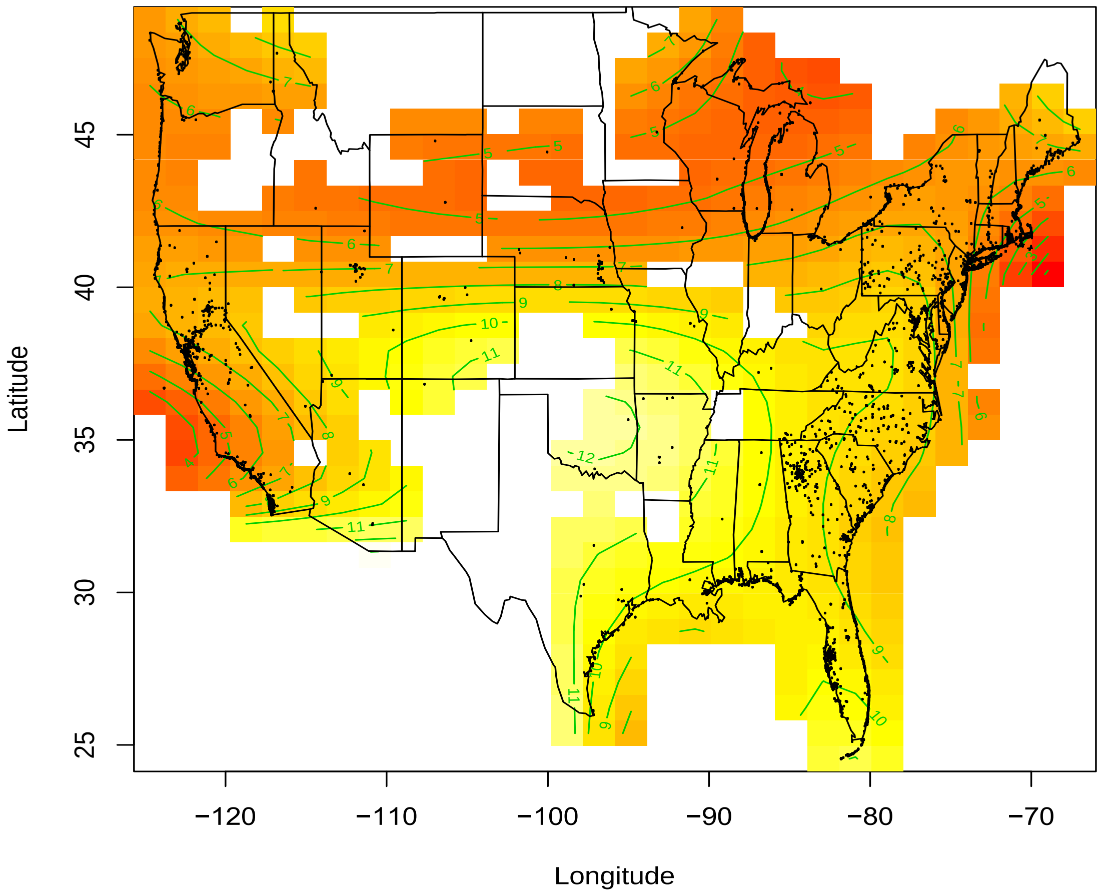


Figure 6.4.3.b Coefficient values for the spatial surface included in the model of ICC data. The response variable is debris collected per mile of survey.
The full model is given in Table 6.4.3.c. The plot shows the values of the coefficient from the spatial surface. These are constrained to have a mean contribution to the model of 0, thus they can be interpreted as spatial deviations from the expected values given the other model components. They can be compared directly with the effect sizes in Table 6.4.3.c. Locations with higher coefficient values (i.e. more debris) are yellow, those with lower values (i.e. less debris) are red. Green lines are isotherms of the coefficient value, with the actual value embedded in the isotherm. The plot is constrained to display the surface only in areas within a certain proximity of a data point.

After incorporating all of the available covariates, more of the states are significantly different in the model. Year still has a negative coefficient, and is one of the strongest terms in the model. Both the fraction of poverty and the population residuals also explain a significant amount of the variability in the model (Table 6.4.3.d and Figure 6.4.3.c),

Table 6.4.3.d Statistical model of debris per mile surveyed from the ICC data from the United States, incorporating all possible covariates.

Included in the intercept are State: AL, SiteType: Fresh Water, and LandUse: Barren. For brevity we have only included states with a significant effect, and have removed the following states: AZ, AK, CO, ID, IN, IA, KS, KY, LA, ME, MN, MS, MO, MT, NV, NM, NY, OK, OR, TN, TX, UT, WA, WV.

A. Parametric Covariates	Coefficient	Std.Error	p Value		Median	Median Effect
(Intercept)	110	21	7.70E-08			110
State						
CA	2.9	1.2	0.014	*		2.9
CT	1.8	0.92	0.057	+		1.8
DE	2.5	0.85	0.0038	*		2.5
DC	4	0.91	1.20E-05	*		4
FL	2.5	0.47	7.70E-08	*		2.5
GA	3.1	0.66	2.40E-06	*		3.1
IL	2.4	1.2	0.048	*		2.4
MD	3.9	0.81	1.10E-06	*		3.9
MA	3.4	0.98	0.00051	*		3.4
NE	6.7	3.2	0.035	*		6.7
NH	3	1	0.0027	*		3
NJ	1.9	0.88	0.028	*		1.9
NC	2.7	0.78	0.00062	*		2.7
OH	2.9	1.1	0.0092	*		2.9
PA	3.3	0.89	0.00018	*		3.3
RI	2.9	0.95	0.0027	*		2.9
SC	3.8	0.77	8.80E-07	*		3.8
VT	3.1	1.4	0.022	*		3.1
VA	4	0.79	3.70E-07	*		4
WI	3	1.2	0.011	*		3
WY	4.2	7.8	0.59			4.2
Total People	0.0014	5.80E-05	5.9E-122	*	25	0.034
Pop 50 km	1.20E-07	4.30E-08	0.0044	*	950000	0.12
Poverty fraction 50km	-5.8	2.1	0.0054	*	0.98	-5.6
Unemployed fraction 50km	-5.1	1.9	0.0068	*	0.28	-1.4
Working fraction 50km	-8.8	2.3	0.00014	*	0.52	-4.6
Site Type						
Inland	-0.41	0.12	0.0011	*		-0.41
Saltwater	-0.13	0.049	0.0073	*		-0.13
Year	-0.049	0.01	3.20E-06	*	2000	-98
Rail Distance	-0.0051	0.0013	9.50E-05	*	19	-0.099
Length Roads in 50km	-4.30E-05	3.70E-05	0.24		3400	-0.14
Land Use						
Cultivated Crops	0.47	0.18	0.0081	*		0.47
Deciduous Forest	0.2	0.095	0.031	*		0.2
Dev., High Intensity	0.13	0.064	0.05	*		0.13
Dev., Low Intensity	0.18	0.055	0.0012	*		0.18
Dev., Med. Intensity	0.38	0.052	4.90E-13	*		0.38

Dev., Open Space	0.59	0.062	1.40E-21	*	0.59
Emer. Herb. Wetlands	0.68	0.075	2.00E-19	*	0.68
Evergreen Forest	0.4	0.11	0.00039	*	0.4
Grass/Herbaceous	0.46	0.11	2.50E-05	*	0.46
Mixed Forest	-0.23	0.2	0.25		-0.23
Open Water	0.26	0.063	3.50E-05	*	0.26
Pasture/Hay	0.52	0.14	0.00024	*	0.52
Shrub/Scrub	0.48	0.12	0.00011	*	0.48
Woody Wetlands	0.66	0.082	6.30E-16	*	0.66
Population Residuals 5 to 50km	-7.30E-07	2.70E-07	0.0074	*	-7.30E-07
Road Residuals 5 to 50km	0.0039	0.00031	1.40E-35	*	0.0039

B. Smooth Covariates	edf	Ref.df	F	p-value	
te(Longitude,Latitude)	254	302	7.9	0	*
s(Month)	1.4	8	0.48	0.05	*

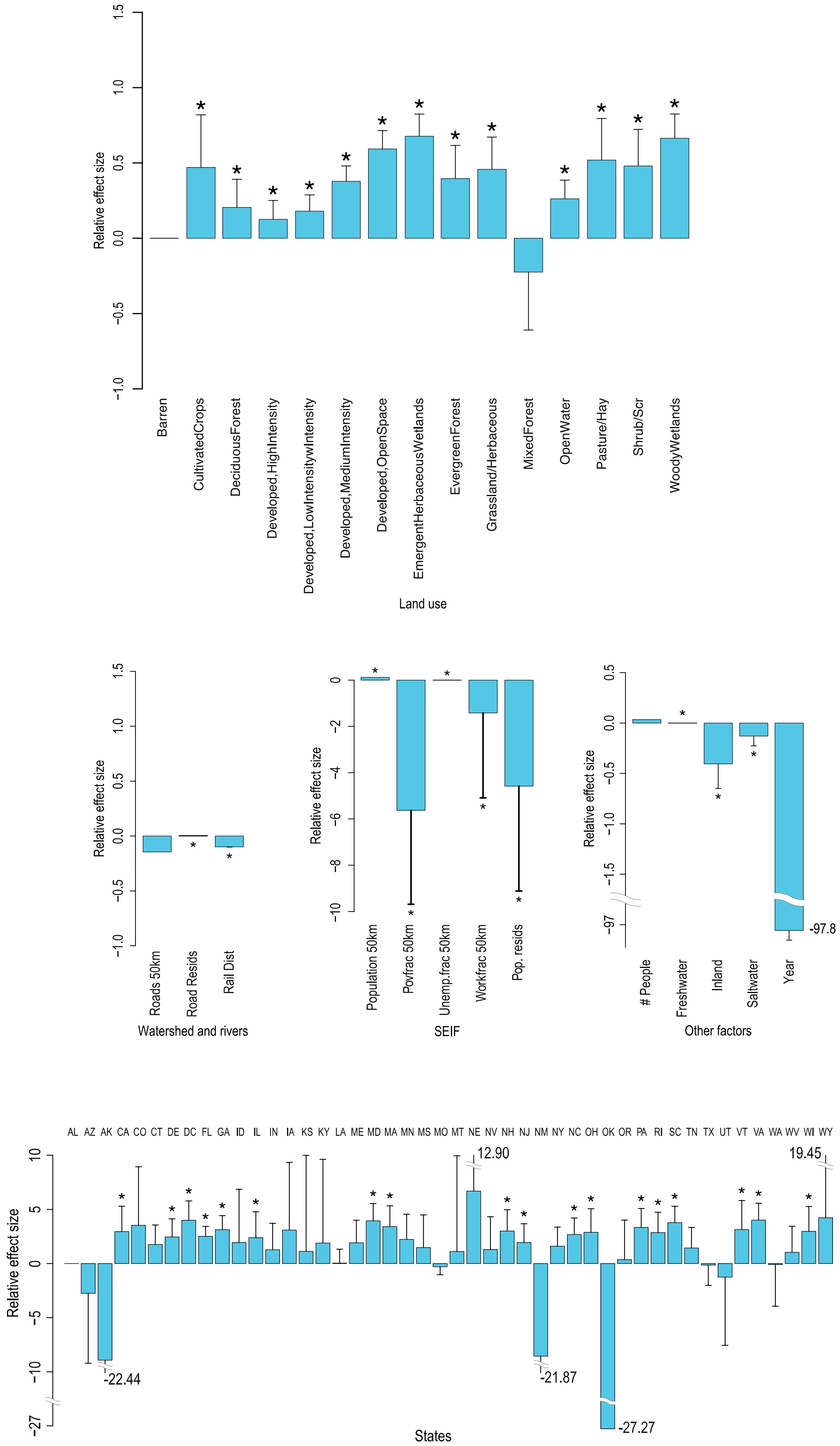


Figure 6.4.3.c Effect sizes for GAM modelling ICC data, with full covariate sets included

Effect sizes give an indication of the relative importance of each model coefficient to the overall results of the model.

6.4.4 Modelling total load in the CSIRO data (2016)

For the second half of this report, we fit a full GAM model to the CSIRO data (as described in section 6.4). Various site level factors, including aspect, backshore, and substrate type were significant explanatory variables. Notably, the urban backshore, found only at sites 26 and 27 (Figure 2.2.b) had a very strong effect. Once again, poverty fraction was significant, as well as work fraction (Table 6.4.4.a and Figure 6.4.4.a).

Table 6.4.4.a Statistical model of debris per mile surveyed from the CSIRO data from the west coast of the United States
This incorporates all possible covariates. Included in the intercept term are Substrate: Clay, State: California, Backshore:Cliff, Aspect: N.

A. Parametric Covariates	Coefficient	Std.Error	p Value		Median	Median Effect
Intercept	-12	13	0.36			-12
Substrate						
Gravel/Pebble	1.4	1.3	0.31			1.4
Rock Slab	-0.65	1.8	0.72			-0.65
Sand	2.8	1.2	0.029	*		2.8
State						
Oregon	-23	6	0.00035	*		-23
Washington	-16	5.9	0.011	*		-16
Distance to nearest road km	-1.3	0.2	2.1E-08	*	0.47	-0.62
Backshore						
Dune	-0.085	0.86	0.92			-0.085
Forest/Tree>3m	-0.58	2.5	0.81			-0.58
Grass/pasture	1.5	0.5	0.005	*		1.5
Grass tussock	-1.6	1	0.12			-1.6
Seawall	4.7	1.1	7.50E-05	*		4.7
Shrub	0.26	0.61	0.67			0.26
Urban	-47	10	4.1E-05	*		-47
Aspect						
NW	10	4.3	0.019	*		10
S	15	5	0.0047	*		15
SE	15	4.9	0.0031	*		15
SW	12	4.6	0.01	*		12
W	11	4.2	0.0089	*		11
Distance to River km	0.048	0.015	0.0025	*	7.4	0.36
Length Roads in 50km	-0.0037	0.0016	0.027	*	930	-3.5
Watershed Population (10)	1.30E-05	4.30E-06	0.0041	*	8200	0.1
Watershed Area (10)	-0.0014	0.00061	0.027	*	920	-1.3
Housing density 50km	0.00041	0.00013	0.0025	*	2900	1.2
Working fraction 50km	89	21	9.9E-05	*	0.47	42
Poverty fraction 50km	-55	8	7.7E-09	*	0.99	-54
Unemployment Fraction 50km	41	16	0.014	*	0.36	15
Median Age 50km	0.077	0.09	0.4		36	2.8

B. Smooth Components	edf	Ref.df	F	p-value	
s(Location)	8.42	8.86	7.3	1.54E-07	*

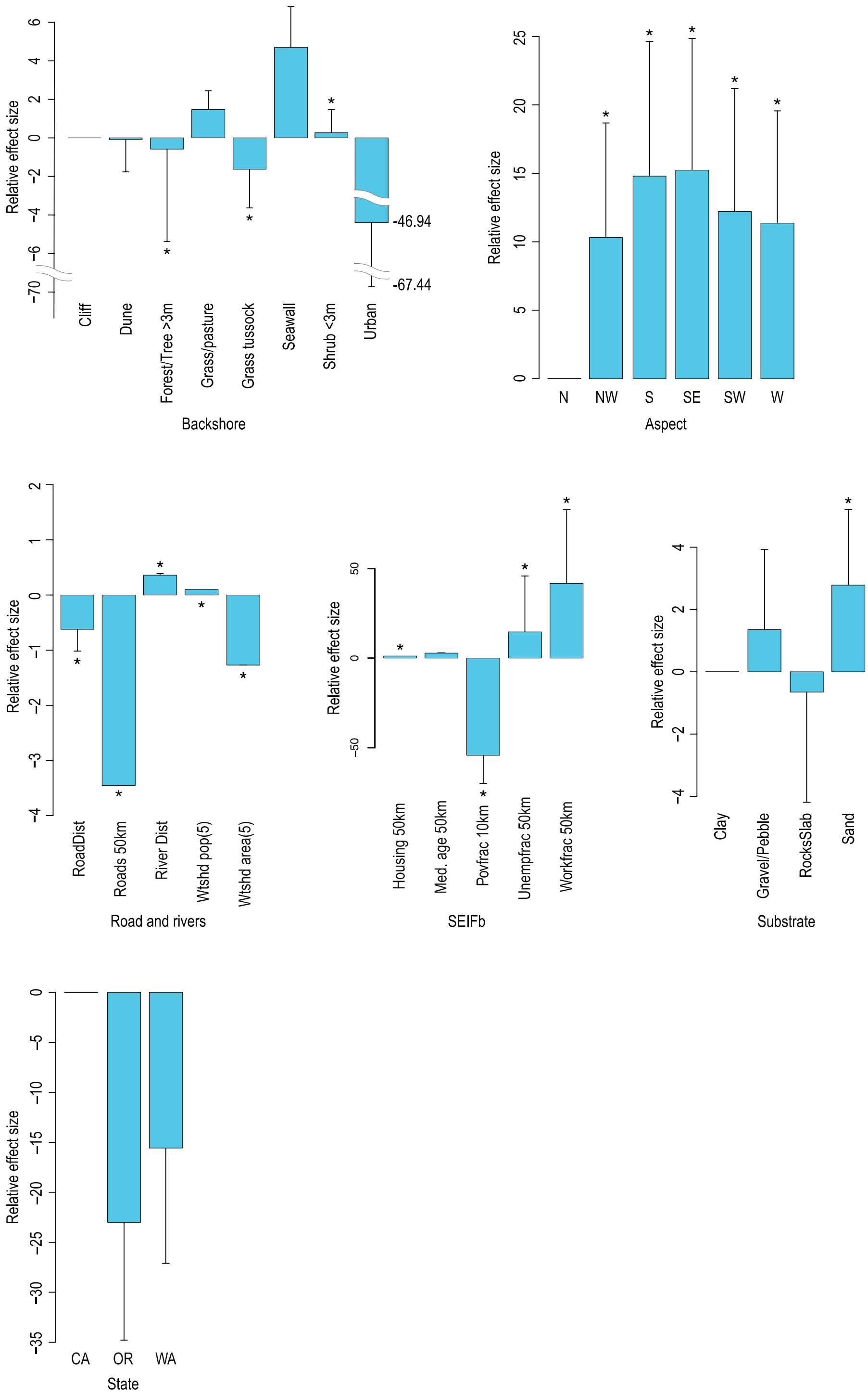


Figure 6.4.4.a. Effect sizes for GAM modelling CSIRO data, with full covariate sets included

Effect sizes give an indication of the relative importance of each model coefficient to the overall results of the model.

6.4.5 Interpretation of analyses

We were able to fit informative models to the three data sets, NOAA's accumulation data, NOAA's standing stock data, and the ICC clean-up data. Each of these datasets presents some challenges in model fitting and interpretation. All three data sets have strongly skewed data that requires non-standard methods for estimation. Moreover, caution is required with model fits, and checking goodness of fit is essential. While we have found models that provide adequate fits to the two NOAA data sets, the model of the ICC data needs improvement as it in its current form it fails a goodness of fit test.

We were able to establish models that provide baseline estimates and change detection for each of the datasets. In all three cases time appears to be a significant component of the models, indicating that debris identified or collected is changing over the analysis period. However, all three data sets appear to be complicated by sampling issues having to do with observer effort or quality. There appears to be a consistent pattern that more debris is identified at a site as more observers are added. In some cases there is a suggestion that observer quality varies among sites in a consistent way, potentially introducing a bias in the data. Ultimately, given that we can estimate the effects of these variables we can standardize the data for them, removing the bias from the estimates. However, this effort will be improved by the inclusion of some additional information, such as the identity of the organization and individuals collecting the data at each site.

We extended the models to incorporate environmental effects, both from metadata collected as part of the surveys and using additional information we obtained from outside sources. Some of these variables appear to be important, in particular we found that road density, which is a proxy for population density and access, has a significant effect on the estimates.

6.5 Categorical Debris Analysis

There is a trade-off among the analytical approaches that can be taken to marine debris data arising from the NOAA surveys and ICC clean-ups. At one extreme, the data can be aggregated up into a total count of items across all categories or a total weight. The advantage of this approach is that by using a single category of data modelling efforts can focus on the full complexity of space and time patterns, incorporating both driving variables, such as local population size, and nuisance variables, such as sampling effort. This is the approach we have taken in Section 6.4 of this report.

At the other extreme one can model the abundance of items in each of the categories, across the tens of categories that are recorded in the NOAA or ICC data. The challenge in this approach is that models describing the abundances in each category may differ, leading to a very complex interpretation of the data. Furthermore, categories may be positively or negatively correlated. For instance, the count of bottles and caps could plausibly be strongly correlated. Thus, while one might initially attempt to extend the approach used for modelling debris loads in the previous section to numerous categories, it is likely prudent to attempt to either simplify the analysis approach or aggregate the categorical data to some extent. After all, if a model is as complex as the data it is trying to describe, it is difficult to extract any additional understanding from the modelling effort.

A range of levels of abstraction are possible in considering models for the categorical data in the NOAA and ICC datasets. Most simply, categories could be aggregated, for instance by material type, likely source, function, or other criteria. This data could then be modelled using equations for each category. A further reduction would be to use the category frequencies to describe the sites, then identify types of sites where the frequencies across the categories differ. One could then attempt to model these site types. This could be done using either an equation per type, or potentially one equation that describes the relative likelihood of the different types across the sites. Finally, the simplest approach would be to consider the frequency distribution across categories, but ignore the identities of the categories themselves. An example of this approach would be to rank order the categories based on their counts in a sample, and then model the rate of decline in frequency from the most to the least common item. This approach could describe changes in sites, such as new sources of debris, or differentiate sites with single versus multiple sources. Although it ignores the identity of the category for the counts, it does have the benefit that it can be described by a single equation, thus greatly reducing the complexity of interpretation.

Here we have implemented analyses for the two of the three approaches. We have used cluster analysis based on polya distributions to examine whether there are identifiable site types, which sites belong to which types, and how those types differ. We extended this analysis using regression trees to investigate explanatory models for the site types. In this case we did not use the site types inferred from the polya distribution approach as our response variable, instead allowing the regression tree to distinguish among sites based on the counts in the categories. Regression trees are useful for complex data where there are no a priori hypotheses about the variables driving the response, as they are data mining tools designed to find (potentially complex) relationships among large numbers of explanatory variables and responses. Finally, we aggregated the categories into a subset of material types and fitted models for the count in each category as a joint random variable using a multivariate normal distribution. This approach allows us to fit a model for the mean value of each category simultaneously, but incorporate covariance among the categories.

Due to the complexity of the analytical task, we focused these analyses on the NOAA accumulation data. This data has the best spatial and temporal coverage of the NOAA data sets, without the sampling uncertainty occurring in the ICC dataset. We present the cluster analysis, followed by the regression tree analysis, and finally the multivariate generalized additive model analysis. We also present an analysis of variance within and among sites, which provides some insight into the magnitude of noise in the accumulation data and its capacity to distinguish differences among sites.

6.5.1 Cluster Analysis

Cluster analysis creates groupings of data points based on raw data counts. It does not incorporate covariates to define the clustering, but is an exploratory technique which identifies structures within the data by combining data points into the most homogeneous groups possible.

We analysed the raw NOAA Accumulation data using the MixPolya package in R. MixPolya fits a multinomial Dirichlet distribution with an expectation-maximisation (EM) algorithm. We removed material category subtotals, and used the raw Accumulation data counts for each category for each transect.

First we determined the appropriate number of clusters using an AIC approach. In this approach the AIC is determined for models fitting an increasing number of clusters to the data. The ideal number of clusters corresponds to the first local minimum AIC value.

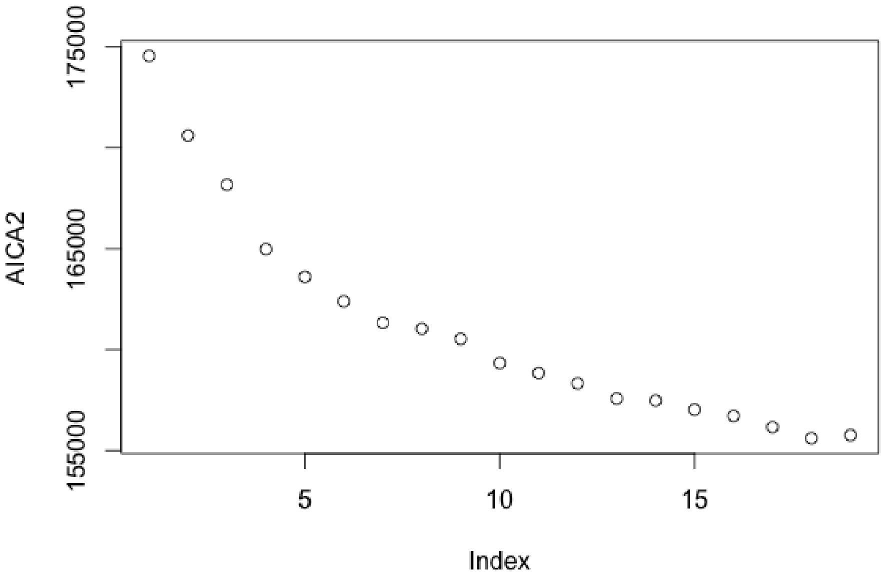


Figure 6.5.1.a AIC values for NOAA Accumulation data fit with MixPolya models.
Index values are equivalent to (number of clusters in model – 1). The minimum index value is 18, which indicates that 19 clusters best describe the data.

The above graph (Figure 6.5.1.a) demonstrates that the ideal number of clusters for the Accumulation data is 19. We can use the cluster analysis to display the membership of each of the groupings (Figure 6.5.1.b, Appendix 1). Accumulation surveys are reasonably well distributed among each of the groups (Table 6.5.1.a) with a few clusterings of rare or uncommon data types.

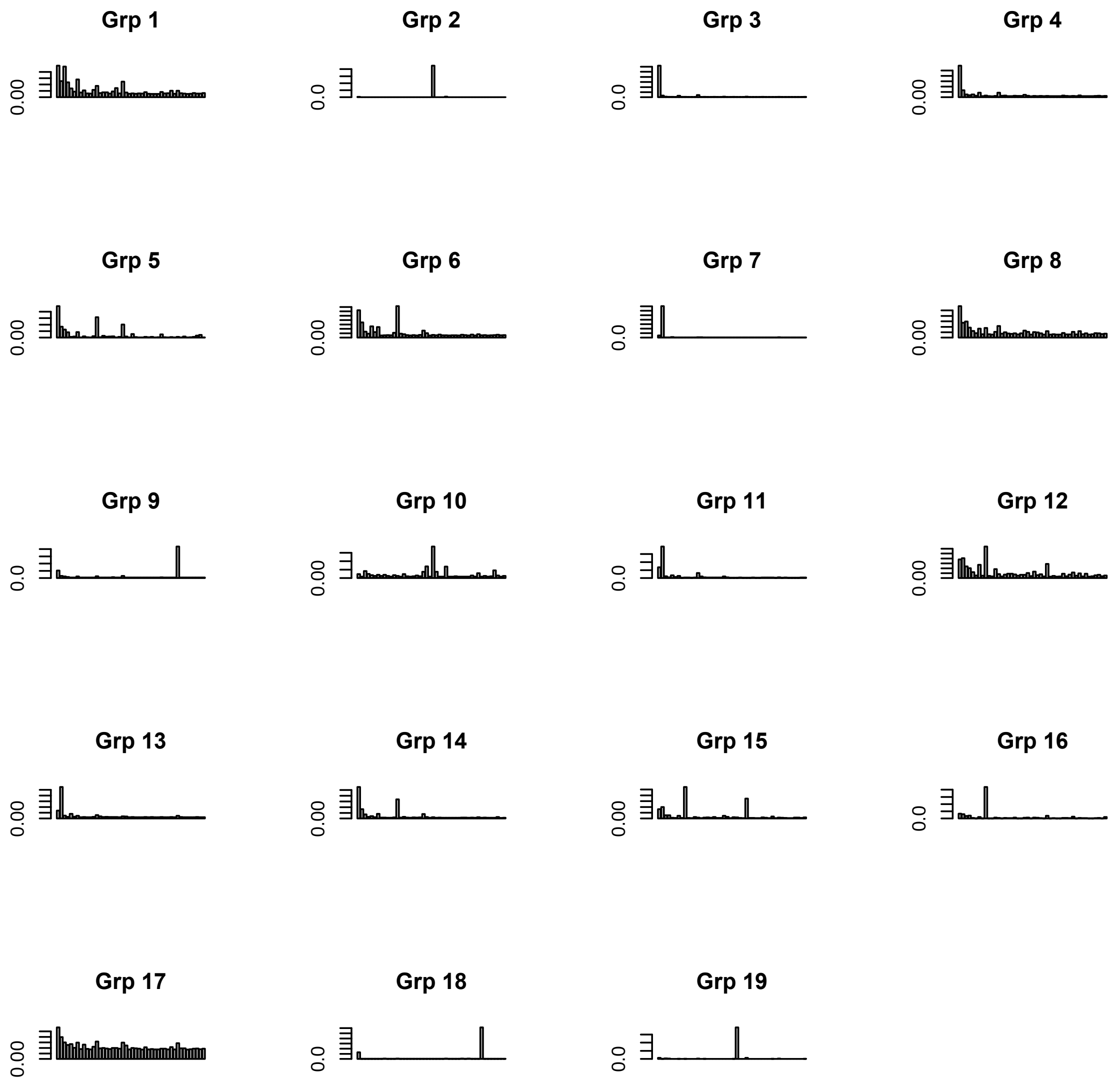


Figure 6.5.1.b Frequency distribution of debris categories across clusters for NOAA Accumulation data.
Category types are across the x axis, and on the y axes are the relative frequency of each one within the cluster. See Supplemental Information (Chapter 7) for expanded figures for each of the 19 groupings presented above.

Table 6.5.1.a Number of surveys in each of the clusters predicted for NOAA Accumulation data

CLUSTER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
# TRANSECTS	57	3	23	106	18	67	25	55	10	4	40	66	96	30	21	16	472	4	3

6.5.2 Tree based analysis

The ctree analysis in the R package partykit is used to implement conditional inference trees. Essentially the analysis uses covariates to determine the best binary split for the data to increase the homogeneity of the response variable(s) within the two subsets of data, then continues to split the data further down the "branches" until it cannot find a significant difference between the data grouping, at which point it stops until subsequent splits along a branch do not produce a significant decrease in heterogeneity among the response variables among the groupings.

Tree based analysis is similar to cluster analysis, in that it partitions the data into groupings, but here the covariates are incorporated into the splitting process, whereas in cluster analysis the groupings are done on the response variables (i.e. category frequencies) alone.

Ctree is a useful tool for data exploration, and can produce a graphical representation of the factors relevant to explaining the variability in the data. However, while trees can identify complex predictive relationships, they can be difficult to interpret due to this complexity. Here we present 3 different tree models (Figure 6.5.2.a to Figure 6.5.2.c). The main point of the figures is to demonstrate the difficulty in delineating clear patterns due to the timing of data collection, the expanse or geographic extent, the method, and, importantly, the variability of the debris items observed and recorded.

For tree analyses, we removed material category subtotals, and used the raw Accumulation data counts for each category for each transect. Similarly to the total flux analysis, we wanted to evaluate the contribution of a spatial component to the model, so we limited the analysis to the west coast of the continental United States. We modelled the spatial component in two different ways; first using the distance southward along the coast from the northernmost site (Interpoint Distance), and also using an index corresponding to the sites in order (PointIndex). The two different analyses yielded substantially similar results, so here we present the PointIndex models.

Our first tree Model (Tree 1) incorporates the covariates month, state, year, pointindex, slope, number of persons assisting, total debris, and road distance. It splits primarily on county, and secondarily on total debris. The county level differences may be geographic, or may relate to policy differences among counties. In order to determine additional driving factors, we created a second tree model, where we removed County. We also left season out, as it should be accounted for within the Month factor.

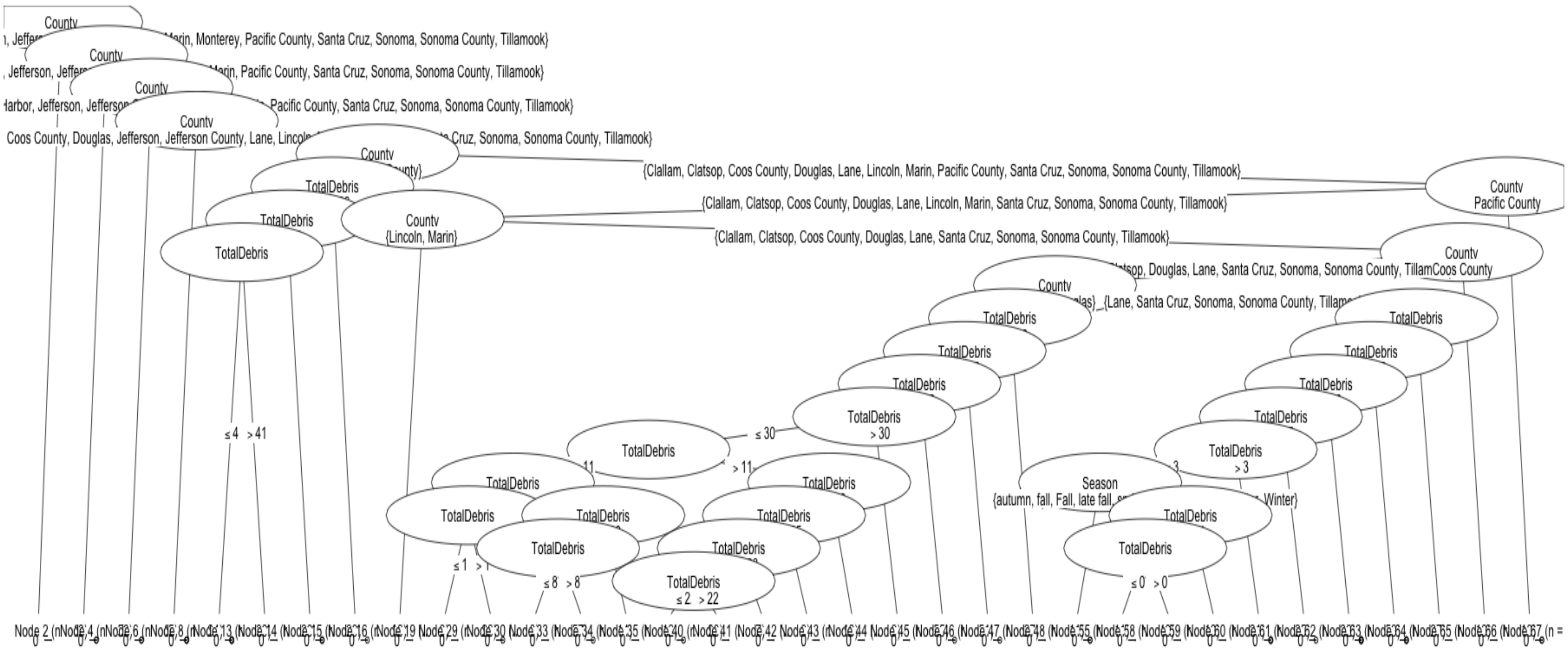


Figure 6.5.2.a Ctree analysis of NOAA Accumulation data.
The model includes all data categories as well as ~ County + Month + State + Year+ PointIndex + Slope + Season + NumberofPersonsAssisting + TotalDebris + Rdist.

Tree 2 splits almost exclusively on Total Debris load, indicating that the categorical composition of the debris will change as the total amount of debris collected changes (Figure 6.5.2.b). We additionally investigated a third model, which does not incorporate total debris load.

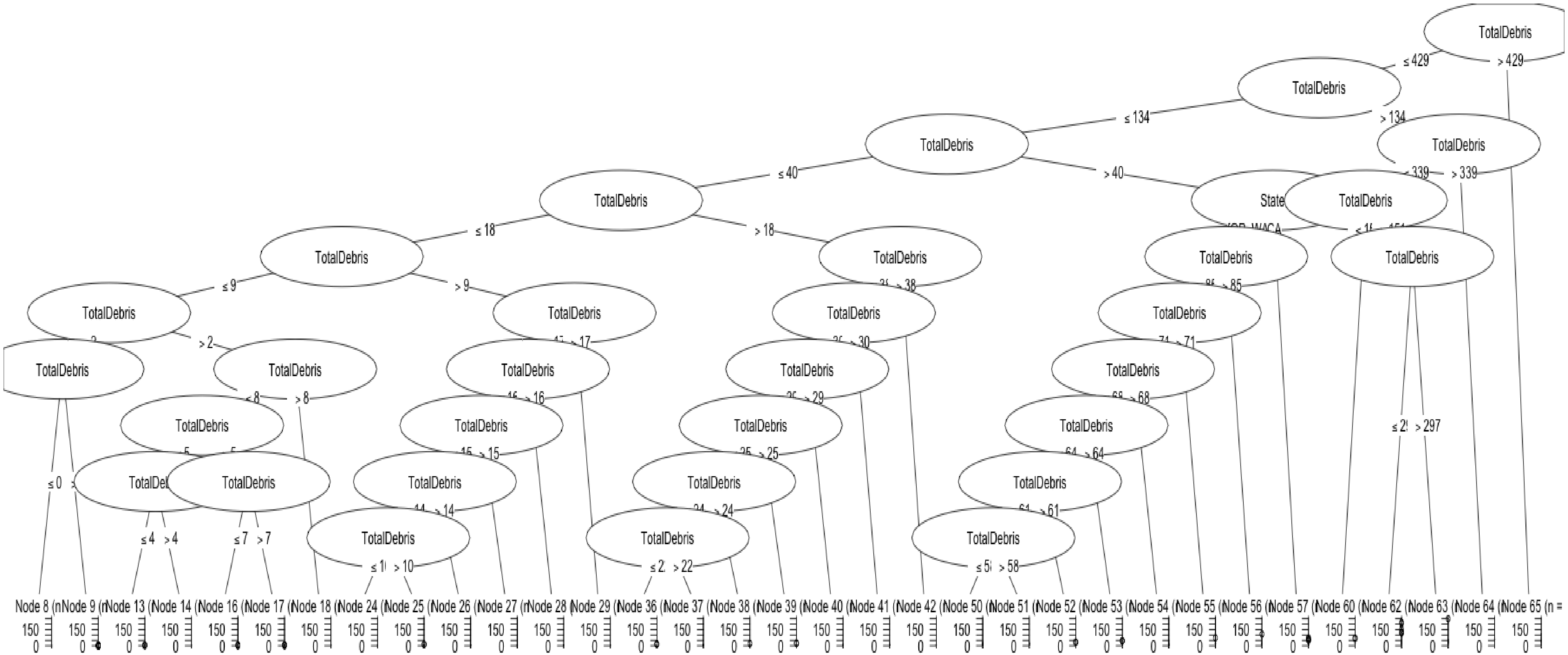


Figure 6.5.2.b Ctree analysis of NOAA Accumulation data.
The model includes all the categories of data reported as well as ~ Month + State + Year+ PointIndex + Slope + NumberofPersonsAssisting + TotalDebris + Rdist. Note that this tree splits almost exclusively on Total Debris load.

In contrast, Tree 3 (Figure 6.5.2.c) splits along a combination of geographic parameters and the distance to the nearest road.

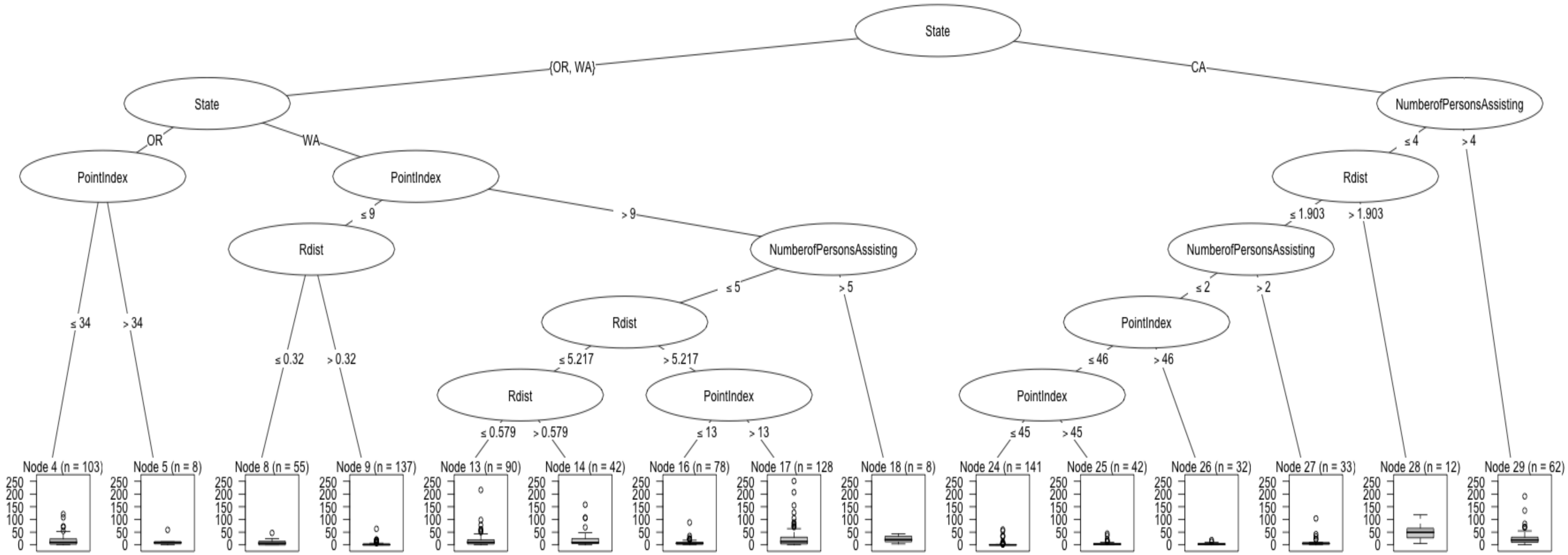


Figure 6.5.2.c Ctree analysis of NOAA Accumulation data.
Model: All categories ~ Month + State + Year+ PointIndex + Slope + NumberofPersonsAssisting + Rdist. Tree 3 splits on a combination of geographical parameters (State, PointIndex), and distance to nearest road.

6.5.3 Generalized Additive Models

Because the GAM analysis has a term for each covariate for each category, using the entire suite of categories was both computationally expensive as well as cumbersome and difficult to interpret. Therefore we analysed only the subtotals of the NOAA Accumulation data for each material category (Plastic, Metal, Glass, Rubber, Processed Lumber, and Cloth). We log transformed the data so that it would meet the requirement for normality. Generalised Additive Models (GAM) are an alternative to generalised linear models (GLM) that allows for non-parametric relationships, called smooths. A GAM can be formed as a combination of parametric terms and smooth terms. The parametric terms are the same as those in a linear regression, which assess whether the data can be fit to a linear function, while the smooth terms allow the model to select any non-linear function that best fits the data. It presumes that data points close to one another (e.g. in time, space, population size or other covariates) will be more similar to each other than those farther apart. Smooths are particularly useful for spatial data and for circular data (such as time of day or time of year) where the beginning value is expected to be the same as the end value (e.g. December 31 and January 1).

Because this model was not hypothesis driven, we used it as an exploratory tool to assess significant factors within the data. We proposed several covariates in our global model, including Number of persons assisting in clean-up, County, Total Debris, Month, Distance to nearest road, and Point Index.

We attempted two different methods of model selection:

First we began with a global model that included all terms as smooths. We then removed all non-significant smoothing terms from the model. Subsequently, we tested whether terms could be moved to a parametric form and still retain significance. The resulting model suggests a linear relationship between the following terms:

Plastic (PointIndex and Rdist) – there is a very small negative relationship between plastic and both point index and Rdist, indicating decreasing amounts of plastic as these values increase (i.e. slightly more plastic in northern surveys, and as you get closer to roads).

Glass (PointIndex and Rdist) – there is a very small positive relationship between glass and both values, indicating increasing amounts of glass in more southern surveys, and further from roads.

Rubber varies negatively with PointIndex and total debris (lower amounts of rubber in northern surveys, and when there is a high amount of debris), and directly with # people assisting and road distance (higher amounts of rubber when more people are conducting the beach clean-up, and higher amounts further from roads).

Processed lumber varies positively with PointIndex (higher values further north) and negatively with road distance (higher amounts closer to roads).

Cloth varies positively with Point Index (higher amounts further south) and negatively with number of people (fewer items of cloth with increasing numbers of surveyors).

The significant smooth terms include County (for all categories) and Month (for Plastic and Metal), as well as total debris and # of people for the plastic category.

Table 6.5.3.a Parameter estimates for a statistical model of debris material types in samples from the west coast of the continental US.
Count values are log(x+1) transformed. The median is the median value of the relevant covariate, multiplying it times the coefficient gives a measure of the effect size of each term. Factors can be taken to have a value of 1 using treatment contrasts, as in this case. Smooth terms in the model are constrained to have mean values of zero, and thus are best interpreted as deviations around the parametric components.

PARAMETRIC TERMS

	Estimate	Std. Error	z value	Pr(>|z|)
PLASTIC				
(Intercept)	7.61E+00	1.51E-04	5.04E+04	0.00E+00
PointIndex	-9.37E-06	3.74E-06	-2.51E+00	1.22E-02
Rdist	-1.37E-05	5.15E-06	-2.66E+00	7.86E-03
METAL				
(Intercept).1	1.86E+00	1.19E-02	1.56E+02	0.00E+00
GLASS				
(Intercept).2	2.17E+00	2.11E-01	1.03E+01	6.85E-25
PointIndex.2	1.19E-02	5.51E-03	2.16E+00	3.09E-02
Rdist.2	1.61E-02	8.60E-03	1.87E+00	6.11E-02
RUBBER				
(Intercept).3	7.49E-01	8.98E-02	8.34E+00	7.20E-17
PointIndex.3	-1.22E-02	1.94E-03	-6.26E+00	3.78E-10
# People	7.68E-03	2.16E-03	3.55E+00	3.87E-04
TotalDebris.3	-7.15E-05	3.09E-05	-2.31E+00	2.07E-02
Rdist.3	2.04E-01	2.49E-03	8.20E+01	0.00E+00
PROCESSED LUMBER				
(Intercept).4	-4.16E-01	9.67E-02	-4.30E+00	1.68E-05
PointIndex.4	1.58E-02	2.10E-03	7.51E+00	5.98E-14
Rdist.4	-1.07E-02	2.62E-03	-4.11E+00	4.02E-05
CLOTH				
(Intercept).5	2.37E-01	1.86E-01	1.28E+00	2.02E-01
PointIndex.5	9.40E-02	1.72E-03	5.47E+01	0.00E+00
# People	-7.15E-03	1.99E-03	-3.60E+00	3.20E-04
Smooth terms	edf	Ref.df	Chi.sq	p-value
PLASTIC				
s(# People)	1.005	1.010	2.598	0.108
s(County)	16.785	19.000	209.950	0.000
s(TotalDebris)	2.538	3.136	7.686	0.061
s(Month)	4.034	8.000	61.950	0.000
METAL				
s.1(Month)	7.610	8.000	1187.945	0.000
GLASS				
s.2(County)	15.223	19.000	557.250	0.000
RUBBER				

s.3(County)	17.999	19.000	13648.506	0.000
PROCESSED LUMBER				
s.4(County)	18.031	19.000	7857.030	0.000
CLOTH				
s.5(County)	18.908	19.000	13337.279	0.000

6.5.4 Variability within and between groups

To provide some insight into the variability within and between sites, we used NOAA Standing Stock data, because multiple transects were collected at each site at each point in time, allowing us to calculate the variability at each individual site and compare that to the variability between all of the sites. We looked at two ways of partitioning the data. First by letting cluster analysis partition the data into similar clusters and second by calculating the within groups variability and the between groups variability.

Cluster analysis came up with an optimal 7 clusters:

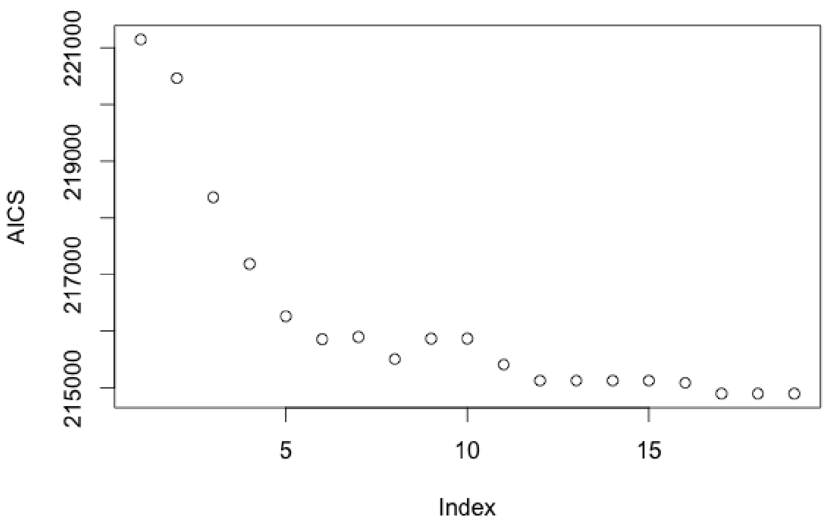


Figure 6.5.4.a AIC values for NOAA Standing Stock data fit with MixPolya models.
Index values are equivalent to the number of clusters in model – 1. The first minimum index value is 6, which indicates that 7 clusters best describe the data.

Table 6.5.4.a Number of individual transects in each of the clusters predicted for NOAA Standing Stock data

CLUSTER	1	2	3	4	5	6	7
# TRANSECTS	1596	158	27	90	39	16	32

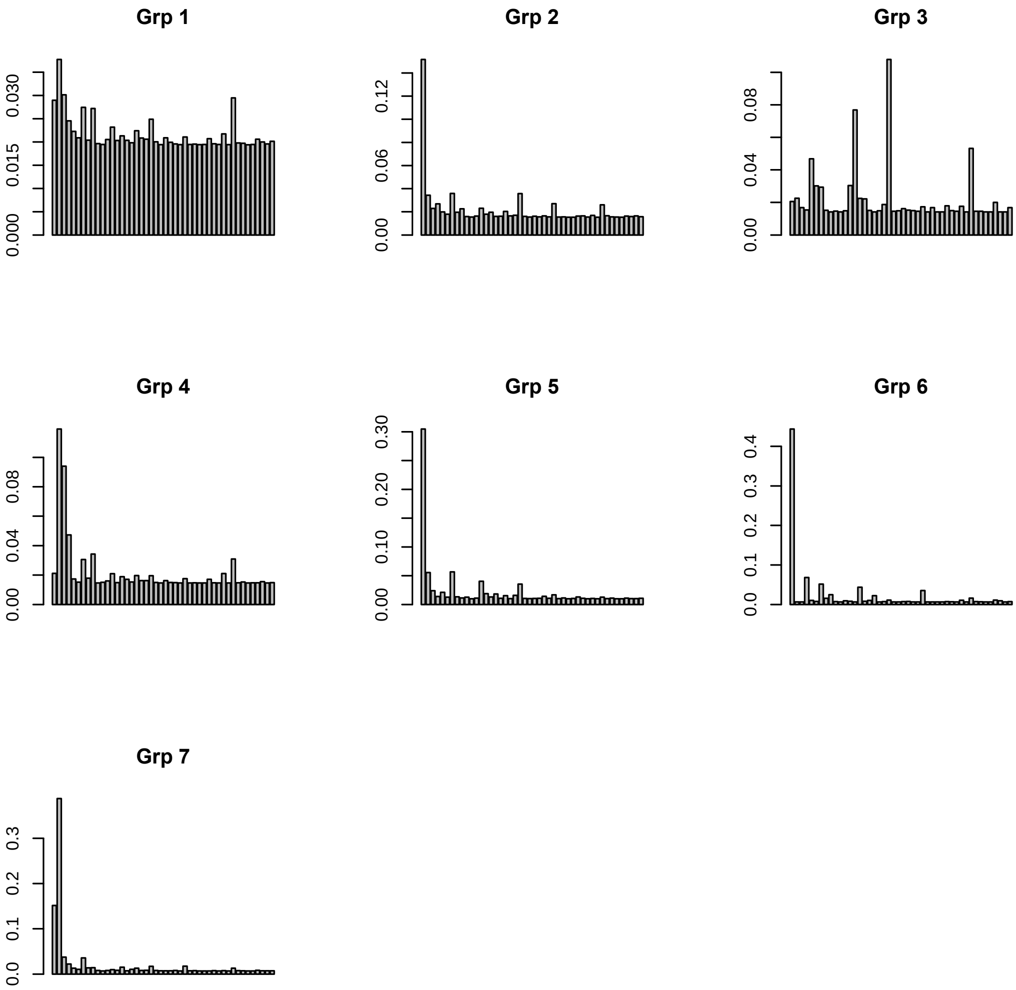


Figure 6.5.4.b Frequency distribution for cluster analysis for NOAA Standing Stock data. Category types are across the x axis, and on the y axes are the relative frequency of each one within the cluster.

To compare the variability within and between sites, we investigated whether different transects from the same survey site/date primarily clustered with each other or were split between clusters. Out of the 550 unique site/date combinations (for the survey sites with at least 3 transects in each), 409 were assigned to a single cluster, 126 to two different clusters, and 15 to three different clusters. This means the majority of site/date combinations are similar enough that the transects for each are all lumped into a single category, whereas a much smaller subset of site/date combinations divide into two or three clusters. Essentially, the results from this analysis finds that for most surveys (unique site/date combinations), there is not a large degree of difference between the individual transects at that site. This analysis gives an indication of whether transects are more similar to one another within a site than between sites. However, because the distribution of the clusters is very uneven, with about 80% of transects falling within a single cluster, we also calculated a mathematical measurement of variability.

For each category of debris we calculated the variability within a site location, and compared it to the variability between sites. We also calculated a measurement of separation (Between group variability divided by within group variability), which indicates how distinct the individual sites are from one another, for each different category. This can be useful in thinking about survey design, in that if sites are strongly separated then their patterns are likely to be informative about large scale differences in sites, for instance urban/rural or near a river outflow versus far. However, if variation at the within site level is as large as between site variation, it suggests that local factors, within the survey site might dominate patterns. For instance, locations nearer a public access point might have more litter items than those further.

The four most highly separated items are glass, glass fragments, buoys, and plastic and hard plastic fragments. There is no clear way to identify factors driving these high separations, but some speculation does seem possible. Glass is typically in high abundance on hard substrates, such as rock outcrops or boulder sites, as breakage appears to be common and people may be less likely to remove fragments. Similarly, buoys and related materials are typically concentrated on coastal sites near fishing or aquaculture industries. Finally, spatial variation in hard plastic may be a result of differential levels of ocean transport in different locations. We previously found that the glass to plastic ratio was a good proxy for land-based versus ocean-based sources of debris at coastal survey sites.

Table 6.5.4.b Measurements of within group variability (VarWithin), between group variability (VarBtwn) and separation (Separation).
Separation is calculated as VarBtwn/VarWithin, and provides a measure of how distinct individual sites are from one another.

Category	Var Within	Var Btwn	Separation
Plastic	249.35	10967.56	43.99
Hard.Plastic	68.25	2613.35	38.29
Foamed.Plastic	61.71	503.92	8.17
Filmed.Plastic	4.48	31.48	7.02
Food.Wrappers	4.39	61.21	13.96
Beverage.Bottles	1.08	11.41	10.51
Other.Jugs.Containers	0.16	3.02	18.64
Bottle.Container.Caps	3.15	67.09	21.27
Cigar.Tips	0.17	3.49	19.97
Cigarettes	1.24	20.71	16.67
Disposable.Cigarette.Lighters	0.03	0.15	4.54
X6.Pack.Rings	0.00	0.00	0.70
Bags	0.08	0.69	8.22
Plastic.Rope.Net	1.44	13.28	9.20
Buoys...Floats	0.68	41.91	61.60
Fishing.Lures...Line	1.98	29.36	14.80
Cups	0.71	1.38	1.94
Plastic.Utensils	0.11	0.38	3.51
Straws	0.37	5.14	13.78
Balloons	0.08	0.35	4.38
Personal.Care.Products	0.12	0.63	5.33
Other	4.15	86.90	20.96

Metal	0.28	0.90	3.22
Aluminum.Tin.Cans	0.04	0.11	2.67
Aerosol.Cans	0.00	0.01	3.97
Metal.Fragments	0.19	0.38	2.02
Other.1	0.04	0.09	2.23
Glass	0.76	39.55	52.38
Beverage.Bottles.1	0.03	0.73	22.61
Jars	0.00	0.04	7.33
Glass.Fragments	0.67	35.68	53.25
Other.2	0.00	0.01	1.52
Rubber	0.18	2.17	12.02
Flip.Flops	0.01	0.10	7.10
Rubber.Gloves	0.00	0.00	0.56
Tires	0.00	0.01	2.68
Rubber.Fragments	0.13	1.64	12.26
Other.3	0.02	0.05	2.43
Processed.Lumber	2.49	32.96	13.22
Cardboard.Cartons	0.00	0.01	1.25
Paper.and.Cardboard	0.80	1.57	1.96
Paper.Bags	0.00	0.00	0.67
Lumber.Building.Material	1.58	29.73	18.79
Other.4	0.04	0.09	2.26
Cloth.Fabric	0.25	4.30	17.48
Clothing...Shoes	0.02	0.10	5.13
Gloves.non.rubber.	0.00	0.00	0.27
Towels.Rags	0.00	0.00	0.74
Rope.Net.Pieces.non.nylon.	0.12	2.68	22.86
Fabric.Pieces	0.04	0.19	4.69
Other.5	0.04	0.03	0.72
Unclassified	0.06	0.31	4.83

6.6 Interpretation of analyses

6.6.1 Cluster Analysis

Cluster analysis is one way to determine whether there are distinct “types” of debris surveys; for example, sites that are dominated by a particular type or types of debris. Here we used a model based on a multinomial distribution to look for patterns in the NOAA Accumulation data, by clustering sites with similar frequency distributions of items. However, likely because of the large number of categories and surveys, the cluster analysis yielded an optimal number of 19 clusters, which is too many to have much functional utility for analysis. With a smaller number of clusters, it would be easier to draw conclusions about site types, and focus specific remediation actions based on the clustering. For example, sites dominated by lumber or large metal objects might be illegal dump areas, and one could target education or signage in the area to combat illegal dumping.

However, in this case, it is difficult to determine strong patterns in the clustering data. There were some clusters comprised almost exclusively of single objects, such as metal fragments, lumber, cigarettes, (e.g. clusters 2, 9, 16). Other surveys had one dominant object and an assortment of other items (e.g. clusters 4, 6, and 8). It is possible that because accumulation surveys are all conducted at very similar sites (e.g. flat sandy shorelines), there is actually relatively little variation between site types, which is why the clusters aren't separating in a more defined way.

In this case, cluster analysis might prove more useful if the items were grouped prior to analysis for simplicity. For instance, items could be grouped by source type, such as consumer items, fishing items, and illegal dumping items. A key consideration in this approach would be to use the subset of items for which categories can be assigned unambiguously.

6.6.2 Tree Models

The tree model shows that both County and Total Debris are significant predictors of the categorical composition of the debris found, however, there are some challenges in interpreting these two variables. While County may represent real differences, however, it is difficult to draw general understanding from this variable. County could be important due to differences in policies, populations, or other factors at the county level. However, it could also be important as it is correlated with spatial proximity among samples, and samples that are close together are likely to be similar. Total Debris is likely appearing due to a mixture of sampling error and true rarity of some categories. As the amount of debris collected at a site decreases, the chance of encountering rare categories decreases rapidly. Thus sites with less debris will likely only have a subset of categories with nonzero counts. Due to this sampling issue, Total Debris is appearing as a highly informative covariate in the first and second tree models, but is likely mostly representing sampling error. Excluding County and Total Debris from the analysis, we found that the spatial factors of Point Index and State, as well distance to the nearest road, were also significant covariates. These variables appeared as important in the earlier analysis of total load, and likely represent real patterns in driving variables among the sites.

Tree models will indicate which covariates are significant predictors of category composition, but it is more challenging to interpret the effects of these covariates on individual categories. The tree model does estimate the frequency distribution across the categories for each of its terminal nodes, which are the final sets of homogenous sites.

The GAM model provides a more nuanced understanding of the significance of these variables to different category types. Certain types of debris (plastic and processed lumber) are more prevalent when the nearest road is closer to the survey site, while others (glass and rubber) are more prevalent with increasing distance to road site (i.e. the further a site is from the road). There are also differences in the prevalence of categories along a spatial gradient. One factor that came out as significant at one node of the tree, as well as in one factor in the model (cloth), was Number of People Assisting. As previously discussed, this may be a result of observer saturation.

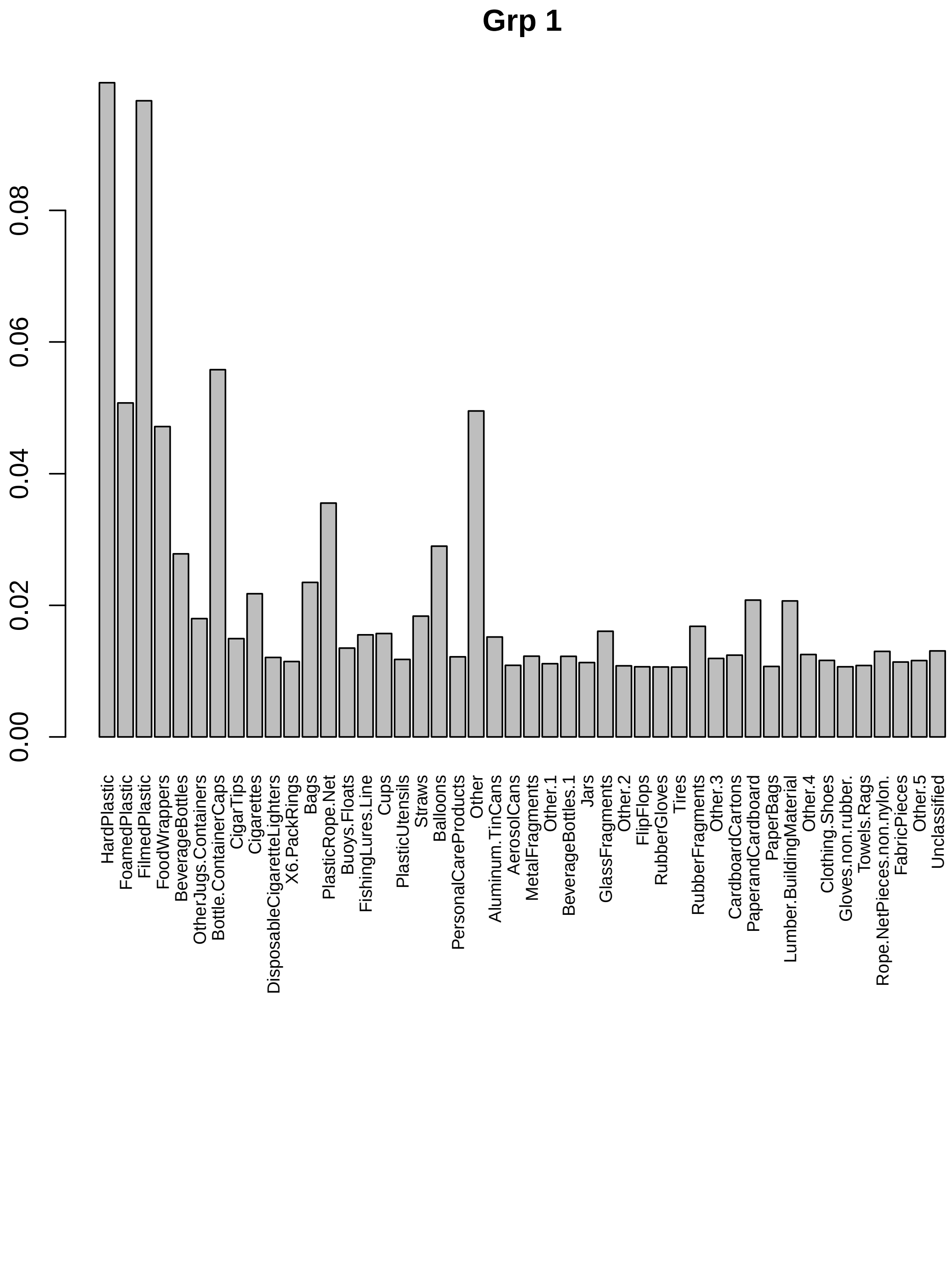
6.6.3 Variation among sites

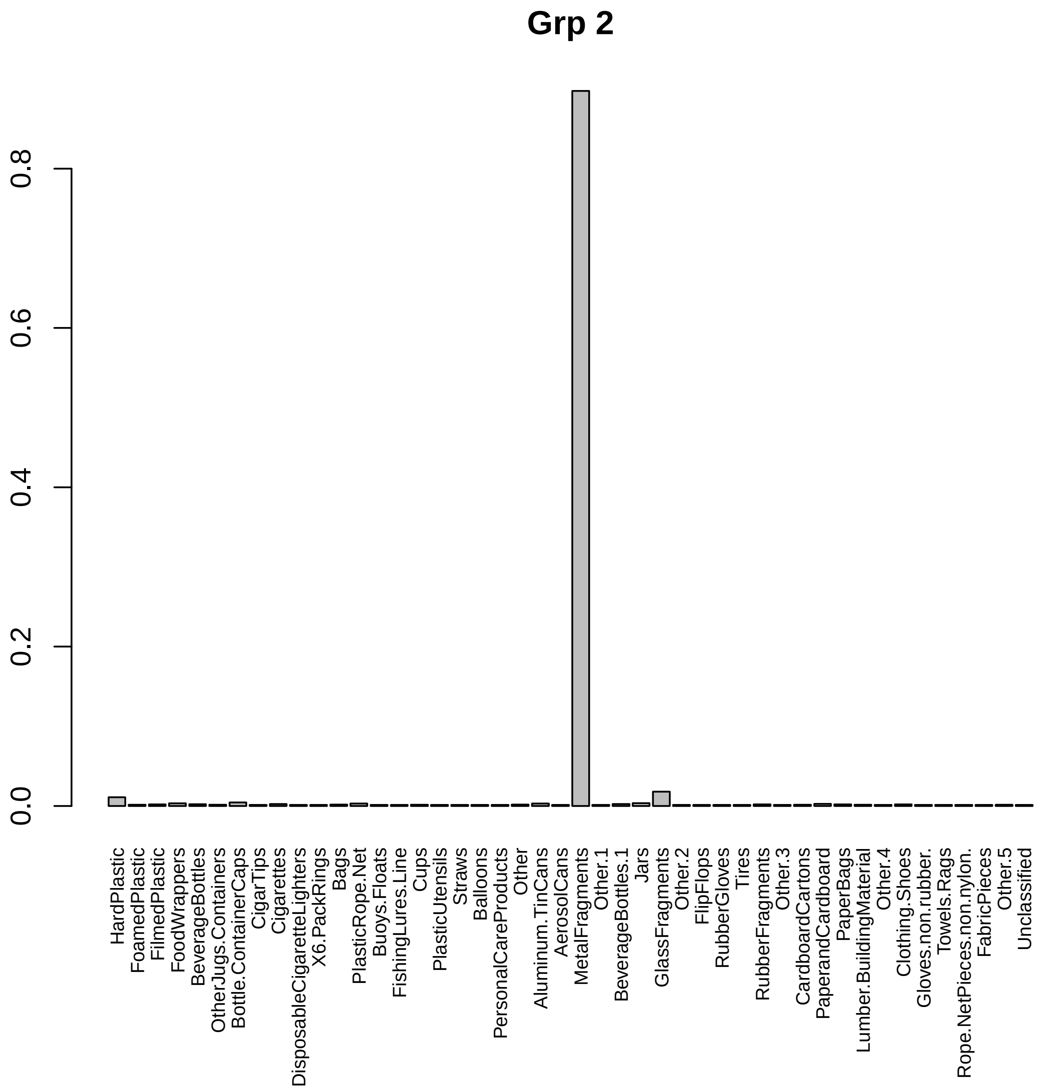
The analysis of cluster membership within and between sites and the separation analysis based on within and among site variance both demonstrate that there is higher variability between sites than within. The cluster analysis suggests that metrics describing the rank order distribution of category frequencies might be useful in describing the sites. Most sites cluster as having relatively even distributions of items across the categories, associated with group 1 in the analysis (Figure 6.5.1.b, c supplement). Other site types are have one or a few particularly common categories, with most other categories at relatively low counts, such as group 7 (Figure 6.5.1.b, c supplement).

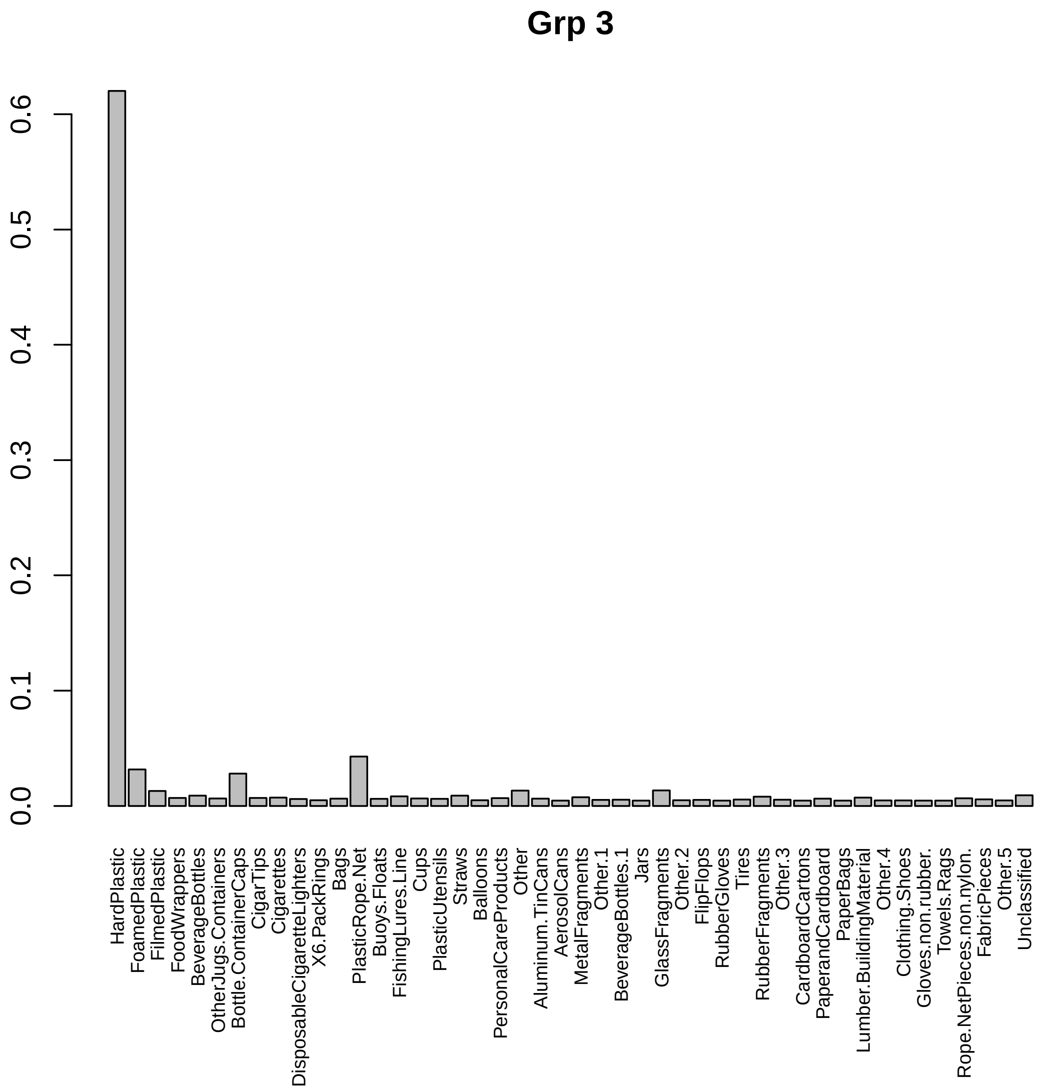
The separation analysis indicates that some categories have quite high separation, including: plastic, hard plastic, other jugs, bottle container caps, cigar tips, cigarettes, buoys and floats, other plastic, glass, beverage bottles, glass fragments, building material, and cloth fabric, rope net pieces. Six categories (six pack rings, rubber gloves, paper bags, non-rubber gloves, towels, other fabric) had a separation of less than zero, meaning that the within site variability was higher than the between site variability. It may be useful to use the category separations to tune data collection and/or analysis, as categories with higher separation have more utility for explaining between site variability. These separation values might also be driven by specific characteristics of the items. For instance, glass has a very high separation. This could be a function of some sites being easy to transport glass to, e.g. near picnic areas, and having a high chance of breakage and thus abandonment, e.g. being particularly rocky. Thus, the characteristics of the sites lead to a process with high variability among sites, but low variability within.

7 Supplemental Information

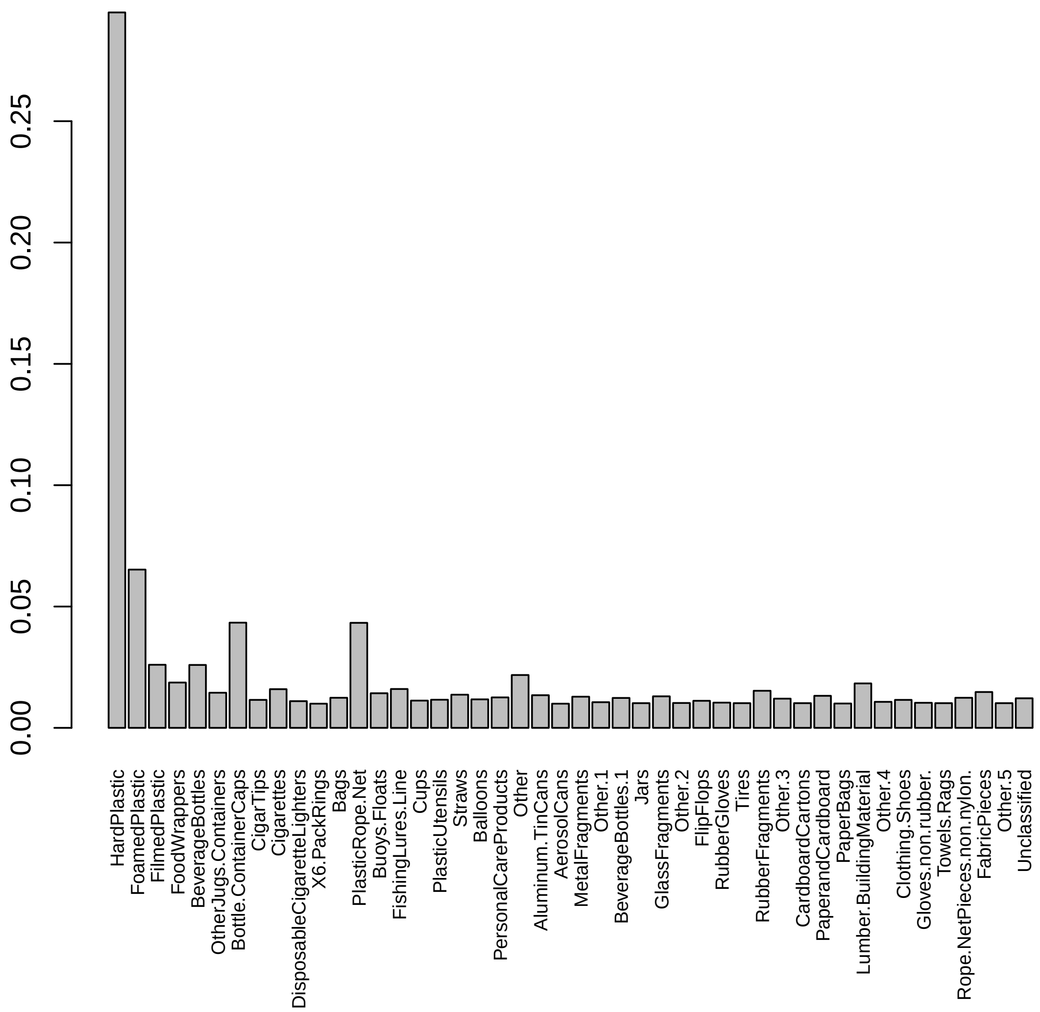
Below and in the pages that follow are each of the 19 groupings from cluster analysis based on the NOAA accumulation data (as per Figure 6.5.1.b). On the x axis are the category types and on the y axes are the relative frequency of each category within the cluster.



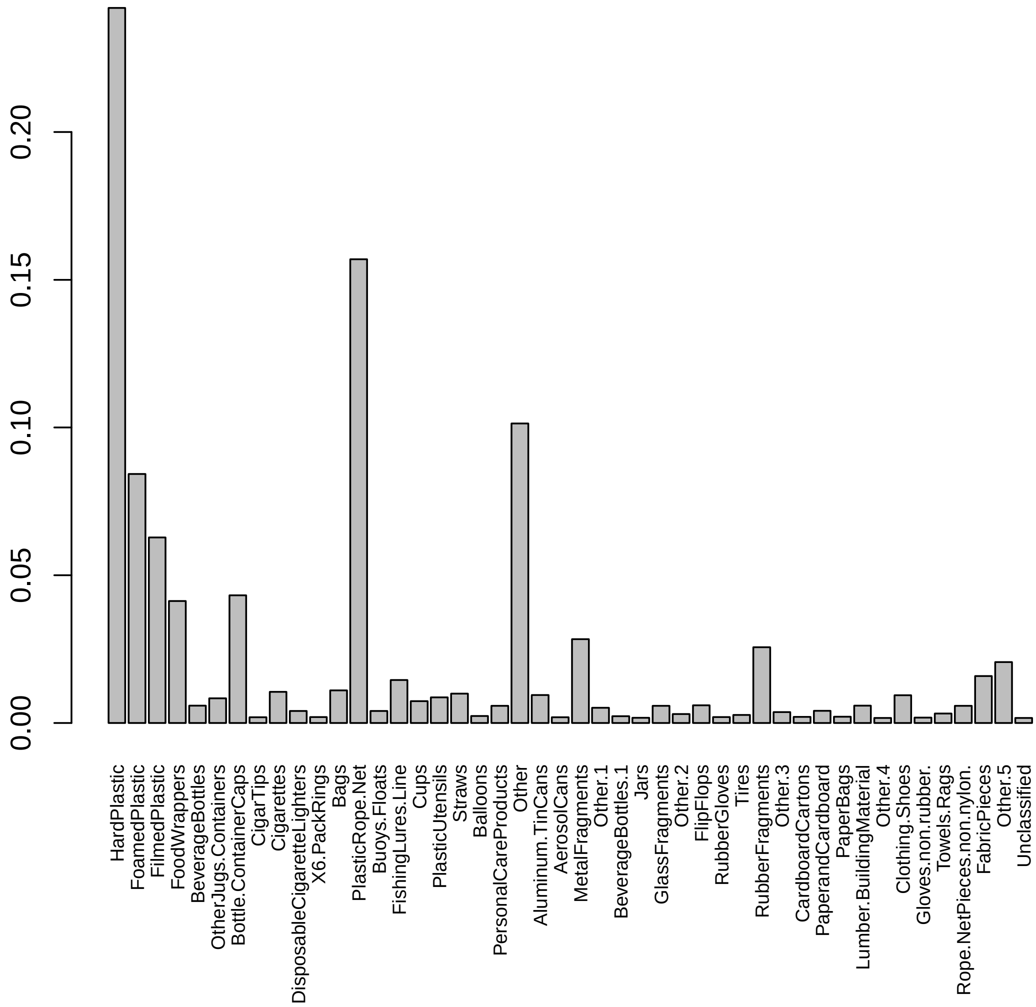

Grp 2
0.0
0.2
0.4
0.6
0.8
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

Grp 4

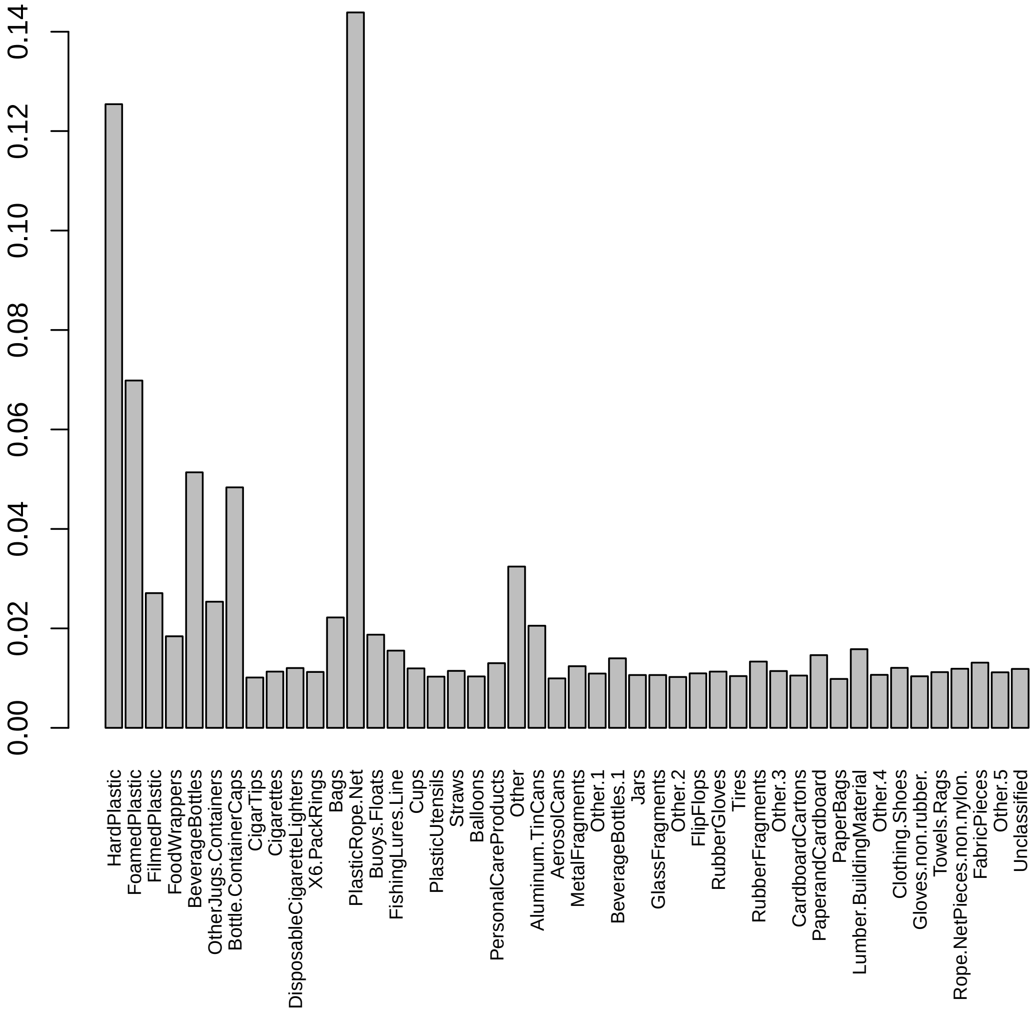

0.00
0.05
0.10
0.15
0.20
0.25
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

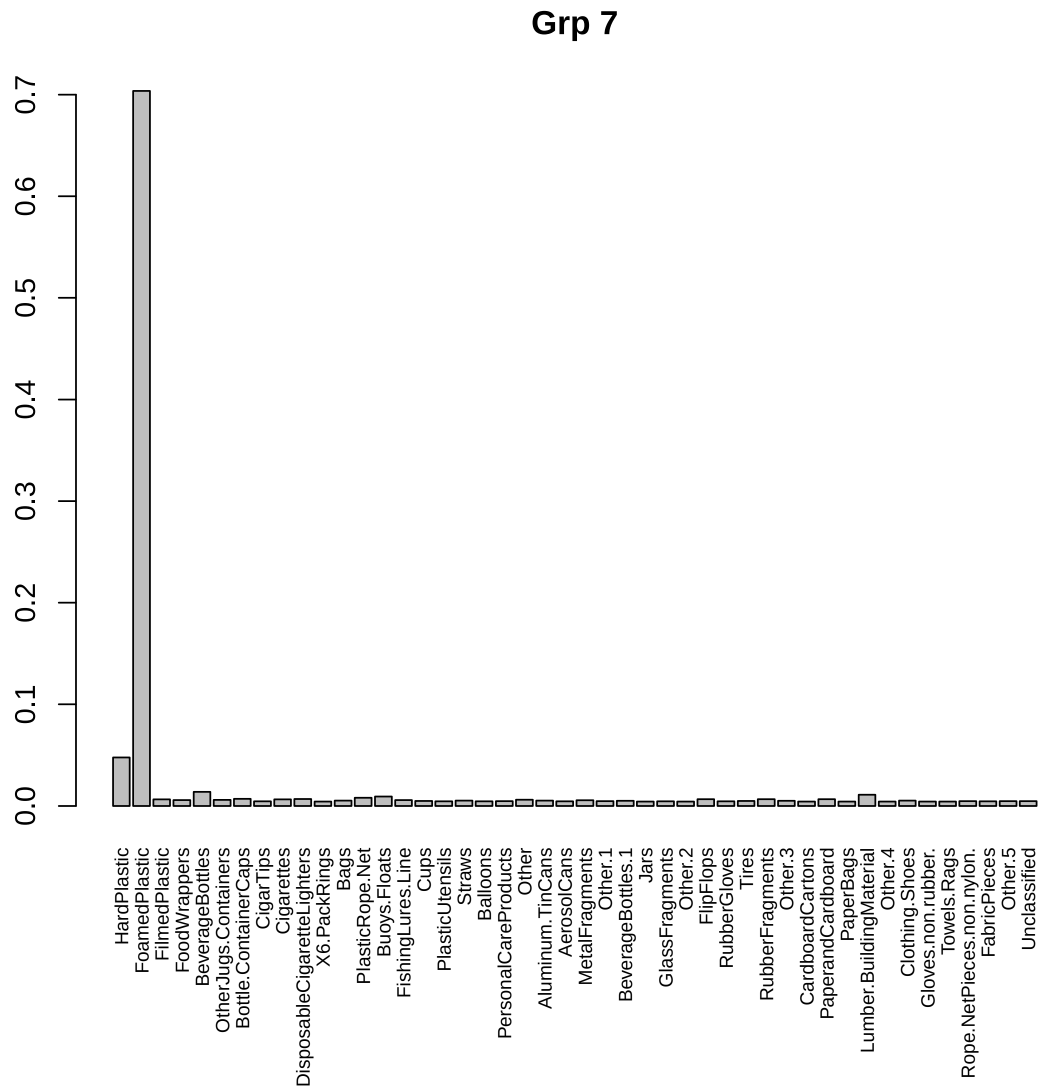
Grp 5



0.00
0.05
0.10
0.15
0.20
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

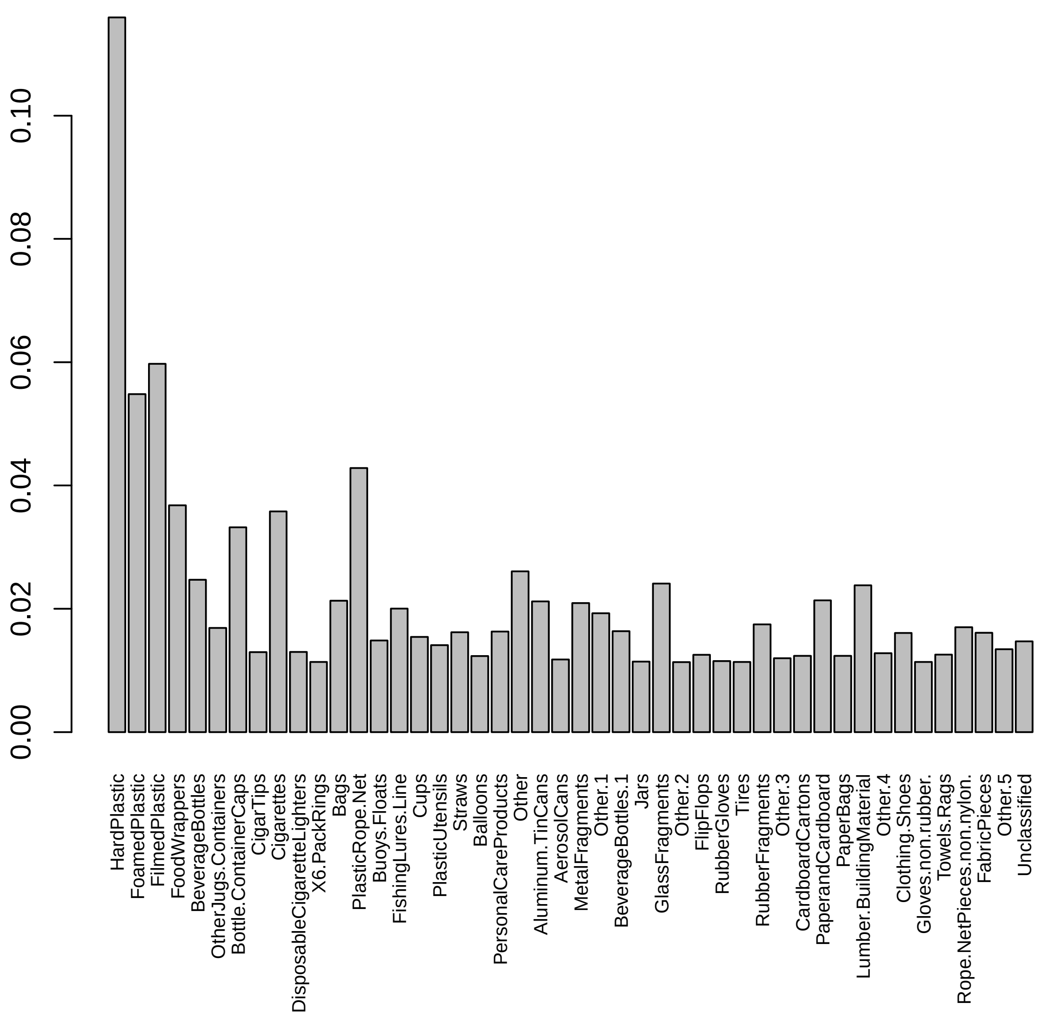
Grp 6


0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

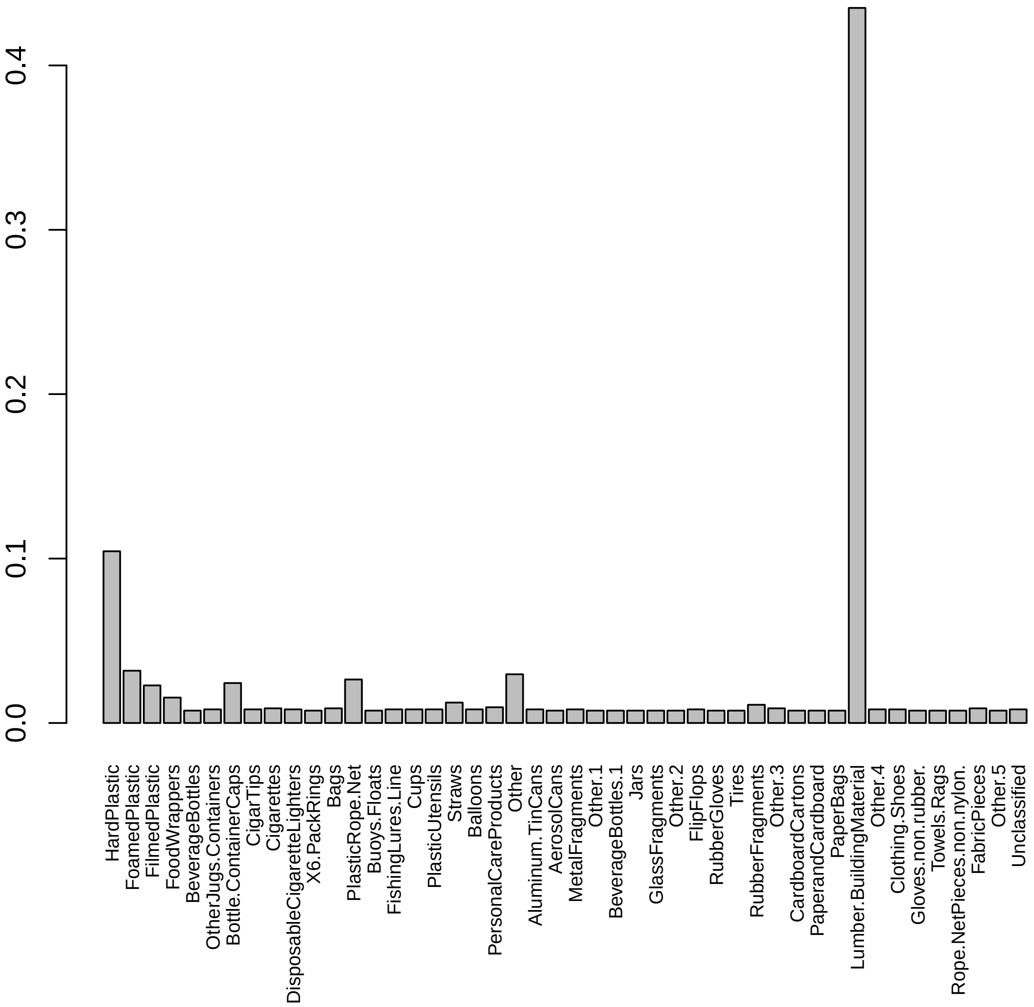


Grp 7
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

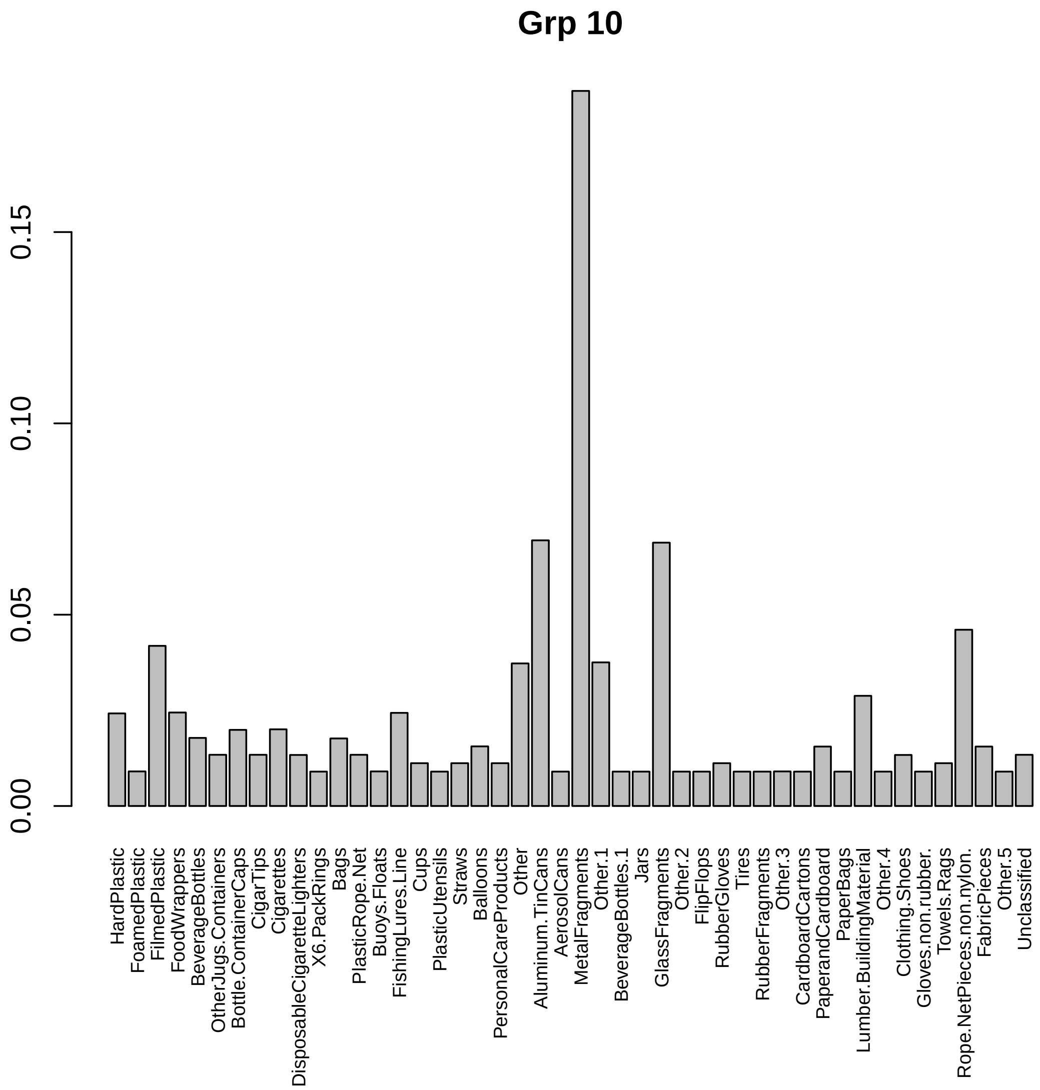
Grp 8


0.00
0.02
0.04
0.06
0.08
0.10
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

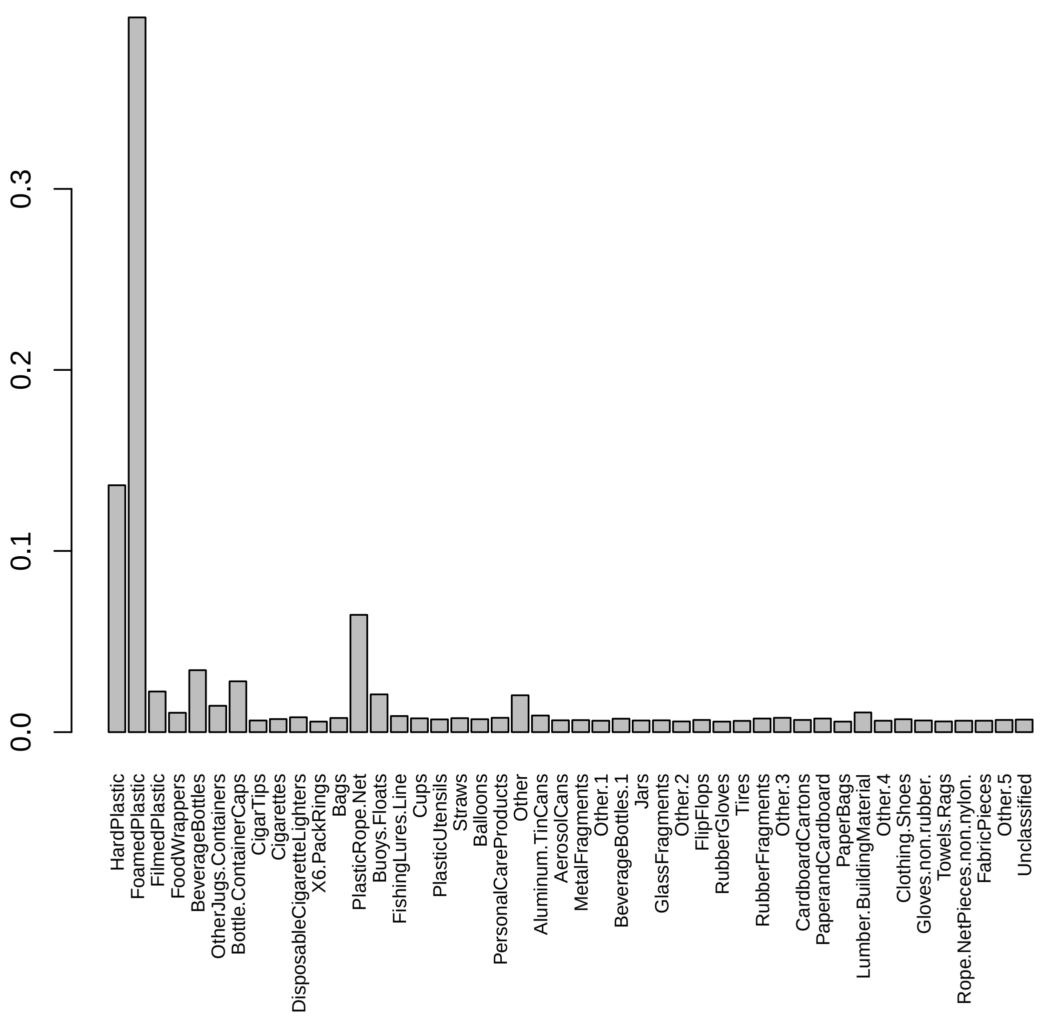
Grp 9



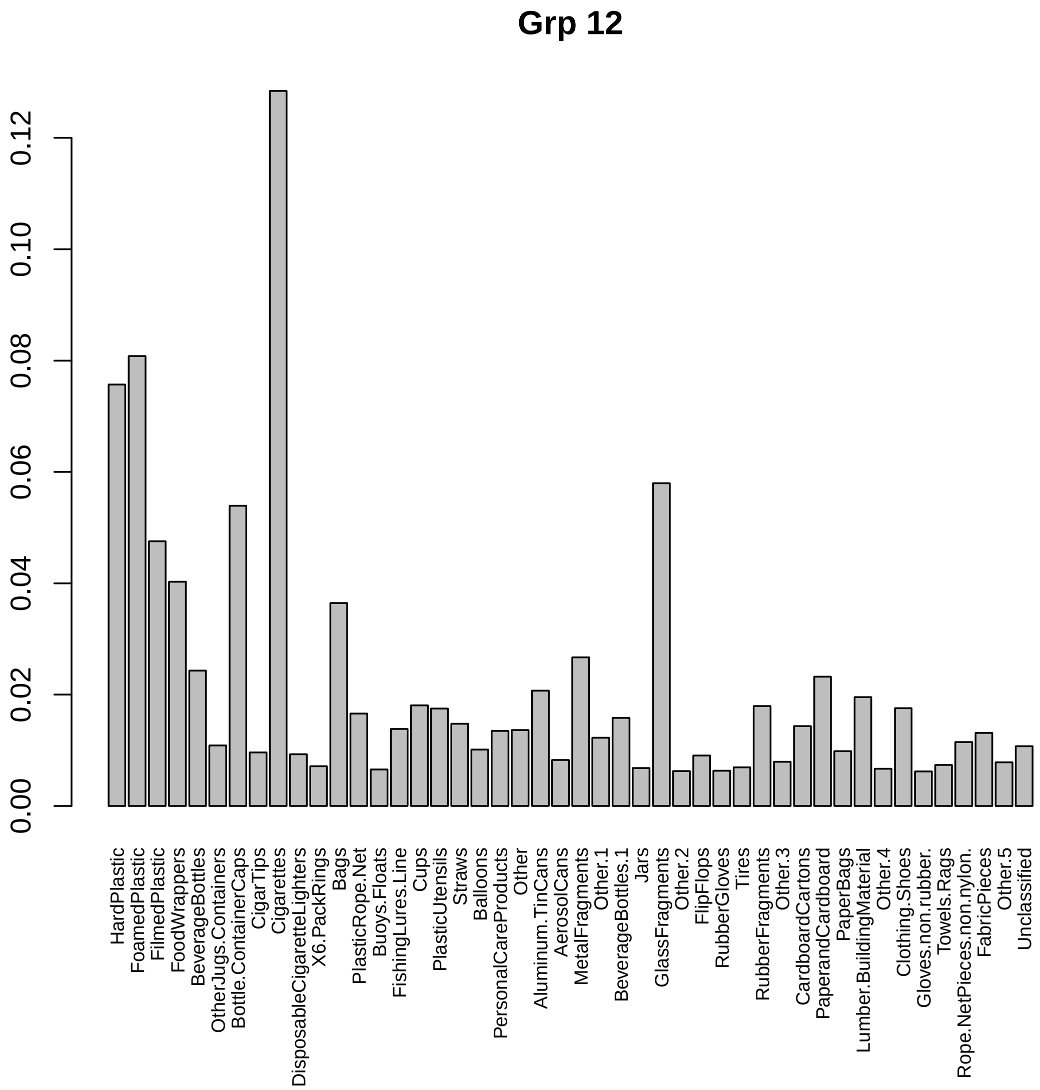
0.0
0.1
0.2
0.3
0.4
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 10
0.00
0.05
0.10
0.15
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

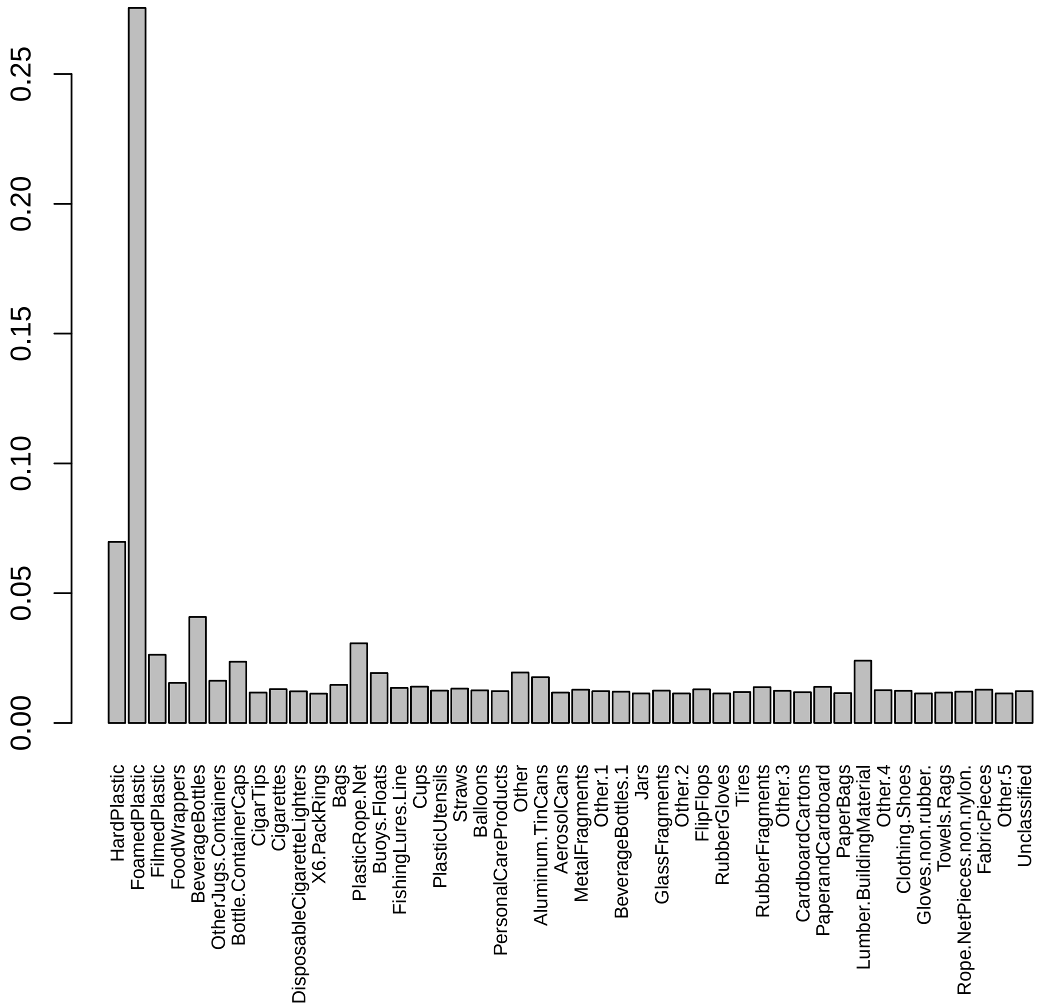
Grp 11

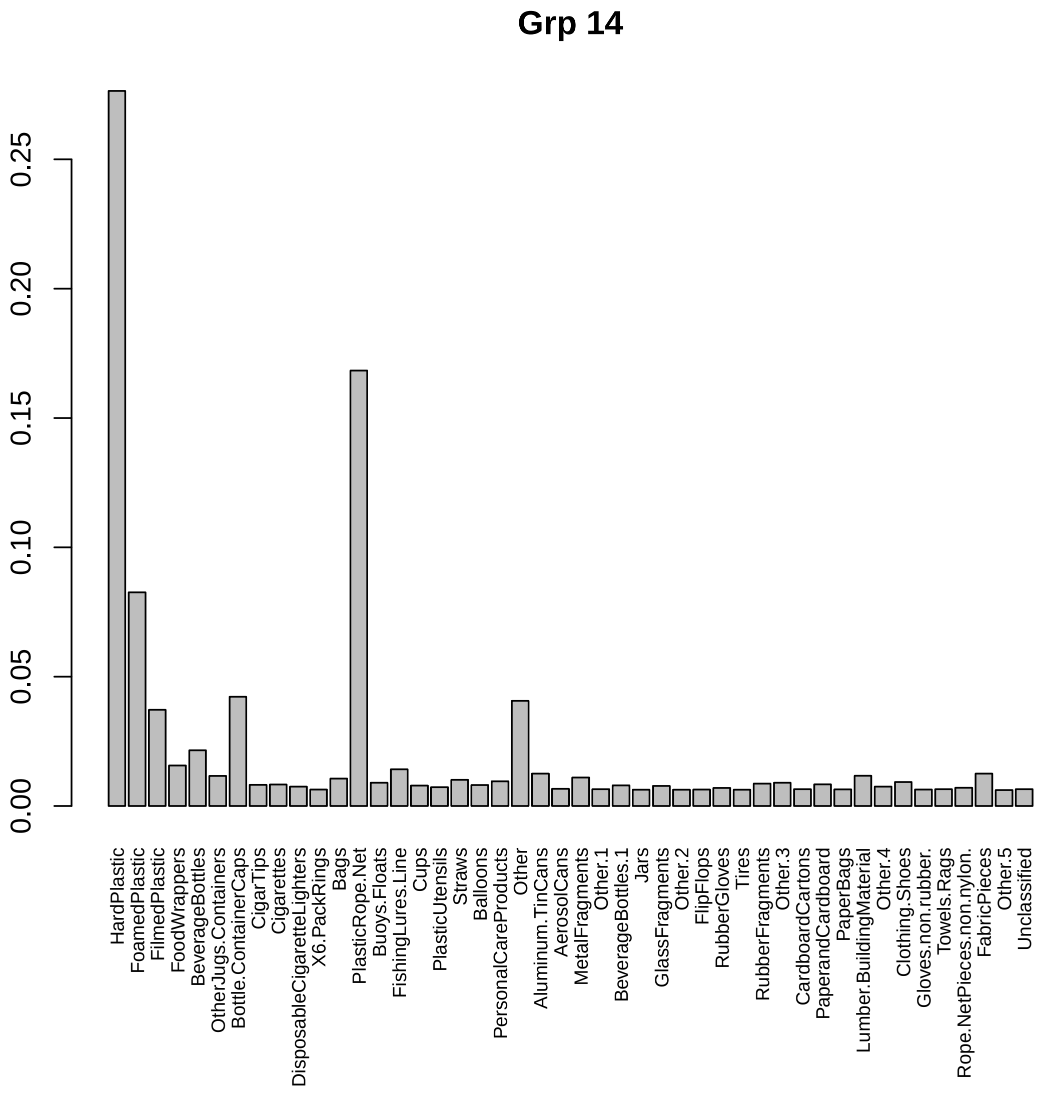


0.0
0.1
0.2
0.3
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

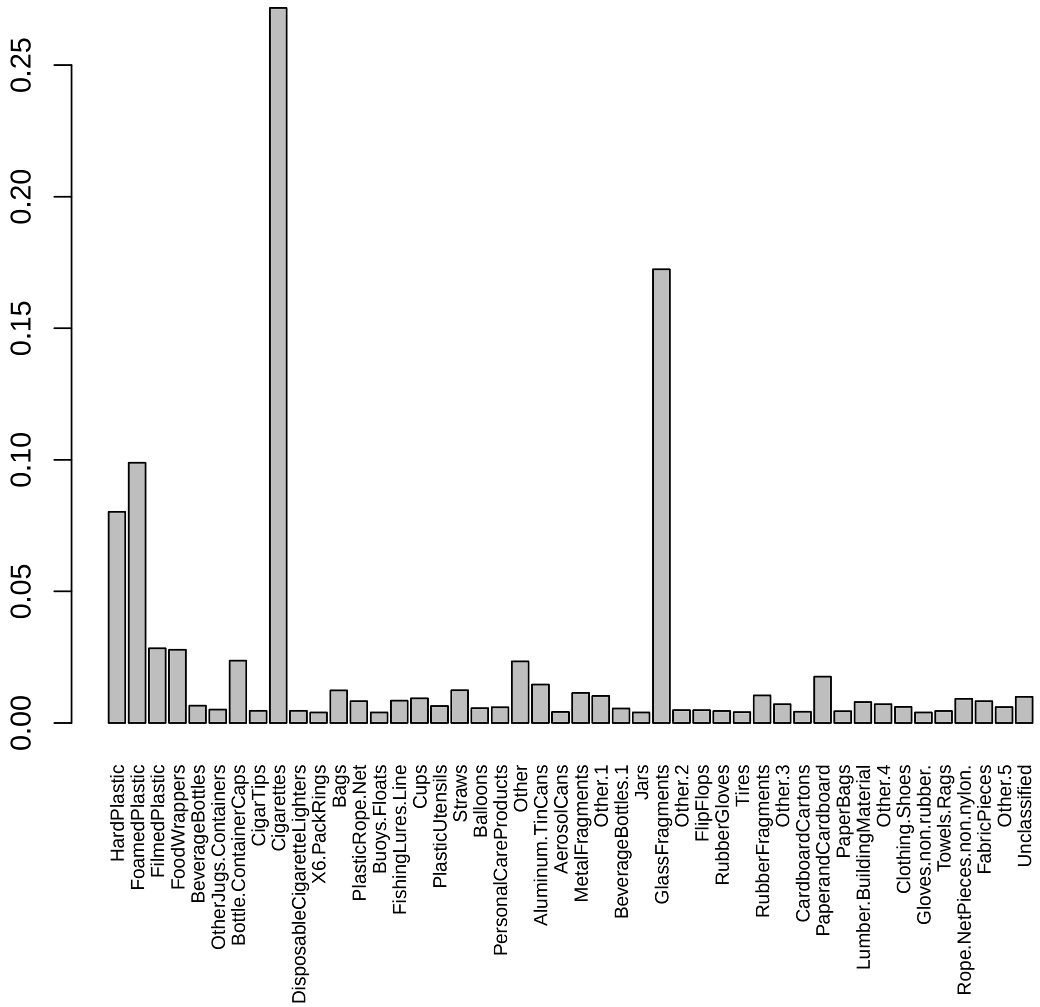

Grp 12
0.00
0.02
0.04
0.06
0.08
0.10
0.12
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

Grp 13

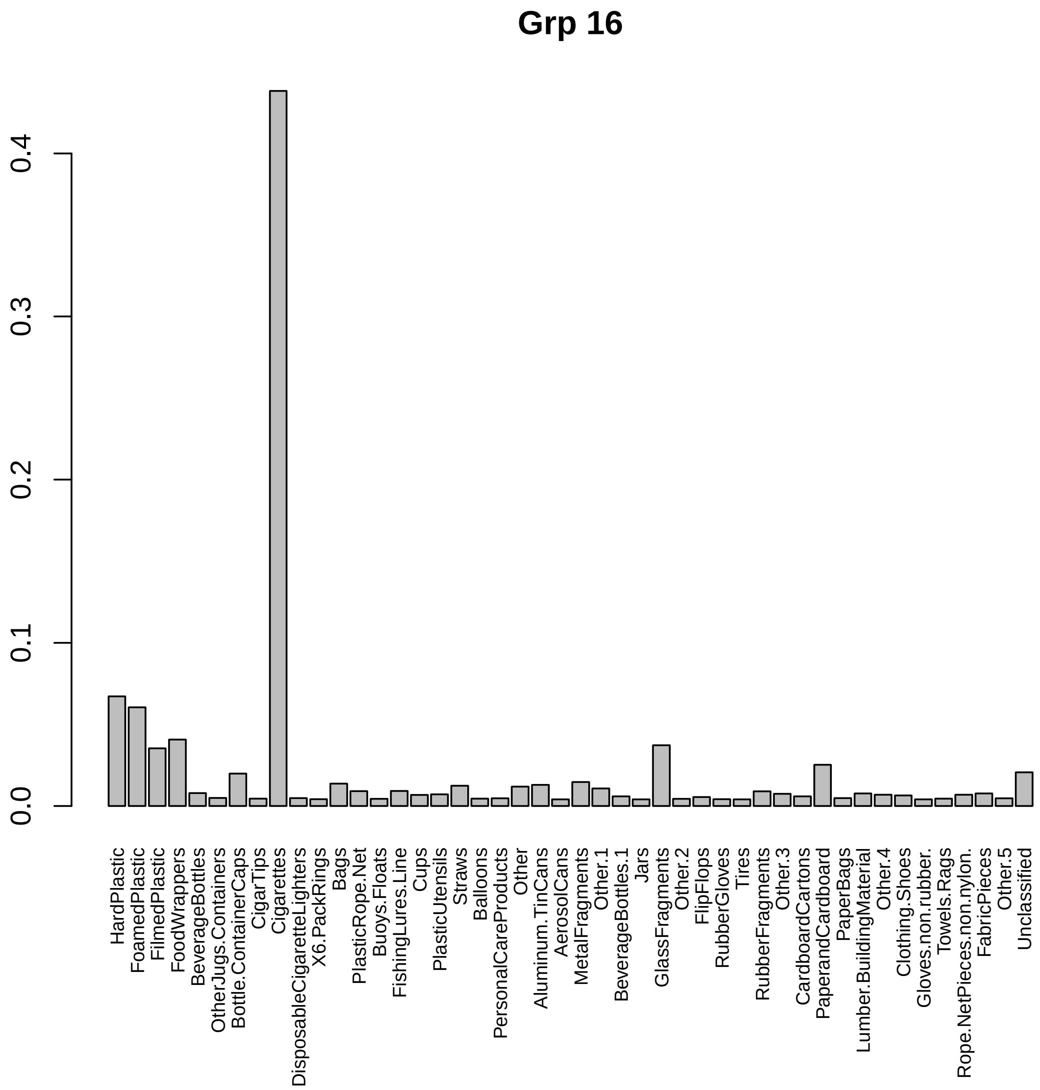

0.00
0.05
0.10
0.15
0.20
0.25
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 14
0.00
0.05
0.10
0.15
0.20
0.25
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

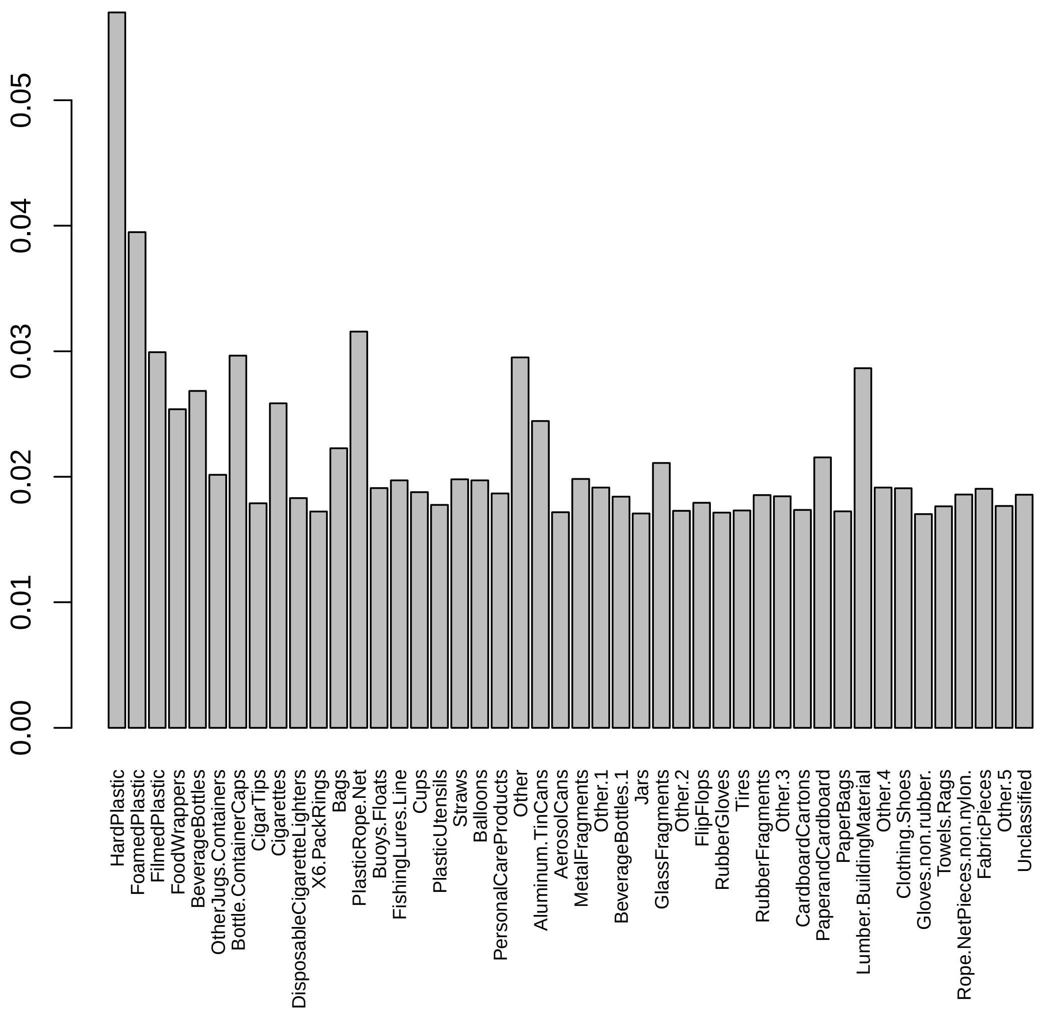
Grp 15

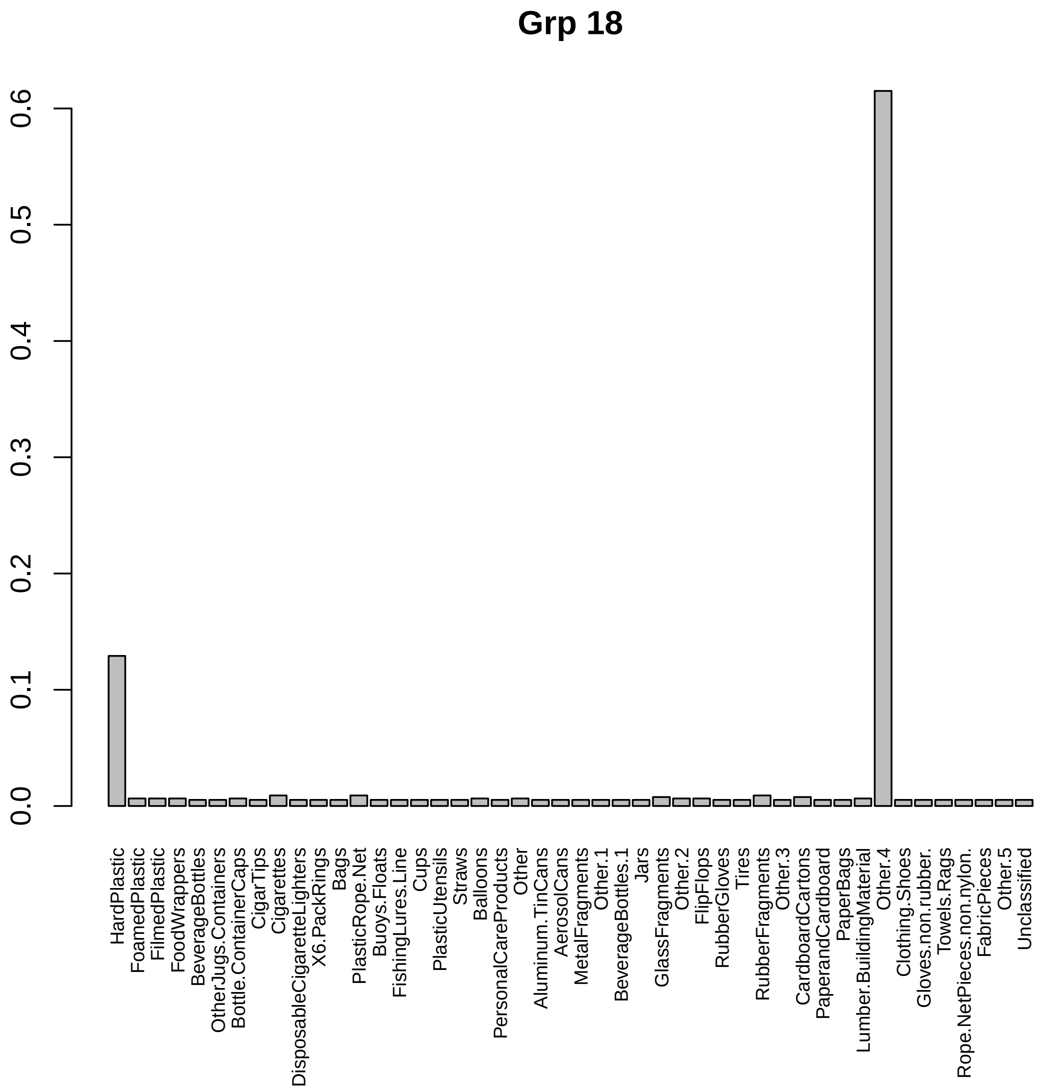


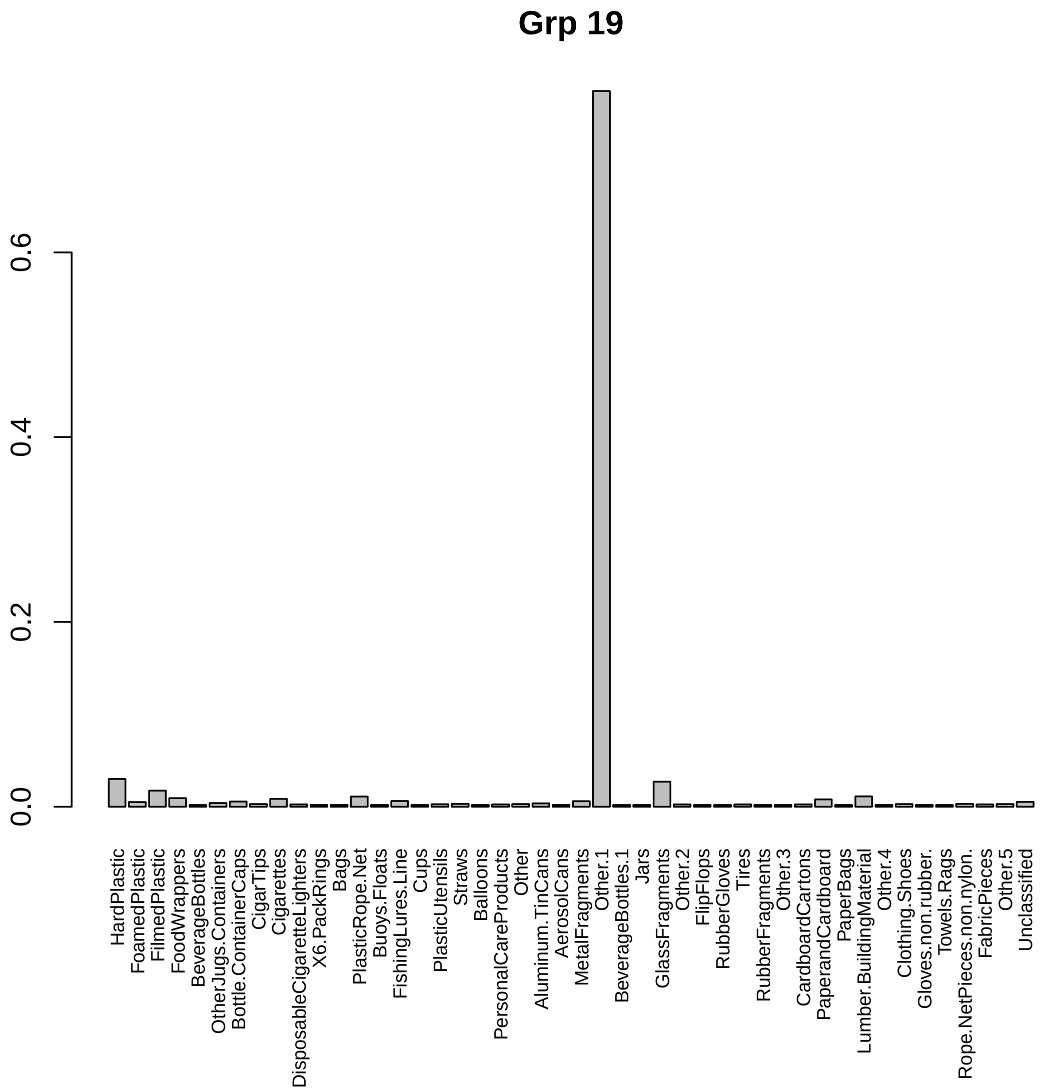
0.00
0.05
0.10
0.15
0.20
0.25
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 16
0.0
0.1
0.2
0.3
0.4
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified

Grp 17


0.00
0.01
0.02
0.03
0.04
0.05
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 18
0.0
0.1
0.2
0.3
0.4
0.5
0.6
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified


Grp 19
0.0
0.2
0.4
0.6
HardPlastic
FoamedPlastic
FilmedPlastic
FoodWrappers
BeverageBottles
OtherJugs.Containers
Bottle.ContainerCaps
CigarTips
Cigarettes
DisposableCigaretteLighters
X6.PackRings
Bags
PlasticRope.Net
Buoys.Floats
FishingLures.Line
Cups
PlasticUtensils
Straws
Balloons
PersonalCareProducts
Other
Aluminum.TinCans
AerosolCans
MetalFragments
Other.1
BeverageBottles.1
Jars
GlassFragments
Other.2
FlipFlops
RubberGloves
Tires
RubberFragments
Other.3
CardboardCartons
PaperandCardboard
PaperBags
Lumber.BuildingMaterial
Other.4
Clothing.Shoes
Gloves.non.rubber.
Towels.Rags
Rope.NetPieces.non.nylon.
FabricPieces
Other.5
Unclassified