Data-Driven Stories Stemming From Public Participation in Science

Jazzmine Waugh

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Julia Parrish, Chair

Jennifer Ruesink

Abigail Swann

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Data-Driven Stories Stemming From Public Participation in Science

Jazzmine Waugh

Chair of the Supervisory Committee:
Julia Parrish
School of Aquatic and Fishery Sciences

Seabirds are facing increasing anthropogenic stress in the current era of climate change. Anthropogenic stressors can be direct, such as oil spills increasing seabird mortality via intake of toxins, or indirect, such as marine heat waves decreasing seabird prey quality and quantity. I present research exploring the impacts of anthropogenic stressors on seabird populations, using data from a long-term, fine-grained citizen science dataset. Citizen science, where members of the public help collect data for scientific research, is a method of obtaining vast quantities of rigorous data. It also acts as a potential path for people who have historically been underrepresented in science to be included in science. In this work, I first explore the impacts of direct and indirect anthropogenic stressors on seabirds. In my first chapter, I identify seabird species that are relatively more susceptible to the direct stressor of oil spills, which can be used to help prioritize species for protection. The second chapter uses the natural history stressor of flight feather molt to explore how natural history and a variety of environmental stressors can combine to influence mortality of common murres (*Uria aalge*) and find that timing is key to

determining how these stressors affect mortality. I then turn to consider the diversity of the citizen science participants who collect data like that used in my ecological research. In the third chapter, I quantify how the demographics of citizen science participants have changed over time and differ across types of projects, highlighting opportunities for increased diversity in citizen science. Taken together, my work underscores the use of citizen science both as a means of monitoring changes in populations of marine organisms and as a method of increasing representation for groups that have historically been underrepresented in science.

TABLE OF CONTENTS

Acknowledgements……..3

Introduction……....4

Chapter 1. Using beached bird data to assess seabird oiling susceptibility ……...……...…13

Abstract…….….13

Introduction……13

Methods……..17

Results……29

Discussion……..37

Supplementary Material……50

Chapter 2. Elevated seabird mortality associated with cumulative environmental and natural history stressors ...…….……..……...…...……..77

Abstract……..77

Introduction……78

Methods……..82

Results……94

Discussion……105

Supplementary Material……...……116

Chapter 3. Demographics of public participation in science: A meta-analytic approach..136

Abstract……136

Introduction……..137

Methods……139

Results……..149

Discussion……161

Conclusion……...166

Supplementary Material……...174

ACKNOWLEDGEMENTS

Thank you first and foremost to my committee: Julia, Abby, Jen, and Beth. The four of you have taken me from a science fledging to a full-grown scientist.

Thank you to my community of scientists, both traditional and not. This includes, but is not limited to, the folks in my lab (Tim, Jackie, Charlie, Hillary, and Yurong), my interns (Lindsey, Max, and William), my cohort, and the citizen science participants who helped collect the data in this work.

Finally, enormous thanks to my husband, cats, and other family. I would never have made it this far if not for your endless support.

INTRODUCTION

As a seabird ecologist and data scientist, I am broadly interested in discovering data-driven stories about when and how ecological stressors impact marine organisms. I define an ecological stressor as pressure coming from unusual environmental conditions that may exert population-level influences (e.g., decreased survival rates of sea otters [*Enhydra lutris*] in Alaska following the *Exxon Valdez* oil spill- Monson et al. 2000; high mortality of Brazilian coastal area coral reefs following a heat stress event- Pereira et al. 2022). When I started this dissertation, I was specifically interested in exploring the impacts of anthropogenic (human-caused) stressors on seabird mortality. Anthropogenic stressors can be direct, such as oil spills that can kill birds via ingestion of toxins (Troisi et al. 2016), or indirect, such as climate change.

In the last decade, climate change has been linked to increased mortality and decreased reproductive success for marine mammals in United States waters (Gulland et al. 2022), declines in recruitment and abundance of intertidal invertebrates in the western North Atlantic (Petraitis & Dudgeon 2020), and changes in abundance and recruitment of fish populations in North America (Lynch et al. 2016). Within seabirds, climate change and specifically the increasing frequency of marine heat waves (Frölicher et al. 2018) has been associated with mass seabird die-offs (e.g., Piatt et al. 2020, Schoen et al. 2022, Jones et al. 2023). I am interested in adding to this corpus exploring how anthropogenic stressors are impacting marine organisms.

Seabirds are a good vehicle for this research for several reasons. They have been proposed as an example of "sentinel indicators" of marine ecosystem health (e.g., Dias et al. 2019, Velarde et al. 2019, Sydeman et al. 2021). This is because marine birds are relatively easy to observe, capture,

and track (Velarde et al. 2019); and changes in their demography have been linked to changes in the physical ocean (e.g., marine heatwaves—Piatt et al. 2020) and in the ocean food web (e.g., abundance of fish stock—Horn & Whitcombe 2015). Second, seabird datasets on the order of a decade or longer are relatively abundant, making it easier to assess patterns of population response to longer-term cyclic patterns (e.g., changes in fledgling and adult survival of Nazca Booby [*Sula granti*] in the Galápagos associated with the Pacific Decadal and El Niño Southern Oscillations—Champagnon et al. 2018) and/or episodic stressor events (e.g., mass mortality of Cassin's auklets [*Ptychoramphus aleuticus*] in the NE Pacific associated with a marine heatwave—Jones et al. 2018).

The chapters in my dissertation incorporate data from the Coastal Observation and Seabird Survey Team (COASST). COASST participants monitor beached seabird carcasses (i.e., dead birds that are found along the coast). Participants survey their local beach monthly, sending data to COASST for expert verification. This dataset is ideal for my purposes because of the spatial extent (from the Chukchi sea area in Alaska south to the Mendocino area in California), longevity (~20 years), expert verification of the data, and the detailed data at the level of the individual carcass (e.g., morphometrics; age, sex, species; carcass condition - see Parrish et al. 2017 for details)

COASST is a citizen science project. Citizen science projects "…focus on, but are not limited to, nonscientists… participating in the processes of scientific research, with the intended goal of advancing and using scientific knowledge" (National Academies of Sciences, Engineering, and Medicine 2018), and can be a way of producing large quantities of data (Hampton et al. 2013, Kobori 2016). Rigorous citizen science data, which I define as data that are effort-standardized

(i.e., collected at specific times, locations and/or for specific durations by a known number of data collectors) and verified for accuracy, has been used to monitor changes in the marine environment (e.g., detecting presence of invasive and native crabs at sites from Maine to New Jersey- Delaney et al. 2008; documenting the scale of marine debris along the coast of New South Wales, Australia- Smith & Edgar 2014; monitoring trends in seabird and marine mammal entanglement rates in California- Donnelly-Greenan et al. 2019). Citizen science data enables me to investigate patterns across broad geographic areas and long time periods, incorporating fine-grain information on the scale of individual organisms. Data across these scales allows me to investigate population-level changes in demographics of marine organisms by teasing out the impacts of environmental stressors across time and geographic space.

I brought to this work several research goals. One goal is from the point of view of the data. I wanted to use large, rigorous datasets to explore how seabirds are impacted by their physical environment. My other goal was one of equity, specifically how to involve more people in scientific research, particularly those historically excluded from science. It was not clear to me at the outset of my PhD how to merge these interests. Eventually, I realized citizen science—explicitly involving the broader public—could act as a bridge between them. My work investigates multiple facets of citizen science in the marine environment, including monitoring of marine ecosystems and exploring the community of people who collect these data.

In my first chapter, I chose to look at a direct anthropogenic stressor in the marine environment, investigating whether seabird species are differentially vulnerable to oil spills. While prior studies have suggested variation in oiling susceptibility across species, findings primarily came from combining expert opinion and best-available information at the species level (e.g., natural

history traits, conservation status, demographic factors) to produce a theoretical value representing species-specific relative susceptibility (e.g., King & Sanger 1979, Williams et al. 1995). To calculate a more empirical index of susceptibility, I compared data on dead seabirds found after two oil spills in the northern California Current System (the *Nestucca* and *Tenyo Maru*) to taxon-specific carcass occurrence data from the same spatial and seasonal window from the COASST dataset. To investigate how taxonomic composition during spills compares to normal variation for the spatial and seasonal window in which the spill occurred, I examined whether taxonomic composition differed between spill and non-spill years. I then explored which species were found at relatively high (or low) levels following spills when compared to non-spill years, as a way of calculating relative susceptibility to oiling. As oil spills can have devastating impacts on seabird populations via decreased breeding success (e.g., Barros et al. 2014) and increased mortality (e.g., Munilla et al. 2011), this research can help identify which birds most need to be prioritized for protection from spills. Where appropriate beached bird data is available, this relative susceptibility method can be extended to other places, other times of year, and potentially other kinds of localized threats (e.g., offshore wind farming, fishery bycatch).

In my second chapter, I investigate how a natural history stressor (flight feather molt) interacts with environmental stressors to impact mortality in common murres (*Uria aalge*, hereafter murres). Murres were selected in part because they are prevalent in the COASST database (n = 28,723 carcasses, or ~29% of all carcasses found by COASST). In addition, murres are a synchronously molting species (Bridge 2006), meaning they drop all flight feathers over a period of days, and COASST data on wing chord measurement can be used to detect molt occurrence over time. Synchronous molting has been hypothesized as more energetically costly per unit time

than molting more slowly (Guillemette et al. 2007). Synchronously molting birds may be particularly stressed by molt due to this high energy cost, making them well-suited for studying interactions between molt as a "natural history stressor" and environmental stressors. Using COASST data, I first created annual models of the proportion of dead birds in molt over the day of the year, to explore interannual variation. Secondly, I created models linking the annual proportion of dead birds found in molt to environmental stressors to test hypotheses that in years with more environmental stressors (e.g., poor habitat quality proxied by cold water habitat availability, high average amount of storminess proxied by wave height), a greater proportion of beached birds during the molt season will have died in molt. This research suggests that in murres, and other synchronously molting birds, the natural history stressor of molt can interact with environmental stressors-- but only up to a point. When environmental stressors early in the year are so stressful that many birds die before reaching molt, molt has a weaker influence as a stressor.

In my third chapter, I pivot to consider the people who collect the kind of data I used in my ecological chapters: citizen science participants. As I considered using citizen science data as a bona fide vehicle for my ecological research, I noticed that it was advanced as a method of increasing diversity in science (e.g., Pandya 2012, Paleco et al. 2021). I began to wonder if citizen science participants could be considered a diverse group. I decided to build upon pre-existing research to get a sense of the diversity in participant demographics, as a function of project type and time since these original claims of increasing diversity were made (e.g., Pandya 2012). Using a literature-based meta-analysis approach with data from citizen science projects, I calculated descriptive statistics, compared project participants against census data for their

relevant geographies, and modeled the relationship between project demographics and project descriptors (e.g., type of scientific data the project collects). Increasing diversity in citizen science could benefit groups that have been historically underrepresented in science (e.g., by giving these groups more of a voice in the research questions that get addressed-Bonney et al. 2016) and scientific research itself (e.g., projects with primarily affluent, well-educated participants may collect data that is not representative of the geographic area over which scientific inference is being made– Blake et al. 2020). My research elucidates how demographics have changed over time, and points towards areas (e.g. particular project types) where further progress could be made.

I believe that bringing in a broad spectrum of people to science should be a goal of the scientific community, both to assist in the collection of high-quality data to monitor changes in a fast-changing world, and to increase the breadth of perspectives and knowledge included in science. My work highlights the potential of citizen science both as a tool to monitor environmental change and as a bridge to a greater diversity of people in science. It contributes to evidence that trained citizen science participants can be a conduit for large quantities of rigorous, relevant data. At the same time, my research suggests that there are opportunities for further inclusion of groups that are currently underrepresented in science. It is my opinion that the scientific community can, and should, use citizen science to further these goals simultaneously.

References

Barros, Á., Álvarez, D., & Velando, A. (2014). Long-term reproductive impairment in a seabird after the Prestige oil spill. Biology Letters, 10(4), 20131041.

Blake, C., Rhanor, A., & Pajic, C. (2020). The demographics of citizen science participation and its implications for data quality and environmental justice. *Citizen Science: Theory and Practice*, *5*(1).

Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public understanding of science*, *25*(1), 2-16.

Bridge, E. S. (2006). Influences of morphology and behavior on wing-molt strategies in seabirds. *Mar Ornithol*, *34*, 7-19.

Champagnon, J., Lebreton, J. D., Drummond, H., & Anderson, D. J. (2018). Pacific Decadal and El Niño oscillations shape survival of a seabird. Ecology, 99(5), 1063-1072.

Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions, 10, 117-128.

Dias, M. P., Martin, R., Pearmain, E. J., Burfield, I. J., Small, C., Phillips, R. A., ... & Croxall, J. P. (2019). Threats to seabirds: a global assessment. *Biological Conservation*, *237*, 525-537.

Donnelly-Greenan, E. L., Nevins, H. M., & Harvey, J. T. (2019). Entangled seabird and marine mammal reports from citizen science surveys from coastal California (1997–2017). *Marine Pollution Bulletin*, *149*, 110557.

Dubansky, B., Whitehead, A., Miller, J. T., Rice, C. D., & Galvez, F. (2013). Multitissue molecular, genomic, and developmental effects of the Deepwater Horizon oil spill on resident Gulf killifish (Fundulus grandis). *Environmental Science & Technology*, *47*(10), 5074-5082.

Frölicher, T. L., Fischer, E. M., & Gruber, N. (2018). Marine heatwaves under global warming. Nature, 560(7718), 360-364.

Guillemette, M., Pelletier, D., Grandbois, J. M., & Butler, P. J. (2007). Flightlessness and the energetic cost of wing molt in a large sea duck. Ecology, 88(11), 2936-2945.

Gulland, F. M., Baker, J. D., Howe, M., LaBrecque, E., Leach, L., Moore, S. E., ... & Thomas, P. O. (2022). A review of climate change effects on marine mammals in United States waters: Past predictions, observed impacts, current research and conservation imperatives. *Climate Change Ecology*, *3*, 100054.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., ... & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156-162.

Horn, M. H., & Whitcombe, C. D. (2015). A shallow-diving seabird predator as an indicator of prey availability in southern California waters: A longitudinal study. *Journal of Marine Systems*, *146*, 89-98.

Jones, T., Parrish, J. K., Peterson, W. T., Bjorkstedt, E. P., Bond, N. A., Ballance, L. T., ... & Harvey, J. (2018). Massive mortality of a planktivorous seabird in response to a marine heatwave. *Geophysical Research Letters*, *45*(7), 3193-3202.

Jones, Timothy, Julia K. Parrish, Jacqueline Lindsey, Charlie Wright, Hillary K. Burgess, Jane Dolliver, Lauren Divine et al. "Marine bird mass mortality events as an indicator of the impacts of ocean warming." Marine Ecology Progress Series (2023).

King, J. G., & Sanger, G. A. (1979). Oil vulnerability index for marine oriented birds. *Conservation of marine birds of Northern North America*, *11*, 227-239.

Kobori, H., Dickinson, J. L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., ... & Miller-Rushing, A. J. (2016). Citizen science: a new approach to advance ecology, education, and conservation. *Ecological research*, *31*, 1-19.

Lynch, A. J., Myers, B. J., Chu, C., Eby, L. A., Falke, J. A., Kovach, R. P., ... & Whitney, J. E. (2016). Climate change effects on North American inland fish populations and assemblages. *Fisheries*, *41*(7), 346-361.

Munilla, I., Arcos, J. M., Oro, D., Álvarez, D., Leyenda, P. M., & Velando, A. (2011). Mass mortality of seabirds in the aftermath of the Prestige oil spill. Ecosphere, 2(7), 1-14.

Monson, D. H., Doak, D. F., Ballachey, B. E., Johnson, A., & Bodkin, J. L. (2000). Long-term impacts of the Exxon Valdez oil spill on sea otters, assessed through age-dependent mortality patterns. *Proceedings of the National Academy of Sciences*, *97*(12), 6562-6567.

National Academies of Sciences, Engineering, and Medicine 2018. Learning Through Citizen Science: Enhancing Opportunities by Design. Washington, DC: The National Academies Press. https://doi.org/10.17226/25183.

Paleco, C., García Peter, S., Salas Seoane, N., Kaufmann, J., & Argyri, P. (2021). Inclusiveness and diversity in citizen science. *The science of citizen science*, *261*.

Pandya, R. E. (2012). A framework for engaging diverse communities in citizen science in the US. Frontiers in Ecology and the Environment, 10(6), 314-317.

Parrish, J. K., Litle, K., Dolliver, J., Hass, T., Burgess, H. K., Frost, E., ... & Jones, T. (2017). Defining the baseline and tracking change in seabird populations: The Coastal Observation and Seabird Survey Team (COASST). In Citizen science for coastal and marine conservation (pp. 19-38). Routledge.

Pecl, G. T., Stuart-Smith, J., Walsh, P., Bray, D. J., Kusetic, M., Burgess, M., ... & Moltschaniwskyj, N. (2019). Redmap Australia: challenges and successes with a large-scale citizen science-based approach to ecological monitoring and community engagement on climate change. Frontiers in Marine Science, 6, 349.

Pereira, P. H., Lima, G. V., Pontes, A. V., Côrtes, L. G., Gomes, E., Sampaio, C. L., ... & Seoane, J. C. S. (2022). Unprecedented coral mortality on Southwestern atlantic coral reefs following major thermal stress. Frontiers in Marine Science, 9, 725778.

Petraitis, P. S., & Dudgeon, S. R. (2020). Declines over the last two decades of five intertidal invertebrate species in the western North Atlantic. *Communications Biology*, *3*(1), 591.

Piatt, J. F., Parrish, J. K., Renner, H. M., Schoen, S. K., Jones, T. T., Arimitsu, M. L., ... & Sydeman, W. J. (2020). Extreme mortality and reproductive failure of common murres resulting from the northeast Pacific marine heatwave of 2014-2016. *PloS one*, *15*(1), e0226087.

Schoen, S. K., Arimitsu, M. L., Marsteller, C. E., & Piatt, J. F. (2022). Lingering impacts of the 2014-2016 northeast Pacific marine heatwave on seabird demography in Cook Inlet, Alaska (USA). Marine Ecology Progress Series.

Smith, S. D., & Edgar, R. J. (2014). Documenting the density of subtidal marine debris across multiple marine and coastal habitats. PloS one, 9(4), e94593.

Sydeman, W. J., Schoeman, D. S., Thompson, S. A., Hoover, B. A., García-Reyes, M., Daunt, F., ... & Watanuki, Y. (2021). Hemispheric asymmetry in ocean change and the productivity of ecosystem sentinels. *Science*, *372*(6545), 980-983.

Troisi, G., Barton, S., & Bexton, S. (2016). Impacts of oil spills on seabirds: Unsustainable impacts of non-renewable energy. international journal of hydrogen energy, 41(37), 16549-16555.

Velarde, E., Anderson, D. W., & Ezcurra, E. (2019). Seabird clues to ecosystem health. *Science*, *365*(6449), 116-117.

Williams, J. M., Tasker, M. L., Carter, I. C., & Webb, A. (1995). A method of assessing seabird vulnerability to surface pollutants. *Ibis*, *137*, S147-S152.

Chapter 1. Using beached bird data to assess seabird oiling susceptibility

Citation: Waugh, J. K., Jones, T., & Parrish, J. K. (2022). Using beached bird data to assess seabird oiling susceptibility. *Marine Pollution Bulletin, 176,* 113437. https://doi.org/10.1016/j.marpolbul.2022.113437

Abstract

Oil spills can cause severe impacts on seabirds, the extent of which varies by species. We investigated taxon-specific susceptibility using data from the *Nestucca* and *Tenyo Maru* oil spills in the northeast Pacific together with seasonally and spatially overlapping baseline beached bird abundance data collected over a 17-year time-period. Multivariate analyses revealed patterns of variation between spill and non-spill data, primarily driven by differences in the relative abundance of common murres (*Uria aalge*) and northern fulmars (*Fulmarus glacialis*). In subsequent susceptibility analyses, alcid (*Alcidae* spp.) carcasses were generally overrepresented in spill data, while gulls (*Larus* spp.), tubenoses (*Procellariformes* spp.), and cormorants (*Phalacrocoracidae* spp.) were generally under-represented. We found that the baseline data had high variability, suggesting a need for many years of baseline data. We propose that where appropriate baseline data exists, this method can be employed to investigate the seabirds most vulnerable to oiling.

1. Introduction

Oil spills are one of the most catastrophic forms of marine pollution in nearshore marine ecosystems, as they can disrupt ecosystem processes via severe mortality of many species

(Chilvers and Battley, 2019; Crain et al., 2009; Halpern et al., 2007). In particular, oil spills have well-documented impacts on seabirds (e.g., Burger and Fry, 1993; Munilla et al., 2011; Castège et al., 2004). These impacts range in scale from the individual bird (e.g., disrupted thermoregulation –Hartung, 1967) to entire populations (e.g., depressed breeding success– Barros et al., 2014). While oil spills are likely to have deleterious effects on all marine bird species, some species appear to be more susceptible to oil exposure and mortality than others (King and Sanger, 1979; Clark, 1984; Lorentsen and Anker-Nilssen, 1993; Wiese and Ryan, 2003; Camphuysen, 2007). Various methods have been employed to assess relative species susceptibility to oiling, including the development of indices of taxon-specific oiling vulnerability (Oiling Vulnerability Indices, or OVIs).

The development of OVIs was pioneered by King and Sanger (1979) in a landmark study assessing 176 marine bird species in the northeast Pacific. Subsequent authors have adopted the OVI approach, variously incorporating different numbers of factors, factor weights, and additive versus multiplicative formulae (e.g., Camphuysen, 1989; Wahl et al., 1981; Williams et al., 1995; Chilvers and Battley, 2019). In general, these methods integrate factors related to natural history (e.g., habitat use, migratory pattern), life history and population dynamics (e.g., reproductive success, distribution) and conservation status (e.g., susceptibility to other population threats, conservation status assessed globally). The benefits of an index approach include the ability to create an assessment of likely risk for many species within a specified location (e.g., a large marine ecosystem). The drawbacks include that the indices necessarily reflect the decisions of the authors as to the importance of the factors incorporated (by deciding whether to weight all factors the same or to assign different weights). An independent evaluation

of taxonomic patterns in oiling vulnerability would provide a means to assess the results OVIs deliver and could provide additional insight for resource management and protection/mitigation.

In this paper, we develop an approach complementary to that of OVIs by comparing taxon-specific spill and non-spill carcass deposition data within the northern California Current System. We used seabird carcass data from two spills, the *Nestucca* (Dec 1988, southern outer coast of Washington, US) and the *Tenyo Maru* (July 1991, northern outer coast of Washington, US). We compared observed patterns of taxonomic composition during those spills to baseline beached bird composition data. This baseline data, collected by a long-term citizen science program, was specific to the season and coastline affected for each spill (Figure 1). Although this method cannot account for the behavior of live birds that might attract or repel them from oil at sea, effectively redistributing the local populations in the spill year, it is an otherwise valid comparison, as Wiese (2003) has shown no effect of oiling on carcass float duration.

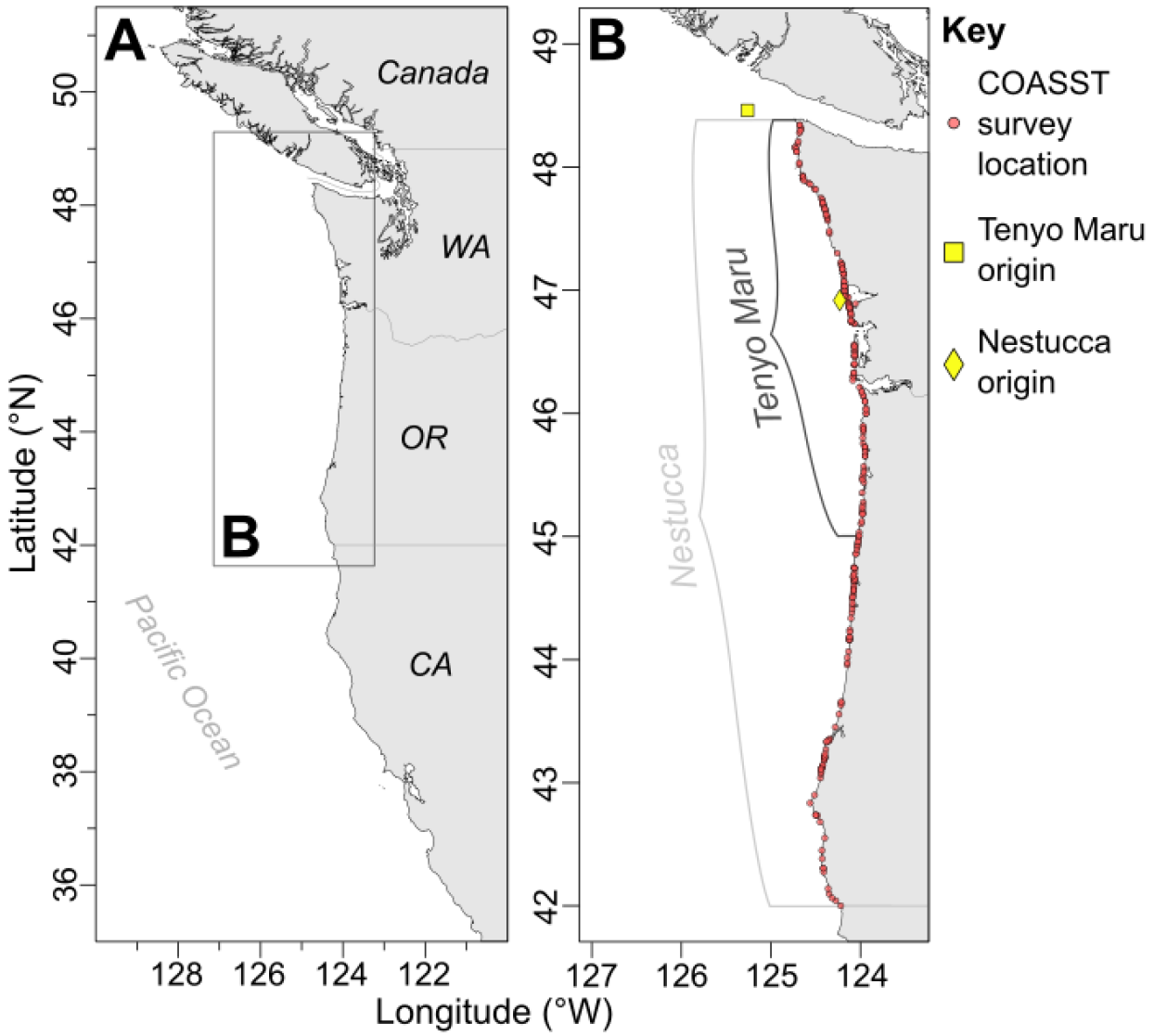


Figure 1. Maps showing (A) the location of the study region within the northeast Pacific and (B) specific locations of spill occurrence and baseline beached bird monitoring sites. The zoomed in map of the study region (B) shows approximate spill points of origin, surveyed beach locations within the COASST program along the outer Washington and Oregon coastlines, and the coastline extent over which COASST surveys were queried to generate baseline estimates of species composition for each spill.

Our goals in this work were firstly to identify whether overall spill taxonomic composition differed from baseline beached bird composition, considering the interannual variability in baseline composition; and secondly to identify which species were relatively

over/underrepresented in the spill dataset as a means of categorizing taxon-specific susceptibility during these two spills. Finally, we compared these relative susceptibilities to the OVI scores for these same taxa presented in King and Sanger (1979) to highlight the similarities/differences of these approaches. We argue that baseline datasets, such as that utilized here, provide additional insight into oiling susceptibility by allowing for a data-informed approach independent from assumptions based on taxonomic traits.

2. Methods

2.1. Spill data

We used spill reports to assemble data on count and lowest possible taxon of oiled beached bird carcasses (hereafter, carcasses) found following the *Nestucca* and *Tenyo Maru* oil spills (Table 1). Following the *Nestucca* spill, carcasses were collected in Washington State coastal areas from the day of the spill (22-Dec-1988) until late Jan-1989 (Ford et al., 1991). Survey efforts were coordinated by the Washington State Conservation Corps (Ford et al., 1991; Burger, 1992). Following the *Tenyo Maru* spill (22-Jul-1991), carcasses were recovered from Vancouver Island south to Lincoln City, Oregon (*Tenyo Maru* Trustee Council & U.S. Fish Wildlife Service 2000). Survey efforts were coordinated by the U.S. Fish and Wildlife Service, the State of Washington, the National Oceanic and Atmospheric Administration, and the Makah Tribe (Summary of the Tenyo Maru oil spill restoration, 2006).

Table 1: For each spill, the date the spill occurred (Date), number of beached bird carcasses in the spill data (Spill Count), latitudinal range which the spill affected and for which we queried the COASST database (Lat. Range), days of the year for which we queried the COASST database (Day-of-year Range), number of beached bird carcasses in the queried subset of the COASST data set (COASST Count), and main data source for each spill (Spill Data Source).

Spill	Spill information		Baseline data (COASST)			Spill Data Source
	Date	Spill Count	Lat. Range (°N)	Day-of-year Range	COASST Count	
Nestucca	22-Dec-88	13,473	42.0–48.4	335–365, 1-14	10,927	Momot (1994)
Tenyo Maru	22-Jul-91	4,300	45.0–48.4	181–225	4,137	Tenyo Maru Trustee Council & U.S. Fish Wildlife Service (2000)

2.2. Baseline data

Data from the Coastal Observation and Seabird Survey Team (COASST) were used to represent taxonomic patterns of baseline abundance, which we define here as taxon-specific carcass sightings by volunteer surveyors during non-spill years, with seasonal and geographic overlap with the spill in question. Established in 1999, COASST is a citizen science program housed at the University of Washington in which expert-trained participants survey a known length of beach (Figure 1) monthly for beached marine birds. Carcasses are scored for body condition, intactness, and the presence of oil or anthropogenic sources of entanglement. All carcasses are identified to the lowest possible taxon with the aid of a key (Hass and Parrish, 2013), photographed, individually marked to distinguish newly found birds from those previously

found, and left on the beach. Identifications are subsequently verified by experts via measurements and photographs. Within the northern California Current large marine ecosystem encompassed by the COASST beached bird monitoring program, recorded chronic oiling rates are extremely low. This area has a median annual chronic oiling rate of 0.1%, ranging from 0.01% to 2.0% among years (2001–2019, n = 656–10,144 birds per year). Within the seasonal and geographic bounds of each spill (i.e., the baseline data used to inform our analyses), oiled birds were only documented in the minority of years (*Nestucca*: 5 of 19 years; *Tenyo Maru*: 2 of 19 years), with recorded annual oiling rates ranging from 0 to 4% (2001–2019, n = 70–5714 birds per year; COASST unpublished data).

While the COASST dataset does not include the years during which either spill occurred, it can provide a post hoc baseline of beached bird community composition. Although both spills had major impacts on local seabird populations in the short-term (*Tenyo Maru*: Thompson et al., 2003, Summary of the Tenyo Maru oil spill restoration, 2006.; *Nestucca*: Burger, 1993, Ford et al., 1991), there is little evidence to suggest that spills by themselves alter seabird populations in the long-term (Clark, 1984). However, factors like exposure to lingering oil and diet can affect recovery time (Esler et al., 2018). Additionally, we found no evidence to suggest that species composition derived from COASST data in the areas affected by either spill changed significantly over decadal timescales (Supplementary Material Table S1), suggesting that overall composition is relatively static.

For each spill analysis, the COASST dataset was restricted according to location and time of year to match the spatial and seasonal extent of the spill (Table 1). To account for the duration of

carcass collection following each spill (*Nestucca*: primary response from 22-Dec to end of Jan-Ford et al., 1991; *Tenyo Maru*: primary response from 22-Jul to 29-Aug National Oceanic and Atmospheric Administration, n.d.) and variation in beaching phenology over the year set encompassed by the baseline dataset, we defined a 45-day seasonal window centered on the day-of-year on which each spill occurred (Table 1). We explored smaller (30-day) and larger (60-day) windows but found that within-analysis proportional composition was largely unchanged (Supplementary Material Figure S1). For each spill analysis we defined the "spill-region" of the baseline dataset as all COASST sites within the northern to southern extent of U.S. shorelines reported as oiled and from which oiled carcasses were recovered (Table 1). However, we did not incorporate the coastline of Vancouver Island, Canada into the analysis, despite oil and oiled carcasses arriving there, because there is no contemporaneous baseline dataset that matches the COASST data in spatial and temporal grain. Because the spatial coverage of COASST expanded into Oregon over time, we used data from 2002 (after which COASST spatial coverage changed by less than 0.1° latitude within the spill-regions) through 2018 (17 years) for both spill analyses.

2.3. Species and group-level data processing

Because there was little-to-no data on survey effort (e.g., number of searchers, beach length or area searched) for either spill, we were unable to compare the rate at which carcasses were found (i.e., carcasses per km) between spill and baseline data. Therefore, we focused on relative taxonomic composition, reasoning that the taxa that are relatively more prevalent in spill body count as opposed to in the baseline dataset are relatively more susceptible to oiling.

We conducted our analyses at two levels of taxonomic resolution: species level, and a variably higher taxonomic level we refer to as the "group level" (Table 2). Species-level analyses allowed us to analyze data at the finest resolution possible but did not include the high proportion of carcasses not identified to species (e.g., in the *Nestucca* dataset ~9.1% of the total carcass count). Group-level analyses were based on semi-aggregated counts, which allowed for maximal inclusion of carcass counts (e.g., incorporating carcasses unknown to species but known to a higher taxonomic level) not presented at the species level. We ran the *Nestucca* data only at the group level, and the *Tenyo Maru* data at both the species and group levels.

Table 2. Species documented in either spill event analysis and relevant taxonomic groupings. Each group was included in both spill analyses unless otherwise indicated (TM = *Tenyo Maru*, N = *Nestucca*). For the *Tenyo Maru* species-level analysis, only species in bold were included; all remaining species did not meet thresholds for analysis inclusion based on rarity (threshold spill data: ε 1 individual; baseline data: ε 3 individuals in two or more years).

Family	Groups	Species
Alcidae	Auklets	**Cassin's auklet (***Ptychoramphus aleuticus*), Parakeet auklet (*Aethia psittacula,* N only)
	Guillemots (TM only)	**Pigeon guillemot (***Cepphus columba*)
	Murrelets	Ancient murrelet (*Synthliboramphus antiquus,* N only), Marbled murrelet (*Brachyramphus marmoratus*)
	Murres	**Common murre (***Uria aalge*)
	Puffins	Horned puffin (*Fratercula corniculata*), **Rhinoceros auklet** (*Cerorhinca monocerata*), Tufted puffin (*Fratercula cirrhata*)
Anatidae	Scoters	Black scoter (*Melanitta americana,* N only), Surf scoter (*Melanitta perspicillata*), **White-winged scoter** (*Melanitta deglandi*)
Diomedeidae	Albatrosses (TM only)	**Black-footed albatross** (*Phoebastria nigripes*)
Gaviidae	Loons	Common loon (*Gavia immer*), Pacific loon (*Gavia pacifica,* N only), Red-throated loon (*Gavia stellata*)
Hydrobatidae	Storm petrels (TM only)	**Fork-tailed storm petrel (***Oceanodroma furcata)*
Laridae	Gulls	American herring gull (*Larus smithsonianus,* N only), California gull (*Larus californicus*, TM only), Kittiwakes (*Rissa* spp., N only), Mew gull (*Larus canus*), Ring-billed gull (*Larus delawarensis*, N only), Western/Glaucous-winged gull (*Larus occidentalis/ Larus glaucescens*)
	Terns (TM only)	**Caspian tern** (*Hydroprogne caspia*)
Phalacrocoracidae	Cormorants	**Brandt's cormorant** (*Phalacrocorax penicillatus*), **Double-crested cormorant** (*Phalacrocorax auritus,* TM only), **Pelagic cormorant** (*Phalacrocorax pelagicus,* TM only)
Podicipedidae	Large Grebes (N only)	Western grebe (*Aechmophorus occidentalis*)
	Small Grebes (N only)	Horned grebe (*Podiceps auritus*), Red-necked grebe (*Podiceps grisegena*)
Procellariidae	Fulmars	**Northern fulmar** (*Fulmarus glacialis*)
	Shearwaters	Short-tailed shearwater (*Ardenna tenuirostris,* TM only), **Sooty shearwater** (*Ardenna grisea,* TM only)

In addition, raw counts from both spill and baseline datasets were processed to account for differences in identification and classification as follows. Several taxonomic categories were excluded a priori due to their ecology and/or scarcity: terrestrial birds, shorebirds, and waterfowl

(except for scoters, *Melanitta* spp.); as well as all completely unidentifiable birds. Across both spill datasets, many gulls (*Larus* spp.) were not identified to species (*Nestucca*: 98.6% of all gulls; *Tenyo Maru*: 17.7%). Furthermore, within COASST, participants group all large (wing chord >33 cm) immature gulls into a single super-taxon (large immature gulls, Hass and Parrish, 2013), which account for 35% (*Nestucca* spill-region) to 59% (*Tenyo Maru* spill-region) of all reported gulls (*Larus* spp.) in the baseline datasets. Therefore, counts of gulls were excluded from species-level analyses, but were combined (including kittiwakes, *Rissa* spp.) into a single group "gull" for group-level analyses.

Once these preliminary exclusions and groupings were made, 0.3% of the *Tenyo Maru* carcasses (14 of 4022) remained unidentified to species (and 2.4% of carcasses in the corresponding COASST baseline dataset), whereas 16% of the *Nestucca* carcasses (1911 of 11,913) remained unidentified to species (and 14.8% of carcasses in the corresponding COASST data). To maximize the inclusion of carcasses from the spill dataset in the analyses, and especially in cases where many carcasses were identified to a taxonomic level above that of our analysis (species for *Tenyo Maru*, group for *Nestucca*), we redistributed unknowns at a level one above the analysis to the next lower level according to the proportion of known carcasses within the dataset (see Supplementary Materials section 3). Redistribution had little effect on overall taxonomic composition (Supplementary Materials section 3), but became important in select taxa (e.g., grebes, *Aechmophorus* and *Podiceps* spp.) where the majority of spill dataset carcasses were categorized at family or sub-family levels. In these cases, redistribution allowed incorporation of the majority of the carcasses, ensuring the relative taxonomic proportions did not under-represent

susceptibility to oiling (i.e., by effectively erasing the carcasses from the total count). All subsequent analyses were based on redistributed counts.

Taxonomic composition was initially examined by plotting proportional abundance (% count/total) by taxa (species for *Tenyo Maru*, group for *Nestucca*) for each spill analysis. To represent average baseline composition corresponding to each spill, we calculated the average proportion for each taxon present in the baseline data by averaging year-specific proportions across baseline data. In subsequent analyses, taxa with low/sporadic counts were excluded to avoid basing conclusions on the presence of only a small number of individual birds in either dataset (spill, baseline). To be included, taxa had to meet both of the following thresholds: spill: ≥ 1 individual; baseline: ≥ 3 individuals in two or more years. For the *Tenyo Maru* spill, species not meeting these thresholds (10 species, 194 carcasses, 4.5% of all carcasses; Supplementary Material Table S6) were omitted from species-level analyses but included in the group-level analyses, excepting large grebes (which included only western grebe, *Aechmophorus occidentalis*) which were still too rare even at the group level. Within the *Nestucca* dataset, guillemots (which included only pigeon guillemot, *Cepphus columba*) were the only group excluded based on rarity. After processing, counts were converted to a proportion of the total for each oil spill and each baseline year, resulting in a dataset of proportional species composition.

2.4. Statistical analysis

2.4.1. Ordination and clustering of spill and baseline datasets

We explored annual variation in taxonomic composition among baseline data and the placement of the spill dataset year among the baseline years via a series of multivariate analyses (including ordination and cluster analyses). This enabled us to identify the degree to which the spill dataset differed from its baseline and to identify taxa driving any differences. To determine the similarity between spill and non-spill years, we calculated distance matrices using Jaccard's distance (Jaccard, 1901) in the vegan package in R (Oksanen et al., 2015). Jaccard's distance was used as it is less affected than other distance metrics (e.g., symmetrical distance metrics like simple matching) by "double absences" (i.e., zero counts for a taxon in multiple years; in our analyses, ~12.0%- *Nestucca* group level- to 18.5%- *Tenyo Maru* species level- of taxa/year combinations had a value of zero, leading to double absences) which may otherwise have skewed measures of similarity among years (Legendre and Legendre, 2013).

We applied Principal Coordinates Analyses (PCoAs) to the spill-specific multivariate distance matrices to visualize relative differences, and potential clustering, in species composition among all years (i.e., a single spill year and multiple baseline years). We explored the number of dimensions needed to adequately represent patterns of taxonomic composition using the percent variation explained for each individual axis and the overall percent variation explained for a given n-dimensional representation. Taxa driving patterns in variability among years were identified by calculating the coefficient of determination, R^2, between an individual taxon's proportional abundance among years (response) and the axis scores (predictors) resulting from the PCoA. A permutation test was then applied to determine the likelihood that the observed variation explained (R^2) by the axis scores could have occurred by chance (i.e., if there was no

relationship with that taxon's proportional abundance). Based on the results of the permutation test we plotted vectors for taxa whose correlation with axis scores was significantly ($\alpha = 0.05$) higher than expected, as an indication of taxa responsible for major modes of taxonomic variation.

We further explored whether years displayed distinct clustering using Ward's Minimum Variance method (Ward, 1963), which aims to group similar points via minimization of the sum of within-cluster distances for a desired number of clusters. We used silhouette width (Rousseeuw, 1987) to determine how many clusters were present in the data. Silhouette width measures how well an observation belongs to a cluster based on distances to observations contained within the cluster to which it is assigned, relative to the next nearest alternative cluster; where widths close to one are well clustered, and widths approaching zero are indicative of points that lie between two or more clusters (Menardi, 2011). We evaluated mean silhouette width for $k = 2–10$ clusters, and the optimum number of clusters was chosen as the value of k maximizing mean silhouette width. For the optimum number of clusters, we also examined the silhouette width for each individual point to identify how well it belonged to its assigned cluster. Cluster analyses were visualized via dendrograms to illustrate grouping structure and to identify whether the spill data, or any baseline datum, would be considered an outlier relative to remaining data for consideration in subsequent analyses. All multivariate analyses were performed in R version 4.0.4 (R Core Team, 2021).

2.4.2. Susceptibility analysis

We compared the proportion of each taxon observed during the spill against proportions observed across the baseline data. Baseline proportions among years were processed via bootstrap resampling to calculate the mean and 95% confidence interval of the mean proportion for each taxon. We also calculated whether proportional abundances in each baseline year were above or below that recorded during the oil spill for each species.

To explore patterns of taxonomic over−/under-representation, we calculated an index of *relative representation* as the ratio of proportional abundance recorded during the oil spill, P_{spill}, to the bootstrapped baseline mean proportion, $\underline{P_{base}}$.

$$Relative\ representation = \frac{P_{spill}}{\underline{P_{base}}}$$

For this index, a value above (below) one indicates that that taxon was over- (under-) represented in the oil spill relative to the baseline. We took the log of this ratio such that equivalent proportions (i.e., $\frac{P_{spill}}{\underline{P_{base}}} = 1$) are close to zero, and over−/under-represented species take positive/negative values, respectively. This also has the advantage that resulting values are symmetric according to the degree of over/under-representation (i.e., oil spill proportions double or half that of the baseline are equally distant, ± 0.693, from the equivalence value of zero when the log-ratio is used).

To examine relative variability in proportional species composition among baseline years, we also identified the number of years for which baseline proportional abundance was exceeded by

proportional abundance observed during the corresponding spill. This was then converted to the proportion of years by dividing by the total number of baseline years (n = 17). We treat the resulting index, $p(P_{spill} > P_{base})$, as a measure of *reliability* given interannual variability in the baseline data, with values close to 0 and 1 indicative of high confidence that a species was under/over-represented in the spill, respectively. Values close to 0.5 convey low *reliability* as corresponding taxa were just as likely to be under-represented as over-represented when examining the spill proportion against any individual baseline year.

Finally, we performed two types of sensitivity analyses to quantify the degree to which our conclusions may have been affected by extremes in our datasets. To account for potential effects of unusual years, we ran all analyses with and without baseline years that had been identified as outliers following multivariate cluster analyses. Second, because proportions are not independent among taxa, there is a risk that susceptibility analyses could have been unduly influenced by dominant taxa (i.e., over-representation by a particularly abundant species may mask signals for less abundant species). To account for this, we re-ran susceptibility analyses in stages, sequentially removing the taxon with the highest perpendicular Euclidean distance from equivalence (i.e., the theoretical point where oil spill and baseline proportions are equal) as this data point is most likely to skew results among remaining taxa. We tracked changes in our indices of *relative representation* and *reliability* through sequential taxon removals and identified whether changes would affect our overall conclusions regarding taxon susceptibility. All species susceptibility analyses were performed in Python 3.9.0 (Python Core Team, 2021).

3. Results

3.1. Community composition during spills

Community composition for spill and corresponding baseline datasets can be described as a Pareto curve (Wood and Wood, 2005), wherein a few taxa overwhelmingly dominated the pattern, with a steep fall off to a long tail of less abundant taxa (Figure 2). However, the dominant species were not entirely consistent between spill analyses, or when comparing spill and relevant baseline data within an analysis. In both spills, common murres (*Uria aalge*) accounted for more than 50% of all carcasses (Figure 2). Spill-specific differences in taxonomic representation were also evident, with puffins (*Fratercula* and *Cerorhinca* spp.) and murrelets (*Synthliboramphus* and *Brachyramphus* spp.) being relatively prevalent in the *Tenyo Maru* spill, but almost completely absent in the *Nestucca* spill; and the opposite being true of grebes and scoters (Figure 2). Demonstrable differences between spill and the relevant baseline data composite included procellariids (*Procellariidae* spp.), cormorants (*Phalacrocoracidae* spp.), and gulls, which were relatively abundant in baseline datasets, and relatively rare in both spills (Figure 2).

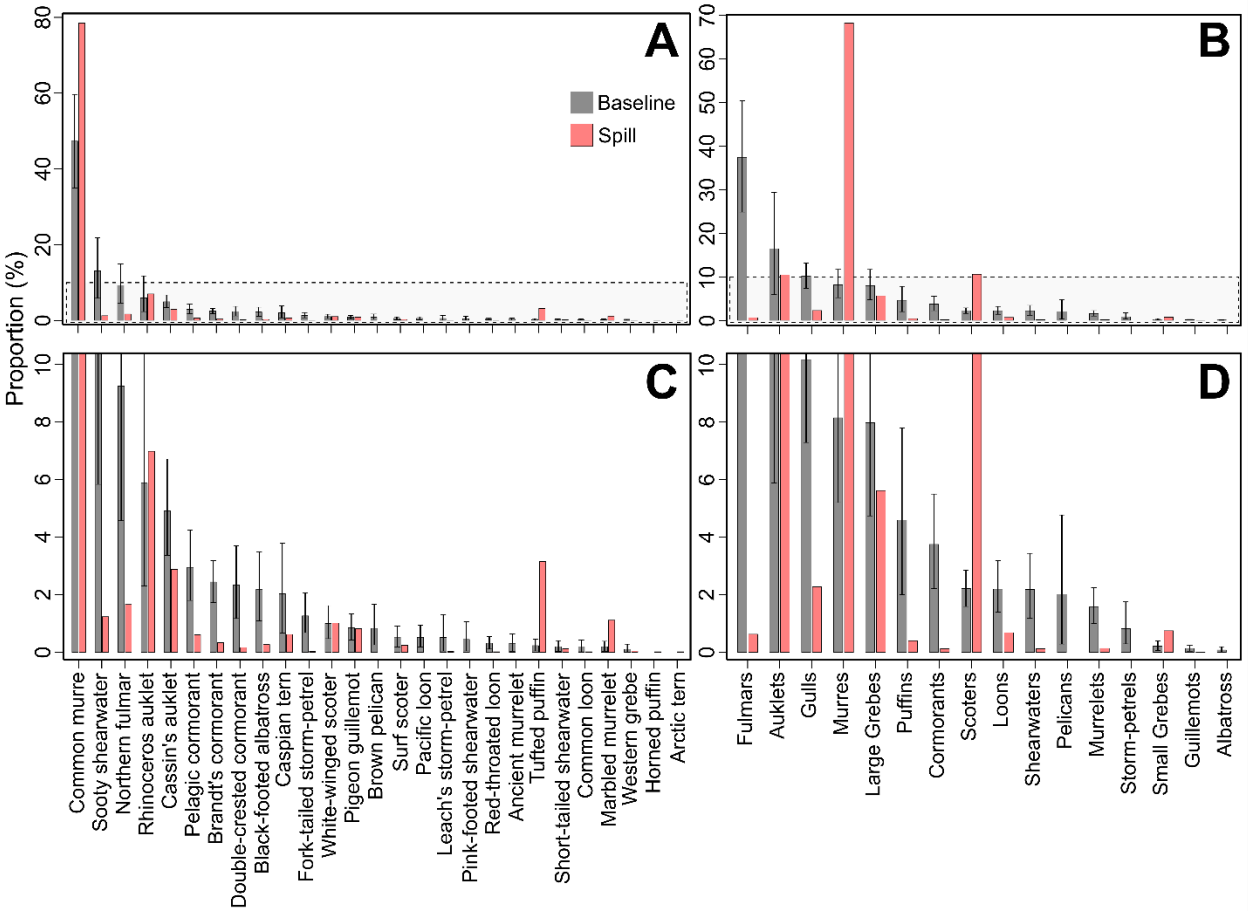


Figure 2. Taxonomic composition of baseline and spill beached bird datasets in decreasing rank order according to baseline taxon abundance. Plotted are the proportional composition (count/total expressed as a %) by species for the *Tenyo Maru* spill (A) and by group for the *Nestucca* spill (B). For each spill, the corresponding baseline proportions representing the average proportion calculated across available baseline years are plotted, along with error bars representing the 95% confidence intervals of bootstrapped mean proportion among years. Lower panels represent a detailed view, truncated at 10% for the *Tenyo Maru* (C) and *Nestucca* (D) taxonomic compositions to show rarer taxa.

3.2. Ordination and clustering of spill and baseline datasets

To examine the interannual variability of spill-specific baseline communities and to locate the spill dataset relative to the baseline years, we used PCoAs based on proportional taxonomic composition. Over half (55–60%) of the variance in annual community composition was explained by the first two axes (Figure 3), with subsequent axes resulting in minimal additional

variance explained (3rd axis: +10–13%, 4th axis: +7.5–8.5%; Supplementary Material Section 5). For both spill analyses, most years were situated along a single cross-axis gradient described by opposing dominance ($\alpha = 0.05$) of common murres (*Tenyo Maru*) and fulmars (including only the species northern fulmar—*Fulmarus glacialis*, in both *Tenyo Maru* and *Nestucca*; Figure 3; Supplementary Material Table S7). Annual exceptions to the dominant gradient were apparent in both analyses and consisted of a loose grouping of mostly consecutive years (*Tenyo Maru*: 2011–2013, *Nestucca* 2014–15 and 2018) creating a secondary dimension, and variably characterized by shearwaters (*Ardenna* spp.; *Tenyo Maru*) or auklets (*Ptychoramphus* and *Aethia* spp.) *Nestucca*; Figure 3). In both analyses, taxonomic composition of the spill data fell within the extremities of the primary gradient, aligning with the dominant baseline species: towards common murres for the *Tenyo Maru* spill, and against northern fulmars for both spills (Figure 3).

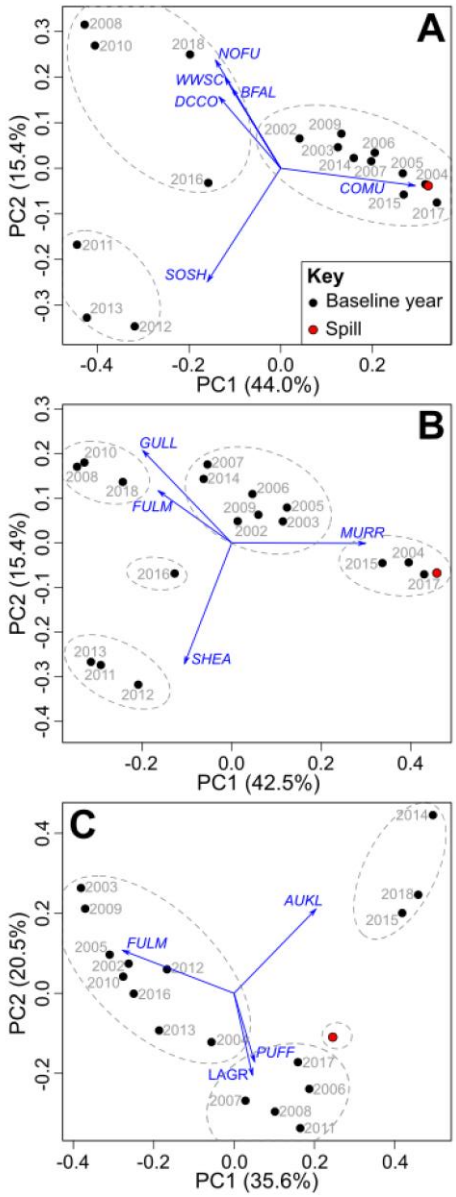


Figure 3. Principal Coordinates Analyses (PCoA) plots showing ordinations of spill and baseline taxonomic composition data for the *Tenyo Maru* spill at the species (A) and group (B) levels, and for the *Nestucca* spill at the group level (C). Percent variation explained along each axis is given in parentheses. Clusters of similar years identified via hierarchical clustering are indicated by dashed ovals. Vectors showing taxonomic loadings are illustrated for those taxa for which variability was significantly ($\alpha = 0.05$) explained by principal axis scores assessed via permutation test. COMU/MURR: Common murre, BFAL: Black-footed albatross, NOFU/FULM: Northern fulmar, WWSC: White-winged scoter, DCCO: Double-crested cormorant, SOSH: Sooty shearwater, GULL: gulls, SHEA: Shearwaters, AUKL: Auklets, PUFF: Puffins, LAGR: Large Grebes.

We applied hierarchical clustering to the multivariate ordination firstly to reveal clusters indicative of 'modes' in taxonomic composition; and secondarily to examine whether the spill year could be considered as an expression of one of those modes or characterized as an outlier. Clustering structure was best described with 3 (*Tenyo Maru* -species), 5 (*Tenyo Maru* – group) or 4 (*Nestucca*- group) clusters determined based on average silhouette widths (although clustering strength was moderate: 0.32–0.39; see Supplementary Material Figure S3) and visualized in Figure 3 as circled year sets. Although clustering easily identified off-gradient years, including outliers, it also partitioned the dominant gradient into 2–3 clusters (Figure 3). The *Nestucca* spill year was characterized as an outlier, whereas the *Tenyo Maru* spill year was strongly grouped with other baseline years (silhouette width 0.5–0.6, Supplementary Material Figure S3 and Table S8). Finally, despite the addition of taxa (e.g., gulls) at the group level *Tenyo Maru* analysis, the pattern of ordination remained largely unchanged (Figure 3 A vs B). Gulls were strongly aligned with fulmars as a driving group and identified by a single cluster, whereas years dominated by murres (which included only common murre) formed an opposing and distinctive cluster, and both were separated from intervening years.

3.3. Susceptibility analysis

We examined taxonomic oiling susceptibility using a 2-axis data visualization (Figure 4). The abscissa shows *relative representation*, or our index of the degree to which taxon-specific proportional abundance in the spill year exceeded (or not) that of the baseline dataset, while the ordinate represents our index of *reliability* in that conclusion. These graphs describe a continuum which can be simply parsed into three regions. The upper right corner includes taxa which

proportionally occur order(s) of magnitude more in the spill relative to the baseline and exceed the baseline in a high proportion of years (e.g., high representation and *reliability*). The lower left corner includes taxa that occur order(s) of magnitude more in the baseline relative to the spill, and whose spill proportion falls below the baseline in a high proportion of years (e.g., low representation and high *reliability*). In the middle are taxa for which patterns of susceptibility are variable based on inclusion/exclusion of dominant taxa in the dataset and possibly length of the baseline dataset.

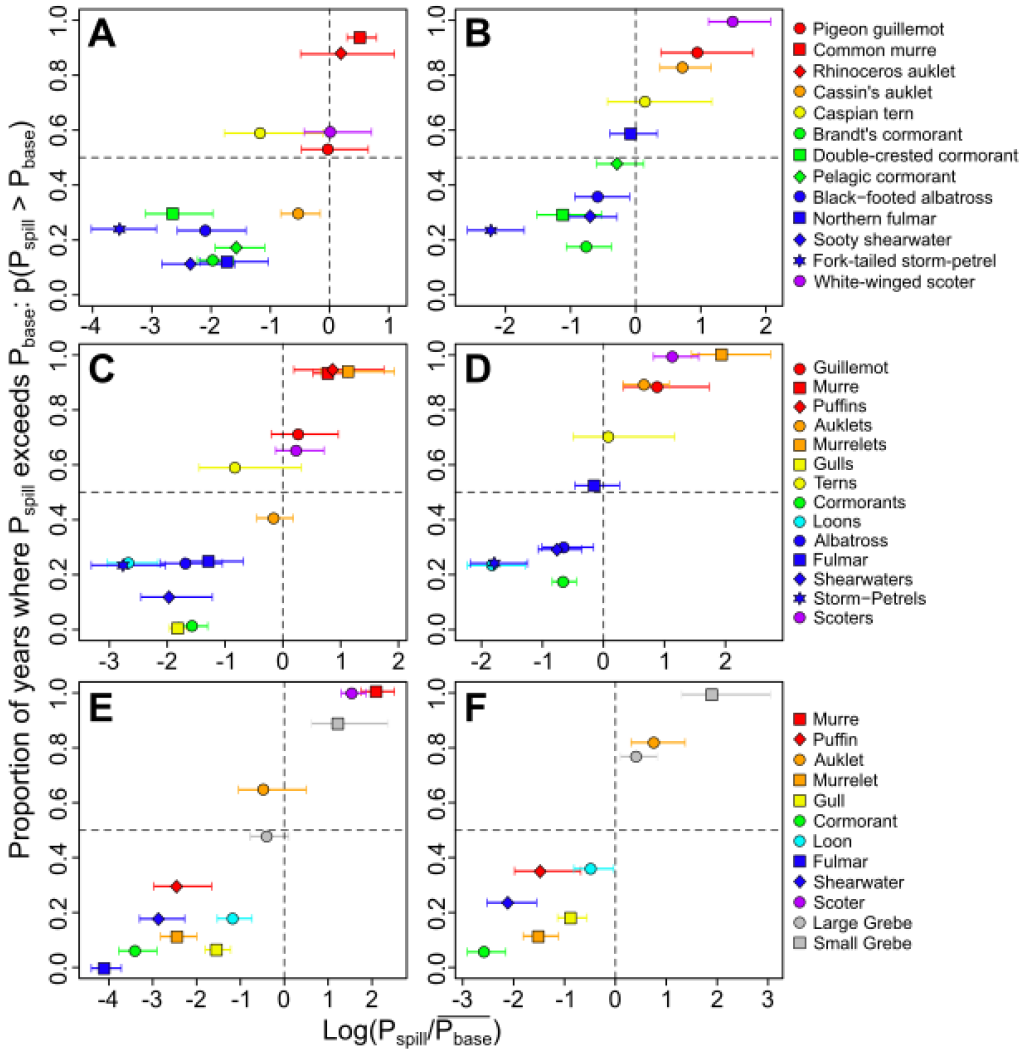


Figure 4. Bi-plots of relative representation ($log\left(\frac{P_{spill}}{P_{base}}\right)$) and reliability ($p(P_{spill} > P_{base})$) metrics for the *Tenyo Maru* spill at the species (A & B) and group levels (C & D) and for the *Nestucca* spill at the group level (E & F). Alternate panels for each spill show metrics calculated before (A, C, E) and after (B, D, F) removal of potentially influential taxa whose influence may have masked patterns in taxonomic space. Taxa removed were: *Tenyo Maru* at the species level: Common murres and Rhinoceros auklets (A – B); *Tenyo Maru* at the group level: Murres, Puffins, and Gulls (C – D), *Nestucca* at the group level: Murre, Fulmars, and Scoters (E – F).

Only a small number of taxa were over-represented on average in spill datasets relative to their baselines, namely common murres, puffins, and murrelets for the *Tenyo Maru* spill analysis, with

the latter two only becoming apparent at the group level; and murres, scoters and small grebes (*Podiceps* spp.) for the *Nestucca* spill analysis (Figure 4). In contrast, across both spills analyses, procellariids, cormorants, loons (*Gavia* spp.), and gulls were under-represented in spill datasets relative to average baseline proportions (Figure 4). Taxa of indeterminate status (i.e., neither over- nor under-represented) were taxonomically inconsistent between spill analyses.

Our *reliability* index indicated that most taxa had some chance of the spill year proportion exceeding at least one baseline year (Figure 4). Notable exceptions were gulls and cormorants for the *Tenyo Maru* group-level analyses and fulmars for the *Nestucca* analyses. Overall, this suggests that for the most over-represented taxa during each spill, years with higher proportional abundance in the baseline may only occur once per decade.

The apparent difference in *reliability* for cormorants in the *Tenyo Maru* spill analysis between species (*reliability* = 0.1–0.3; Figure 4A) and group-level (*reliability* = 0; Figure 4C) analyses arose due to species-specific absences in alternating baseline years (pelagic cormorant *Phalacrocorax pelagicus*: 2 years; Brandt's cormorant- *Phalacrocorax penicillatus*: 2 years; double-crested cormorant- *Phalacrocorax auritus*: 5 years), resulting in some years where the spill exceeded the baseline. When aggregated across the species group, all baseline years became nonzero and always exceeded that observed during the spill.

When numerically dominant taxa were sequentially removed, *relative representation* changed considerably for the *Tenyo Maru* analyses (Figure 4) with five species (~38.46%) and three groups (~21.43%) changing signs from relatively susceptible to relatively immune, or vice versa

(Figure 4; see Supplementary Material Figs. S5–S8 for full sensitivity results). *Nestucca* analyses were also influenced by the inclusion of numerically abundant taxa, but to a lesser extent, with three groups (loons, large grebes, and auklets) changing sign upon taxon removal (Figure 4, and Supplementary Material Figs. S5–S8). Taken together, this suggests that the occurrence of particularly dominant taxa may mask signals presented by less abundant species when using analyses based on proportions.

4. Discussion

4.1. Main findings

Our analyses demonstrate the value of long-term beached bird monitoring programs for taxon-specific impact assessment of oil spills. Although the composition of beached birds following two major spills along the outer coast of Washington State arguably fell within the space described by the baseline dataset (Figure 3), there was selective over/under-representation among particular taxa suggesting differential susceptibility to oiling (Figure 4). Such spill-mediated shifts in taxonomic composition have been found in other taxa (e.g., in phytoplankton, Parsons et al., 2015; and bacteria, Engel et al., 2017).

Species that tend to have high carcass counts in spills are often singled out as being highly vulnerable (e.g., murres in the *Exxon Valdez* spill - Piatt et al., 1990). By contrast, our approach focuses on taxa that are highly over- and under-represented in spills regardless of raw carcass count. This brings out relatively rare taxa that appear to have a higher susceptibility to oiling, (e.g., marbled murrelets, *Brachyramphus marmoratus*, and tufted puffins, *Fratercula cirrhata* in

our analyses, Figure 2, Figure 4), which might otherwise be overlooked because of the predominance of majority taxa.

4.2. Spills occurring in the same location but different seasons

Not surprisingly, there was a classic Pareto curve (Wood and Wood, 2005) evident in both the baseline and spill datasets (Figure 2). Although the *Tenyo Maru* spill originated further offshore than the *Nestucca* spill, oil from both spills quickly spread along the entire outer coastline of Washington State, reaching variably into Oregon (Figure 1). While differences in the species diversity of the spill datasets may have been influenced by the offshore-onshore distribution of the spill origin, inter-spill differences in the associated baseline dataset patterns are likely indicative of seasonally driven relative abundance (*Tenyo Maru* - summer; *Nestucca* - winter), and/or seasonally specific mortality of specific species, as a function of life history and natural history. For instance, locally breeding common murres predominated in the summer baseline datasets and experience a beaching peak following the end of the breeding season (Parrish et al., 2007), whereas seasonal immigrant northern fulmars predominated during the winter (Figure 2) following dispersal south from the Gulf of Alaska (Hatch et al., 2010). However, within the spill datasets, murres were overwhelmingly dominant, upwards of 70% of the total (Figure 2), suggesting that - at least for this species - oil susceptibility may not depend on season.

4.3. Comparison with oil vulnerability indices

To address the paucity of spatio-temporally relevant independent data needed to describe relative susceptibility directly following major spills in the northeast Pacific, King and Sanger (1979)

developed a trait-based approach to assessing the relative vulnerability of marine-associated birds known as the Oil Vulnerability Index, or OVI. Their approach used a 1–5 (low-to-high) scale to rank the importance of each of 20 factors variously describing status and trends in species' abundance and distribution in addition to a range of natural history and life history traits.

The OVI approach has since been adapted across a range of subsequent oil vulnerability analyses (e.g., Camphuysen, 1998; Wahl et al., 1981; Williams et al., 1995; Wiese and Ryan, 2003; Chilvers and Battley, 2019) as well as vulnerability to other anthropogenic mortality sources (e.g., wind farms- Garthe and Hüppop, 2004). Factor numbers have varied, as has factor weighting (e.g., Anker-Nilssen, 1987; Williams et al., 1995). Additional factor types have typically included susceptibility to anthropogenic mortality sources, including oil; and conservation status assessments (e.g., Webb et al., 2016).

The utility of King and Sanger's (1979) work is the ability to create a single set of OVI scores for a wide range of species, which can then be applied to any number of spill, or spill scenario, situations throughout the northeast Pacific (e.g., Wahl et al., 1981), and more broadly (e.g., the North Sea - Camphuysen, 1989; New Zealand - Chilvers and Battley, 2019). The drawbacks were, and are, variable data quality across all species and necessary reliance on expert opinion as a raw data stand-in.

King and Sanger (1979) favored individual traits (9) over population (3) factors, and also included five factors focused specifically on exposure to oil. In general, mortality factors other than oiling (3) and factors related to population size were less influential in generating their

higher OVI scores, both because of the skew in factor number, and because these latter factors tended to have lower scores (e.g., 1–3) assigned to them. Furthermore, because individual traits and oiling exposure scores tended towards identical within family, OVI rankings produced phylogenetic clustering.

Our data suggest that there is an effect of phylogeny as more closely related taxa tended to be clustered together (most apparent in the *Tenyo Maru* species level analysis; Figure 4A). One exception are the alcids (*Alcidae* spp.), which ranged across our susceptibility categories, especially in the winter spill (Figure 4E). Despite this singular difference, we suggest an interpretation of this general similarity is that shared traits expressed in the ecology and behavior of related species do matter in placing individual birds in harm's way and may trump population and broader-scale conservation factors, at least as assessed at the local (immediate to spill) level.

However, our results also suggest that there are differences between the OVI for northeast Pacific taxa (i.e., King and Sanger, 1979) and a spill-specific baseline data-driven approach, both in the relative placement of taxa (i.e., the rank order of species or higher taxonomic groupings), and in seasonal shifts in apparent susceptibility. Alcids, scoters, and grebes emerged as the most susceptible taxonomic groups in our analysis (Figure 4). King and Sanger (1979) also ranked alcids (scores of 70+ out of 100) and scoters (score of 72) as highly vulnerable, although this was not particularly the case with grebes (scores of 44–56). In general, alcids tend to score as highly vulnerable in OVIs, regardless of geolocation (Wahl et al., 1981; Camphuysen, 1989; Camphuysen, 1998; Williams et al., 1995; Webb et al., 2016; Webb et al., 1995; Tasker et al., 1990; Wiese and Ryan, 2003). Scoters are also highly susceptible (Camphuysen, 1989;

Camphuysen, 1998; Williams et al., 1995; Webb et al., 2016; Webb et al., 1995; Tasker et al., 1990). By contrast, grebes tend to be relatively unsusceptible (but see Williams et al., 1995, Webb et al., 1995).

At the individual species level, differential results between the approaches are more apparent. At the scale of the northeast Pacific, King and Sanger (1979) ranked marbled murrelets as the most vulnerable alcid, and common murres as the least. Wahl et al. (1981), working in an adjacent geography - the Puget Sound within Washington State - scored rhinoceros auklets (*Cerorhinca monocerata*), pigeon guillemots, and common murres as most vulnerable. We found that murres were consistently the single most susceptible species (see also Wiese and Ryan, 2003), whereas murrelets were among the most variable taxa in terms of winter (low susceptibility) to summer (extremely high susceptibility) shifts.

At the other end of the scale, the procellariids (fulmars, shearwaters, storm-petrels - which included only fork-tailed storm-petrels, *Oceanodroma furcata*, and albatrosses - which included only black-footed albatrosses, *Phoebastria nigripes*) were consistently among the least susceptible groups in our analysis (Figure 4). Although King and Sanger (1979) did rank shearwaters and albatrosses as relatively less vulnerable, with scores well below their vulnerability threshold of 61, they ranked storm-petrels as quite vulnerable (scores of 63–67). In general, procellariids are considered less vulnerable to oiling than many other groups, albeit with taxon- and location-specific exceptions (e.g., fulmars: Camphuysen, 1989; shearwaters: Williams et al., 1995, Tasker et al., 1990).

Other major groups were variably similar in ranking between our analysis and King and Sanger (1979), with differences often emerging as a function of spill, and presumably season. That is, our results indicate that some groups may experience differential susceptibility: cormorants shifting from medium to low susceptibility from summer to winter; and loons displaying the opposite pattern (Figure 4).

Fox et al. (2016), examining risk of oiling of 12 taxon groups in an area of the northern British Columbia, Canada coastline found grebes and pigeon guillemots to be relatively at risk, but also listed large and small gulls, and cormorants as having a relatively high oiling risk, whereas our work indicated these taxa as relatively safe (Figure 4). We suggest these differences arise not only from methodological differences (mapping overlays versus beached carcass counts) but also from the fact that their highest risk of chronic oiling was consistently located immediately adjacent to the shoreline, effectively placing all nearshore species at higher risk than those with more shelf-wide distributions. Adapting this approach to spill tracks may create more coincident results.

In sum, we suggest that both the mapping and OVI approaches can provide excellent predictions of relative risk, whereas our post-hoc beached bird method allows for a more comparative, control-treatment, approach. Because long-term beached bird baseline data is lacking for many spill locations (e.g., Jones et al., 1970; Paine et al., 1996), and at-sea marine bird data are variably available, an OVI approach may be more globally useful.

4.4. Caveats of this method

We found that the baseline dataset is a dynamic assemblage of species, where taxonomic composition varies year over year, and occasionally across stanzas of years (e.g., *Tenyo Maru* 2011–2013, Figure 3). While spill years are found within the margins, or extremities, of the entire baseline set, they are not necessarily outliers. Although this may indicate that the taxonomic distribution of mortality following a spill is not necessarily that different from at least a portion of the baseline, another interpretation is that other forcing factors are provoking significant shifts in the mortality distribution across the species present in the system, and that those factors are variably evident across years. Significant mortality factors in the northeast Pacific within the past twenty years that have differentially affected marine bird species with concomitant shifts in the frequency distribution of beach-cast carcasses include harmful algal blooms (Jones et al., 2017) and the marine heatwave (Jones et al., 2018; Piatt et al., 2020). The northeast Pacific has undergone regime shifts (Overland et al., 2008) and different phases of El Niño Southern Oscillation (Wang et al., 2019) between the *Nestucca* spill and today, with the - untested - potential to permanently shift the taxonomic distribution of seabirds.

Our baseline data were collected after the spills had occurred, eliminating the possibility of a simple before-after design (e.g., McDonald et al., 2000; Roth and Baltz, 2009). To address this potential source of bias, we compared taxa composition over the first half of the COASST baseline data (2002−2010) with the taxa composition over the second half (2011–2018; Supplementary Material Table S1). We found no significant difference between the two time periods, suggesting that species composition and baseline mortality factors in the spill-regions

did not change substantially. However, we cannot confirm that individual taxa did not change significantly due to factors other than the spill within the study interval.

Where baseline data are available, either the spill or baseline data may be lacking effort control (e.g., number of searchers, area searched, person-hours spent searching), as was the case for the spill data used here. If effort control metadata are missing, our approach of using proportion data may be appropriate, with full acknowledgment that taxa proportions are interdependent (Douma and Weedon, 2019). Dominant taxa can "swamp" the analysis, making it difficult to assess patterns in less prevalent taxa. We addressed this by sequentially removing the most dominant taxa and observing how this affected the susceptibility of less prevalent taxa (Figure 4, Supplementary Material Figs. S5–S8).

4.5. Conclusion

In summary, we propose the use of rigorous beached bird data as a valid baseline against which to assess the impacts of anthropogenic forcing on marine bird populations, including oil spills; and argue that baseline datasets created a posteriori to an event in question can be appropriate. However, we also caution that the creation of a credible baseline dataset can take years to decades, and may require dozens of data collection sites, to encompass the variability inherent in the system. Nevertheless, we advocate for more long-term, rigorous beached bird datasets collected and maintained for the purpose of determining the impacts of anthropogenic forcing on marine systems. Where such baseline data are available, our method can serve as a valuable tool

alongside OVIs to provide a spatial and seasonal view of which seabirds are most vulnerable to spills.

Acknowledgments

This work was supported by Washington Sea Grant R/RCE-9 to JKP and a CICOES graduate fellowship to JKW. This research would not be possible without the tireless effort of oil spill response volunteers and thousands of COASST participants. The idea seed of this paper was developed by the master's thesis of Kate Litle and assisted by initial analyses developed by André Punt. Julian Olden assisted with the initial statistical design and coding. Francis Wiese, William Sydeman, and three anonymous reviewers provided crucial and critical feedback on this manuscript.

References

Anker-Nilssen, T. (1987). *Methods for consequence analysis oil/sea-birds.* Viltrapport 44, Directorate for Nature Management, Trondheim, 100 pp. (in Norwegian).

Barros, Á., Álvarez, D., & Velando, A. (2014). Long-term reproductive impairment in a seabird after the Prestige oil spill. *Biology Letters*, *10*(4), 20131041.

Burger, A. E. (1992). *The effects of oil pollution on seabirds off the west coast of Vancouver Island* (No. CWS--93-02107).

Burger, A. E. (1993). *Effects of the Nestucca oil spill on seabirds along the coast of Vancouver Island in 1989* (No. CWS/PYR--69-5/179E). Canadian Wildlife Service.

Burger, A. E., & Fry, D. M. (1993). Effects of oil pollution on seabirds in the northeast Camphuysen, C. J. (1989). *Beached bird surveys in the Netherlands 1915/1988. Seabird mortality in the southern North Sea since the early days of oil pollution.* Technisch Rapport Vogelbescherming 1. Werkgroep Noordzee, Amsterdam.

Camphuysen, C. J. (1998). Beached bird surveys indicate decline in chronic oil pollution in the North Sea. *Marine pollution bulletin*, *36*(7), 519-526.

Camphuysen, C. J. (2007). *Chronic oil pollution in Europe, a status report.* Report Royal Netherlands Institute for Sea Research, commissioned by International Fund for Animal Welfare (IFAW), Brussels.

Castège, I., Hémery, G., Roux, N., d'Elbée, J., Lalanne, Y., d'Amico, F., & Mouchès, C. (2004). Changes in abundance and at-sea distribution of seabirds in the Bay of Biscay prior to, and following the "Erika" oil spill. *Aquatic Living Resources*, *17*(3), 361-367.

Chilvers, B. L., & Battley, P. F. (2019). Species prioritization index for oiled wildlife response planning in New Zealand. *Marine pollution bulletin*, *149*, 110529.

Clark, R. B. (1984). Impact of oil pollution on seabirds. *Environmental Pollution Series A, Ecological and Biological*, *33*(1), 1-22.

Crain, C. M., Halpern, B. S., Beck, M. W., & Kappel, C. V. (2009). Understanding and managing human threats to the coastal marine environment. *Annals of the New York Academy of Sciences*, *1162*(1), 39-62.

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*(9), 1412-1430.

Engel, A. S., Liu, C., Paterson, A. T., Anderson, L. C., Turner, R. E., & Overton, E. B. (2017). Salt marsh bacterial communities before and after the Deepwater Horizon oil spill. *Applied and environmental microbiology*, *83*(20).

Esler, D., Ballachey, B. E., Matkin, C., Cushing, D., Kaler, R., Bodkin, J., ... & Kloecker, K. (2018). Timelines and mechanisms of wildlife population recovery following the Exxon Valdez oil spill. *Deep Sea Research Part II: Topical Studies in Oceanography*, *147*, 36-42.

Ford, G., Varoujean, D., Warrick, D., Williams, W., Lewis, D., Hewitt, C., & Casey, J. (1991). *Seabird mortality resulting from the Nestucca oil spill incident winter 1988-89.* Ecological consulting.

Fox, C.H., O'Hara, P.D., Bertazzon, S., Morgan, K., Underwood, F.E. and Paquet, P.C., 2016. A preliminary spatial assessment of risk: Marine birds and chronic oil pollution on Canada's Pacific coast. *Science of the Total Environment*, 573, pp.799-809.

Garthe, S., & Hüppop, O. (2004). Scaling possible adverse effects of marine wind farms on seabirds: developing and applying a vulnerability index. *Journal of applied Ecology*, *41*(4), 724-734.

Halpern, B. S., Selkoe, K. A., Micheli, F., & Kappel, C. V. (2007). Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology*, *21*(5), 1301-1315.

Hartung, R. (1967). Energy metabolism in oil-covered ducks. *The Journal of Wildlife Management*, 798-804.

Hass T., Parrish J. K. (2013) *Beached birds: a COASST field guide* (4th ed.). Wavefall Press.

Hatch, S. A., Gill, V. A., & Mulcahy, D. M. (2010). Individual and colony-specific wintering areas of Pacific northern fulmars (Fulmarus glacialis). *Canadian Journal of Fisheries and Aquatic Sciences*, *67*(2), 386-400.

Jaccard, P. (1901). Comparative study of floral distribution in a portion of Alps and Jura. *The Company Vaudoise Bulletin of Natural Sciences, 37*(5), 547-579.

Jones, P. H., Howells, G., Rees, E. I. S., & Wilson, J. (1970). Effect of 'Hamilton Trader' oil on birds in the Irish Sea in May 1969. *British Birds*, *63*(3), 97-110.

Jones, T., Parrish, J. K., Peterson, W. T., Bjorkstedt, E. P., Bond, N. A., Ballance, L. T., ... & Harvey, J. (2018). Massive mortality of a planktivorous seabird in response to a marine heatwave. *Geophysical Research Letters*, *45*(7), 3193-3202.

Jones, T., Parrish, J. K., Punt, A. E., Trainer, V. L., Kudela, R., Lang, J., ... & Hickey, B. (2017). Mass mortality of marine birds in the Northeast Pacific caused by Akashiwo sanguinea. *Marine Ecology Progress Series*, *579*, 111-127.

King, J. G., & Sanger, G. A. (1979). Oil vulnerability index for marine oriented birds. *Conservation of marine birds of northern North America*, *11*, 227-239.

Legendre, P., & Legendre, L. (Eds.). (2013). *Developments in numerical ecology* (Vol. 14). Springer Science & Business Media.

Lorentsen, S.H. & Anker-Nilssen, T. (1993). Behaviour and oil vulnerability of Fulmars Fulmarus glacialis during an oil spill experiment in the Norwegian Sea. *Marine pollution bulletin*, *26*(3), pp.144-146.

McDonald, T. L., Erickson, W. P., & McDonald, L. L. (2000). Analysis of count data from before-after control-impact studies. *Journal of Agricultural, Biological, and Environmental Statistics*, 262-279.

Menardi, G. (2011). Density-based Silhouette diagnostics for clustering methods. *Statistics and Computing*, *21*(3), 295-308.

Momot, J. J. (1994). *Nestucca oil spill restoration plan*. US Fish and Wildlife Service, Ecological Services, Olympia State Office.

Munilla, I., Arcos, J. M., Oro, D., Álvarez, D., Leyenda, P. M., & Velando, A. (2011). Mass mortality of seabirds in the aftermath of the Prestige oil spill. *Ecosphere*, *2*(7), 1-14.

National Oceanic and Atmospheric Administration (n.d.) *F/V Tenyo Maru*. Retrieved March 12, 2021, from https://incidentnews.noaa.gov/incident/5017#!

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., ... & Wagner, H. (2015). Vegan community ecology package: ordination methods, diversity analysis and other functions for community and vegetation ecologists. *R package ver*, 2-3.

Overland, J., Rodionov, S., Minobe, S., & Bond, N. (2008). North Pacific regime shifts: Definitions, issues and recent transitions. *Progress in Oceanography*, *77*(2-3), 92-102.

Paine, R. T., Ruesink, J. L., Sun, A., Soulanille, E. L., Wonham, M. J., Harley, C. D., ... & Secord, D. L. (1996). Trouble on oiled waters: lessons from the Exxon Valdez oil spill. *Annual Review of Ecology and Systematics*, *27*(1), 197-235.

Parrish, J. K., Bond, N., Nevins, H., Mantua, N., Loeffel, R., Peterson, W. T., & Harvey, J. T. (2007). Beached birds and physical forcing in the California Current System. *Marine Ecology Progress Series*, *352*, 275-288.

Parsons, M. L., Morrison, W., Rabalais, N. N., Turner, R. E., & Tyre, K. N. (2015). Phytoplankton and the Macondo oil spill: A comparison of the 2010 phytoplankton assemblage to baseline conditions on the Louisiana shelf. *Environmental Pollution*, *207*, 152-160.

Piatt, J. F., Lensink, C. J., Butler, W., Kendziorek, M., & Nysewander, D. R. (1990). Immediate impact of the 'Exxon Valdez' oil spill on marine birds. *The Auk*, *107*(2), 387-397.

Piatt, J. F., Parrish, J. K., Renner, H. M., Schoen, S. K., Jones, T. T., Arimitsu, M. L., ... & Sydeman, W. J. (2020). Extreme mortality and reproductive failure of common murres resulting from the northeast Pacific marine heatwave of 2014-2016. *PLoS One*, *15*(1), e0226087.

Python Core Team (2021). Python: A dynamic, open source programming language. *Python Software Foundation*, *78*. https://www.python.org/.

R Core Team (2021). R: A language and environment for statistical computing. URL https://www.R-project.org/.

Roth, A. M. F., & Baltz, D. M. (2009). Short-term effects of an oil spill on marsh-edge fishes and decapod crustaceans. *Estuaries and Coasts*, *32*(3), 565-572.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Summary of the Tenyo Maru oil spill restoration [electronic resource]. (2006). Natural Resource Trustees, Makah Indian Tribe, Washington Dept. of Ecology, National Oceanic and Atmospheric Administration, U.S. Fish and Wildlife Service.

Tasker, M. L., Webb, A., Harrison, N. M., & Pienkowski, M. W. (1990). Vulnerable concentrations of marine birds west of Britain. *Nature Conservancy Council, Peterborough*, 1-45.

Tenyo Maru Trustee Council & U.S. Fish Wildlife Service. (2000). *Final restoration plan and environmental assessment for the Tenyo Maru oil spill.* Lacey, WA: U.S. Dept. of the Interior, Fish and Wildlife Service.

Thompson, C. W., Ryan, K., Parrish, J. K., Zador, S., & Warheit, K. I. (2003). Restoration of common murre colonies in the Copalis National Wildlife Refuge--Phase I: draft final report to the *Tenyo Maru* Trustee Council.

Wahl, T. R., Speich, S. M., Manuwal, D. A., Hirsch, K. V., & Miller, C. (1981). *Marine bird populations of the Strait of Juan de Fuca, Strait of Georgia and adjacent waters in 1978 and 1979* (No. PB-82-132622). Washington Univ., Seattle (USA). Coll. of Forest Resources.

Wang, B., Luo, X., Yang, Y. M., Sun, W., Cane, M. A., Cai, W., ... & Liu, J. (2019). Historical change of El Niño properties sheds light on future changes of extreme El Niño. *Proceedings of the National Academy of Sciences*, *116*(45), 22512-22517.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236-244.

Webb, A., Elgie, M., Hidef, C. I., Pollock, C., Barton, C., Burns, S., & Hawkins, K. (2016). Sensitivity of offshore seabird concentrations to oil pollution around the United Kingdom: Report to oil & gas UK. *Document No HP00061701. Available online at http://jncc. defra. gov. uk/page-7373*.

Webb, A., Stronach, A., Tasker, M. L., & Stone, C. J. (1995). *Vulnerable concentrations of seabirds south and west of Britain*. Joint Nature Conservation Committee.

Wiese, F. K. (2003). Sinking rates of dead birds: improving estimates of seabird mortality due to oiling. *Marine Ornithology*, *31*(1), 65-70.

Wiese, F. K., & Ryan, P. C. (2003). The extent of chronic marine oil pollution in southeastern Newfoundland waters assessed through beached bird surveys 1984–1999. *Marine Pollution Bulletin*, *46*(9), 1090-1101.

Williams, J. M., Tasker, M. L., Carter, I. C., & Webb, A. (1995). A method of assessing seabird vulnerability to surface pollutants. *Ibis*, *137*, S147-S152.

Wood, M. C., & Wood, J. C. (2005). *Joseph M. Juran: critical evaluations in business and management*. Routledge.

Chapter 1- Supplementary Materials

Waugh J.K. et al. (2022) Using Beached Bird Data to Assess Seabird Oiling Susceptibility.

Section 1. Assessment of decadal-scale shifts in species composition

Given the baseline dataset begins 11-14 years after the spills being examined, there is a potential concern that any differences in species composition between the spill and baseline datasets are not due to differential oiling susceptibility but because species composition changed over time between the years in which the spills occurred (1988 and 1991) and the start of the COASST dataset (2002).

To partially address this, we examined the baseline dataset for decadal-scale shifts in species composition, reasoning that if there was no indication of shifts in species composition within the ~ 20-year baseline dataset then patterns of species composition are likely also applicable to the years in which the spills occurred, 11-14 years prior to the start of the COASST dataset.

We split the dataset of baseline proportional composition (see main text) into two groups (2002-2010 and 2011-2018) and tested for differences in taxonomic composition between these two groups using a perMANOVA test (Anderson 2014) and the adonis function in the vegan R package (Oksanen 2015; **Table S1**). Results of the perMANOVA test were indicative of no significant differences in overall taxonomic composition between early (2002-2010) and late (2011-2018) baseline periods for either the regions associated with the *Tenyo Maru* ($p = 0.3292$; **Table S1**) or *Nestucca* ($p = 0.106$; **Table S1**) spill, judged based on p-value's > 0.1 in both instances (calculated based on 10,000 permutations).

Table S1. Results of perMANOVA analysis for regions associated with the *Tenyo Maru* and *Nestucca* spills with p-values calculated based on 10,000 permutations.

	df	Sum of squares	Mean squares	F	R^2	p(>F)
Tenyo Maru						
Group	1	0.195	0.195	1.110	0.061	0.329
Residuals	17	2.984	0.176		0.939	
Total	18	3.179				
Nestucca						
Group	1	0.382	0.382	1.698	0.091	0.106
Residuals	17	3.822	0.225		0.909	
Total	18	4.204				

Section 2. Sensitivity of baseline proportional abundance to time window width

In order to sample the baseline dataset, we assumed a seasonal window for each spill, and used only those surveys that fell within the spill-specific seasonal window. To investigate the degree to which our choice of window duration may have influenced our results, we calculated the correlation between measures of taxon proportional abundance calculated for our 45-day window (centered on the date of the spill), relative to the same measures calculated for a shorter time window (30-day) and a longer time window (60-day). For each of the three window sizes (30-, 45-, and 60-day) we calculated the mean taxon proportional abundance among years and its 95% confidence interval based on bootstrap resampling (see main text).

For both spills, there was a high degree of correlation (Pearson's correlation coefficient > 0.99 for all comparisons) in proportional abundances between our chosen window versus measures calculated based on shorter and longer durations (**Figure S1**). Differences in average proportional abundance were minimal, amounting to a mean change in % abundance of 0.44 (45-day window relative to 60-day) to 0.48% (45-day window relative to 30-day window), and a

maximum change of 3.6%. All absolute differences that constituted a greater than 1% change between alternate window durations were for taxa with high proportional abundance (45-day mean percent composition ranging from 13-51%), and so would not strongly affect our results. Therefore, collectively we conclude that our results would not be meaningfully altered given an alternate sampling window duration.

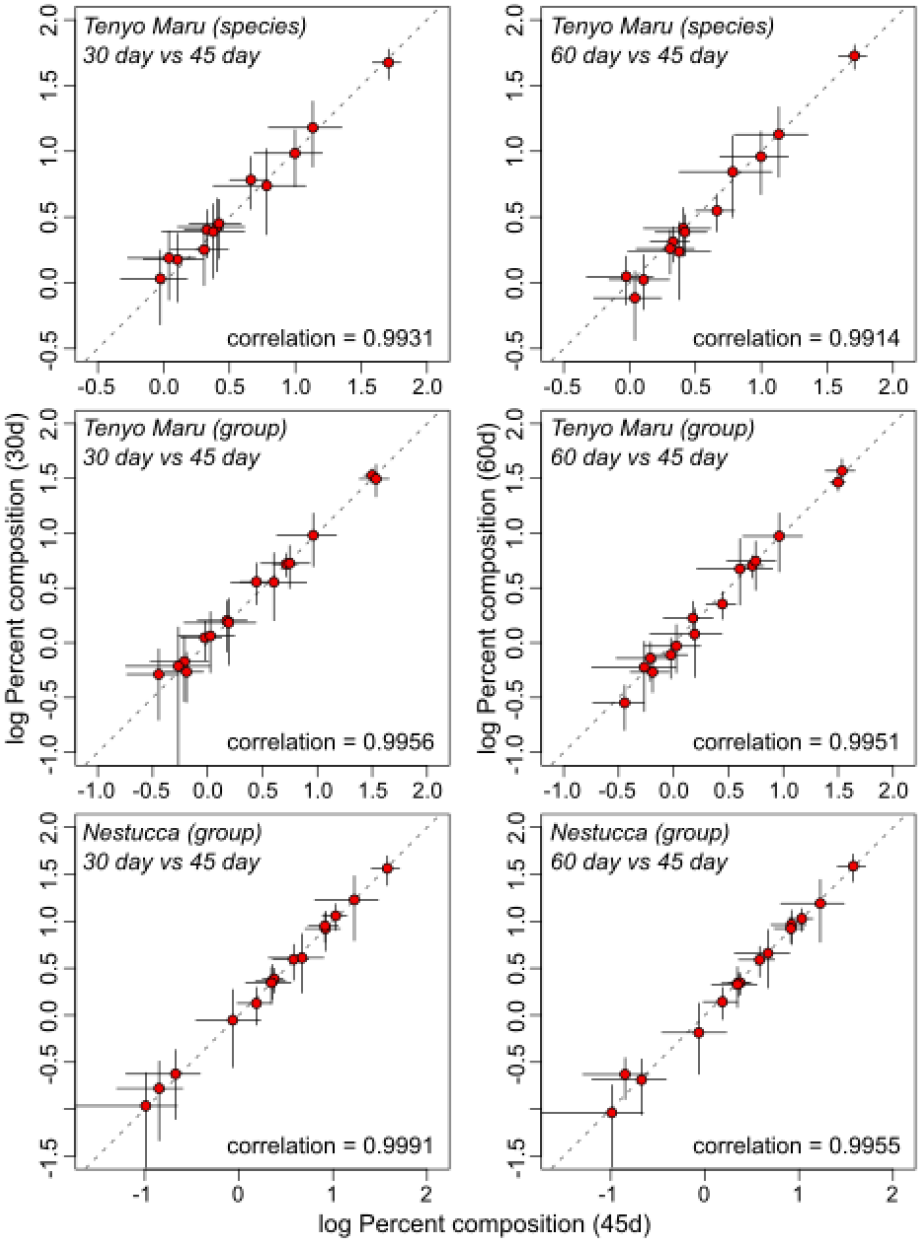


Figure S1. Comparison of proportional composition among taxa calculated based on 30-, 45- and 60-day seasonal windows. Each panel represents bootstrapped mean percent composition (log-transformed) and its 95% confidence interval based on among year differences, calculated for each taxon based on alternate seasonal window durations (30 and 60) relative to the 45-day window used in the main text. Each panel references the spill and level of analysis for which these analyses were performed, and the correlation value in the lower left is the Pearson's correlation coefficient for the points shown. In each panel the dashed line represents $y = x$.

Section 3. Redistribution of unknown counts

3.1 Tenyo Maru spill dataset

During the *Tenyo Maru* spill, after the exclusion of shorebirds, terrestrial birds, and gulls, there remained only 14 individuals not identified to species. These individuals were redistributed proportionally among species within their respective group based on their abundance expressed as a proportion of the known (i.e., identified to species) total for that group (**Table S2**).

The redistribution of counts had a minimal effect on calculated proportions (mean absolute difference = 2.96×10^{-4}, range = 0 to 2.74×10^{-3}), with the largest relative differences, calculated as change divided by raw proportions expressed as a percentage, for storm-petrels (proportions were 50% higher) and cormorants (36.0% higher), although this is largely due to their relatively low abundances. For the remaining taxa, there was a < 2% relative change in proportions after counts were redistributed.

Table S2. Redistribution of unknown birds observed during the *Tenyo Maru* oil spill.

Group (n unknown)	Species	Count	% of known in group	Redistribution of unknown	Final
Cormorants (12)	Pelagic cormorant	18	55%	+6.55	24.55
	Brandt's cormorant	10	30%	+3.64	13.64
	Double-crested cormorant	5	15%	+1.82	6.82
Shearwater (1)	Sooty shearwater	49	91%	+0.91	49.91
	Short-tailed shearwater	5	9%	+0.09	5.09
Storm-petrel (1)	Fork-tailed storm-petrel	1	50%	+0.5	1.5
	Leech's storm-petrel	1	50%	+0.5	1.5

3.2 Nestucca spill dataset

For the *Nestucca* spill, after the exclusion of shorebirds and terrestrial birds there remained 282 alcids, 669 grebes and 5 tubenoses not identified to group. These were redistributed proportionally among groups based on their abundance expressed as a proportion of the known (i.e., identified to group) total for that super-group (**Table S3**).

The redistribution of counts had a minimal effect on calculated proportions (mean absolute difference = 8.4×10^{-3}, range = 0 to 4.74×10^{-2}), with the largest relative differences, calculated as change divided by raw proportions expressed as a percentage, for small and large grebes, which underwent 5-6-fold increases in proportional representation due to the large relative number of unidentified individuals. Without redistribution of unknown grebes, both small and large grebe groups would be under-represented in the *Nestucca* spill data in any subsequent analyses, effectively biasing any analyses towards concluding that grebes were less susceptible to oiling. Therefore, we used redistributed counts to mitigate against such bias, but recommend that results for large and small grebes be interpreted with some caution due to the relatively large proportion of unknown grebes. For all remaining taxa, there was a < 8% relative change in proportions after counts were redistributed.

Table S3. Redistribution of unknown birds (group level) observed during the *Nestucca* oil spill.

Super-Group (n unknown)	Group	Count	% of known in super-group	Redistribution of unknown	Final
Alcids (282)	Guillemots	1	0.01%	+0.03	1.03
	Murres	8091	86.1%	+242.9	8333.9
	Puffins	48	0.5%	+1.4	49.4
	Auklets	1238	13.2%	+37.2	1275.2
	Murrelets	17	0.2%	+0.5	17.5
Grebes (669)	Large Grebes	97	88.2%	+589.9	686.9
	Small Grebes	13	11.8%	+79.1	92.1

Tubenoses (5)	Fulmars	69	84.1%	+4.2	73.2
	Shearwaters	13	15.9%	+0.8	13.8

3.3 Baseline dataset – Tenyo Maru spatio-seasonal window

The redistribution of unknown species for the baseline dataset extracted for the *Tenyo Maru* georegion is given in **Table S4.**

Table S4. Redistribution of unknown birds included in the baseline dataset, restricted to the *Tenyo Maru* spatio-seasonal window.

Year	Group (n unknown)	Species	Count	% of known in group	Redist.	Final
2002	Small alcid (1)	Cassin's auklet	2	100%	+1	3
	Cormorant (1)	Pelagic cormorant	1	50%	+0.5	1.5
		Double-crested cormorant	1	50%	+0.5	1.5
2003	Large alcid (1)	Pigeon guillemot	3	5.3%	+0.05	3.05
		Common murre	49	86.0%	+0.86	49.86
		Rhinoceros auklet	4	7.0%	+0.07	4.07
		Tufted puffin	1	1.8%	+0.02	1.02
2004	Small alcid (2)	Cassin's auklet	14	93.3%	+1.87	14.87
		Ancient murrelet	1	6.7%	+0.13	1.13
	Cormorants (3)	Pelagic cormorant	6	75%	+2.25	8.25
		Brandt's cormorant	2	25%	+0.75	2.75
2005	Large alcid (1)	Pigeon guillemot	3	1.7%	+0.02	3.02
		Common murre	171	95.5%	+0.96	171.96
		Rhinoceros auklet	5	2.8%	+0.03	5.03
	Small alcid (2)	Cassin's auklet	3	75%	+1.5	4.5
		Marbled murrelet	1	25%	+0.5	1.5
2006	Small alcid (1)	Cassin's auklet	11	100%	+1	12
	Cormorants (1)	Pelagic cormorant	1	50%	+0.5	1.5
		Brandt's cormorant	1	50%	+0.5	1.5
2007	Small alcid (2)	Cassin's auklet	5	83.3%	+1.7	5.7
		Marbled murrelet	1	16.7%	+0.3	1.3
	Cormorants (4)	Pelagic cormorant	3	50%	+2	5
		Brandt's cormorant	3	50%	+2	5
2008	Loon (1)	Common loon	0	/	+0.33	0.33
		Pacific loon	0	/	+0.33	0.33
		Red-throated loon	0	/	+0.33	0.33
2009	Murrelet (1)	Ancient murrelet	1	100%	+1	2
	Small Alcids (2)	Cassin's auklet	2	50%	+1	3
		Ancient murrelet	2	50%	+1	3
	Large cormorant (1)	Brandt's cormorant	4	66.7%	+0.67	4.7
		Double-crested cormorant	2	33.3%	+0.33	2.3
	Cormorant (3)	Pelagic cormorant	9	60%	+1.8	10.8
		Brandt's cormorant	4	26.7%	+0.8	4.8
		Double-crested cormorant	2	13.3%	+0.4	2.4
	Albatross (1)	Black-footed albatross	1	100%	+1	2
2010	Large Alcid (1)	Common murre	6	66.7%	+0.67	6.7
		Rhinoceros auklet	3	33.3%	+0.33	3.3
	Large cormorants (1)	Brandt's cormorant	2	66.7%	+0.67	3.7
		Double-crested cormorant	1	33.3%	+0.33	1.3

	Shearwaters (1)	Sooty shearwater	5	100%	+1	6
2012	Shearwater/Fulmar (1)	Northern fulmar	3	4.4%	+0.04	3.04
		Sooty shearwater	63	92.6%	+0.93	63.9
		Pink-footed shearwater	2	2.9%	+0.03	32.03
2013	Large Alcid (1)	Pigeon guillemot	1	5.8%	+0.06	1.06
		Common murre	13	76.5%	+0.76	13.8
		Rhinoceros auklet	3	17.6%	+0.18	3.2
	Large cormorants (1)	Brandt's cormorant	4	50%	+0.5	4.5
		Double-crested cormorant	4	50%	+0.5	4.5
	Cormorants (2)	Pelagic cormorant	7	43.8%	+0.88	7.9
		Brandt's cormorant	4.5	28.1%	+0.56	5.1
		Double-crested cormorant	4.5	28.1%	+0.56	5.1
	Shearwater (1)	Sooty shearwater	55	100%	+1	56
2014	Puffins (1)	Rhinoceros auklet	12	100%	+1	13
	Small Alcid (1)	Cassin's auklet	8	100%	+1	9
	Alcid (1)	Common murre	61	73.5%	+0.74	61.74
		Rhinoceros auklet	13	15.7%	+0.16	13.16
		Cassin's auklet	9	10.8%	+0.11	9.11
	Cormorants (2)	Pelagic cormorant	3	37.5%	+0.75	3.75
		Brandt's cormorant	1	12.5%	+0.25	1.25
		Double-crested cormorant	4	50%	+1	5
	Storm-petrel (1)	Fork-tailed storm-petrel	1	100%	+1	2
2015	Large Alcid (2)	Pigeon guillemot	1	0.3%	+0.01	1.01
		Common murre	348	96.4%	+1.93	349.9
		Rhinoceros auklet	11	3.0%	+0.06	11.1
		Tufted Puffin	1	0.3%	+0.01	1.01
	Large cormorant (5)	Brandt's cormorant	6	50%	+2.5	8.5
		Double-crested cormorant	6	50%	+2.5	8.5
	Cormorants (3)	Pelagic cormorant	5	22.7%	+0.68	5.7
		Brandt's cormorant	8.5	38.6%	+1.16	9.7
		Double-crested cormorant	8.5	38.6%	+1.16	9.7
	Shearwater/fulmar (2)	Northern fulmar	12	29.3%	+0.59	12.6
		Sooty shearwater	27	65.9%	+1.32	28.3
		Short-tailed shearwater	2	4.9%	+0.10	2.1
2016	Puffins (1)	Rhinoceros auklet	72	100%	+1	73
	Small alcid (1)	Cassin's auklet	7	100%	+1	8
	Large cormorants (1)	Brandt's cormorant	1	100%	+1	2
	Cormorant (1)	Pelagic cormorant	1	33.3%	+0.33	1.33
		Brandt's cormorant	2	66.7%	+0.67	2.67
2017	Large Alcid (1)	Pigeon guillemot	4	1.6%	+0.02	4.02
		Common murre	248	98.0%	+0.98	249
		Rhinoceros auklet	1	0.4%	+0.00	1.00
		Brandt's cormorant	3	75%	+0.75	3.75

	Large cormorants (1)	Double-crested cormorant	1	25%	+0.25	1.25
	Cormorants (1)	Pelagic cormorant	1	16.7%	+0.17	1.2
		Brandt's cormorant	3.75	62.5%	+0.63	4.4
		Double-crested cormorant	1.25	20.8%	+0.21	1.5
	Shearwater/fulmar (1)	Sooty shearwater	2	66.7%	+0.67	2.7
		Northern fulmar	1	33.3%	+0.33	1.33
2018	Large cormorants (1)	Double-crested cormorant	2	100%	+1	3
	Cormorants (2)	Pelagic cormorant	2	40%	+0.8	2.8
		Double-crested cormorant	3	60%	+1.2	4.2
	Loons (1)	Common loon	0	/	+0.33	0.3
		Pacific loon	0	/	+0.33	0.3
		Red-throated loon	0	/	+0.33	0.3

3.4 Baseline dataset – Nestucca spatio-seasonal window

The redistribution of unknown species for the baseline dataset extracted for the *Nestucca* georegion is given in **Table S5**.

The largest redistribution was in 2014, where many individuals were only identified as small alcid. This was during the Cassin's auklet mass mortality event (Jones et al. 2018), where due to verification standards (i.e., photographic quality) and the need for photographing groups of birds due to the number of birds per survey, individual birds could not be verified as Cassin's auklets. However, due to the high proportion of auklets in the identified set of individuals (98.8%), we believe that redistribution is valid in this case.

Table S5. Redistribution of unknown birds included in the baseline dataset, restricted to the *Nestucca* spatio-seasonal window.

Year	Super-group (n unknown)	Group	Count	% of known in super group	Redist.	Final
2003	Large alcid (1)	Murres	2	100%	+1	3
2005	Small alcid (3)	Auklets	10	83.3%	+2.5	12.5
		Murrelets	2	16.7%	+0.5	2.5
	Shearwater/fulmar (3)	Fulmars	185	98.9%	+3.0	188.0
		Shearwaters	2	1.1%	+0.0	2.0
2006	Small alcid (1)	Auklets	23	85.2%	+0.85	23.9
		Murrelets	4	14.8%	+0.15	4.2
2007	Small alcid (1)	Auklets	5	83.3%	+0.83	5.8
		Murrelets	1	16.7%	+0.17	1.2
2010	Small alcid (1)	Auklets	11	61.1%	+0.61	11.61
		Murrelets	7	38.9%	+0.39	7.39
	Alcid (1)	Murres	27	26.2%	+0.26	27.3
		Guillemots	2	1.94%	+0.02	2.0
		Puffins	55	53.4%	+0.53	55.5
		Auklets	11.6	11.3%	+0.11	11.7
		Murrelets	7.4	7.2%	+0.07	7.5
2011	Large alcid (1)	Murres	97	50.8%	+0.51	97.5
		Puffins	93	48.7%	+0.49	93.5
		Guillemot	1	0.5%	+0.00	1.00
	Shearwater/fulmar (1)	Fulmar	44	84.6%	+0.85	44.85
		Shearwater	8	15.4%	+0.15	8.15
2012	Shearwater/fulmar (2)	Fulmar	154	99.4%	+0.99	155
		Petrels	1	0.6%	+0.01	1.01
2013	Shearwater/fulmar (5)	Fulmar	115	89.8%	+0.9	115.9
		Shearwater	13	10.2%	+0.1	13.1
2014	Small alcid (1287)	Auklets	3538	98.8%	+1272	4810
		Murrelets	42	1.2%	+15	57
	Large alcid (2)	Murre	93	57.1%	+1.1	94.1
		Puffin	68	41.7%	+0.8	68.6
		Guillemot	2	1.2%	+0.02	2.0
	Shearwater/fulmar (5)	Fulmar	229	98.7%	+4.9	233.9
		Shearwater	3	1.3%%	+0.1	3.1
2015	Small alcid (22)	Auklets	177	97.3%	+21.4	198.4
		Murrelets	5	2.7%	+0.6	5.6
2016	Small alcid (1)	Auklet	10	71.4%	+0.71	10.7
		Murrelet	4	28.6%	+0.29	4.3
	Alcid (1)	Murres	21	45.7%	+0.46	21.5

		Puffins	10	21.7%	+0.22	10.2
		Auklets	10.7	23.3%	+0.23	10.9
		Murrelets	4.3	9.3%	+0.09	4.4
	Shearwater/fulmar	Fulmar	175	96.7%	+6.77	181.8
	(7)	Shearwater	6	3.3%	+0.23	6.2
2017	Small alcid (2)	Auklet	22	95.7%	+1.91	23.9
		Murrelet	1	4.3%	+0.09	1.1
2018	Small alcid (1)	Auklet	69	92%	+0.92	69.9
		Murrelet	6	8%	+0.08	6.1
	Shearwater/fulmar	Fulmar	3	60%	+0.6	3.6
	(1)	Shearwater	2	40%	+0.4	2.4

Section 4. Species omitted based on rarity (*Tenyo Maru*)

Following the steps described in the manuscript, we omitted 10 species, with a total of 194 individuals (**Table S6**) due to low numbers of individuals in the corresponding baseline dataset.

Table S6. Species documented in the *Tenyo Maru* spill dataset omitted from representation analyses based on rarity in the COASST dataset. Also shown are the number (N) and % of carcasses documented during the *Tenyo Maru* spill for each species, as well as a summary representation of baseline abundances for the corresponding spatio-seasonal window among years, and the overall % (averaged among years) composition for those species.

	Spill data		**Baseline data**	
Species	**N**	**%**	**Abundances (n year)**	**Average proportion (%)**
Tufted puffin	127	3.15%	1 (5), 0 (12)	0.24%
Marbled murrelet	45	1.11%	4 (1), 1 (3), 0 (13)	0.20%
Surf scoter	10	0.25%	2 (2), 1 (6), 0 (9)	0.53%
Short-tailed shearwater	5	0.12%	2 (3), 1 (1), 0 (13)	0.20%
Western grebe	2	0.05%	1 (2), 0 (15)	0.11%
Arctic tern	1	0.02%	0 (17)	0%
Common loon	1	0.02%	2 (1), 1 (3), 0 (13)	0.20%
Horned puffin	1	0.02%	0 (17)	0%
Leach's storm-petrel	1	0.02%	3 (1), 2 (1), 1 (3), 0 (12)	0.52%
Red-throated loon	1	0.02%	1 (4), 0 (13)	0.32%

Section 5. Change in variation explained with additional PCoA axes

We examined scree plots of the percent variation explained by PCoA axes, examining for inflection where the addition of further axes resulted in relatively little increase in variation explained. For all datasets and levels of analysis (species, group), the first two PCoA axes explained more than 55% of variation (see main text) with subsequent axes each accounting for < 13% of total variation (**Figure S2**), indicating that the first two axes provided a suitable visualization of the majority of the variation exhibited in each dataset.

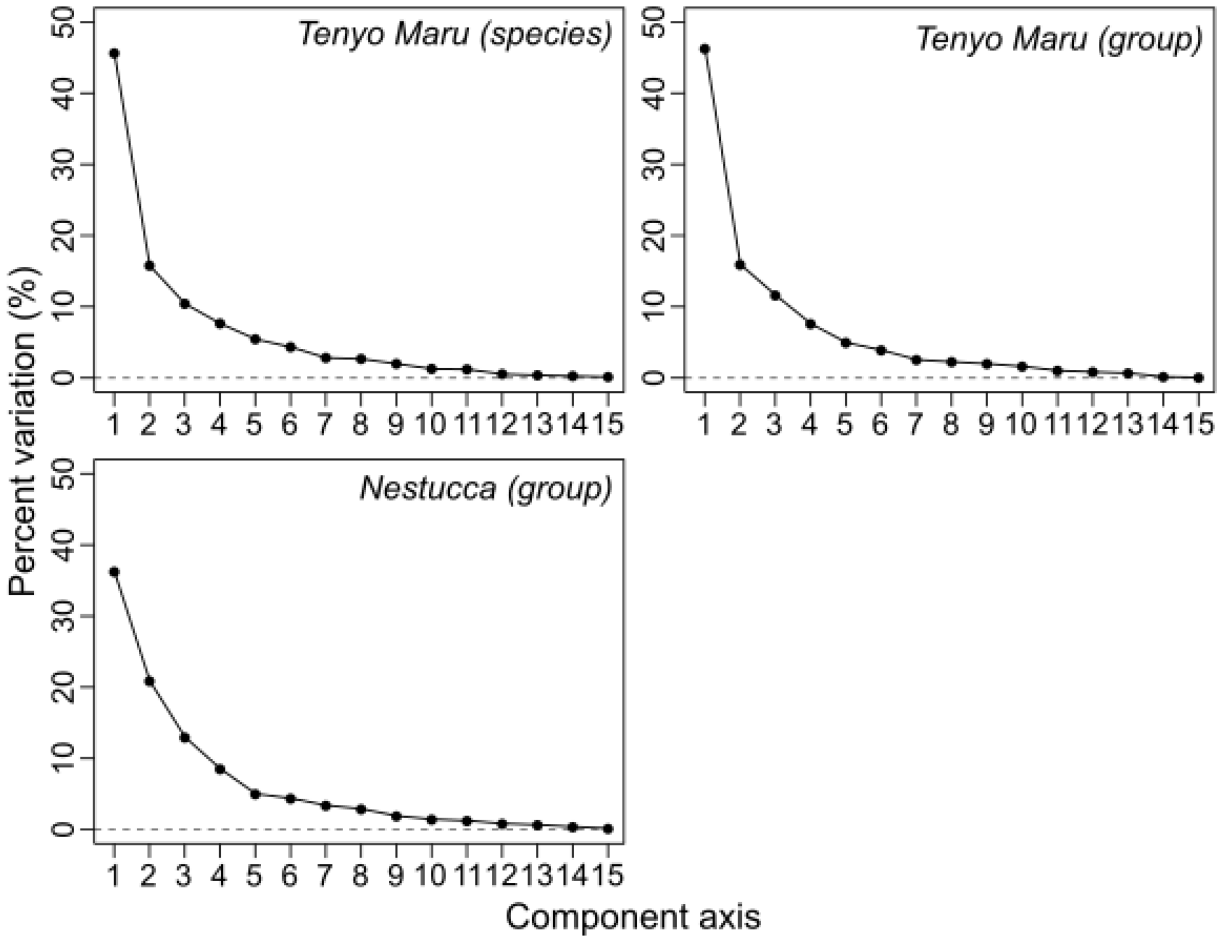


Figure S2. Scree plot of change in percent variation explained by each principal component axis when applied to proportional composition data for *Tenyo Maru* and *Nestucca* spill-centric datasets.

Section 6. Taxonomic loadings on PCoA axes

Species responsible for predominant patterns in taxonomic variation were identified by calculating the coefficient of determination, R^2, between an individual taxon's proportional abundance among years (response) and the axis scores (predictors) resulting from the PCoA. A permutation test was then applied, which generated a null distribution of expected R^2 values if there was no relationship between taxon proportional abundance and axis scores. Observed and expected distributions for R^2 were then evaluated to calculate a p-value corresponding to the probability of obtaining an R^2 at least as extreme as the observed value if there was no relationship between taxon proportional abundance and axis scores. We plot vectors representative of individual taxon loadings onto PCoA ordination plots for taxa with p-values < 0.05 to indicate taxa responsible for major modes of taxonomic variation.

Table S7. Species loadings (or principal weights in the eigenvectors) on each principal coordinate for the *Tenyo Maru* (species, group) and *Nestucca* (group) analyses. Statistical significance was evaluated based on permutations of no relationship with primary/secondary PCoA axes (n_{perm} = 1000). Taxa with $p < 0.05$ are indicated in **bold**.

		Loading			
Analysis	Taxon	Axis 1	Axis 2	R^2	p
Tenyo Maru (species)	Pigeon guillemot	0.95	0.32	0.04	0.76
	Common murre	**0.99**	**-0.13**	**0.97**	**< 0.01**
	Rhinoceros auklet	-1.00	0.10	0.01	0.94
	Cassin's auklet	-0.45	0.89	0.04	0.71
	Caspian tern	-0.01	1.00	0.14	0.31
	Brandt's cormorant	-1.00	-0.03	0.17	0.23
	Double-crested cormorant	**-0.66**	**0.76**	**0.47**	**0.02**
	Pelagic cormorant	-0.14	0.99	0.03	0.82
	Black-footed albatross	**-0.52**	**0.85**	**0.46**	**0.02**
	Northern fulmar	**-0.52**	**0.86**	**0.84**	**< 0.01**
	Sooty shearwater	**-0.54**	**-0.84**	**0.96**	**< 0.01**
	Fork-tailed storm-petrel	-0.63	0.78	0.29	0.08
	White-winged scoter	**-0.53**	**0.85**	**0.60**	**< 0.01**
Tenyo Maru (group)	Guillemot	0.98	0.22	0.05	0.68
	Murre	**1.00**	**0.00**	**0.99**	**< 0.01**
	Puffin	-0.42	-0.91	0.00	0.97
	Auklet	-0.50	0.87	0.05	0.69
	Murrelet	0.84	0.54	0.12	0.40
	Gull	**-0.70**	**0.71**	**0.91**	**< 0.01**
	Tern	0.21	0.98	0.06	0.65
	Loon	-0.54	-0.84	0.02	0.85
	Cormorant	-0.90	0.44	0.23	0.14
	Albatross	-0.86	0.51	0.27	0.08
	Fulmar	**-0.82**	**0.58**	**0.45**	**0.02**
	Shearwater	**-0.37**	**-0.93**	**0.93**	**< 0.01**
	Storm-petrel	-0.91	0.42	0.17	0.27
	Scoter	**-0.69**	**0.72**	**0.34**	**0.04**
Nestucca (group)	Murre	0.34	-0.94	0.24	0.11
	Puffin	**0.29**	**-0.96**	**0.35**	**0.03**
	Auklet	**0.69**	**0.72**	**0.94**	**< 0.01**
	Murrelet	0.92	-0.38	0.13	0.35
	Large Grebe	**0.22**	**-0.98**	**0.48**	**0.02**
	Small Grebe	1.00	-0.10	0.08	0.54
	Gull	0.04	-1.00	0.27	0.09
	Loon	-0.26	-0.97	0.11	0.44
	Cormorant	0.55	-0.84	0.32	0.06
	Fulmar	**-0.93**	**0.36**	**0.97**	**< 0.01**
	Shearwater	-0.08	-1.00	0.18	0.21
	Scoter	-0.14	-0.99	0.07	0.59

Section 7. Identification and evaluation of the optimum number of clusters

We explored how many clusters were present in the data by examination of Silhouette information/width (Rousseeuw 1987; Menardi 2011). Silhouette width measures how well an observation belongs to a cluster based on distances to observations contained within the cluster to which it is assigned, relative to the next nearest alternative cluster (Menardi 2011). Observations with silhouette widths close to one are well clustered, whereas observations with silhouette widths close to zero indicate observations that lie between two or more clusters (Menardi 2011). Average silhouette width therefore represents the degree of clustering in the data, and how well it is captured by a given number of partitions. Mean silhouette width was evaluated for k = 2-10 clusters, and the optimum number of clusters was chosen as the value which maximized mean silhouette width. For the optimal k we also examined the silhouette width for each individual year to identify how well it belonged to its assigned cluster.

For the *Tenyo Maru* spill, mean silhouette widths peaked at 3 clusters (species level) and 5 clusters (group level), whereas for the *Nestucca* data, mean silhouette width peaked at 3-4 clusters, with 4 clusters performing marginally better (**Figure S3**). However, the overall clustering structure was relatively weak, with average Silhouette widths peaking at 0.32 (*Nestucca*) to 0.39 (*Tenyo Maru* – species) (**Figure S3**).

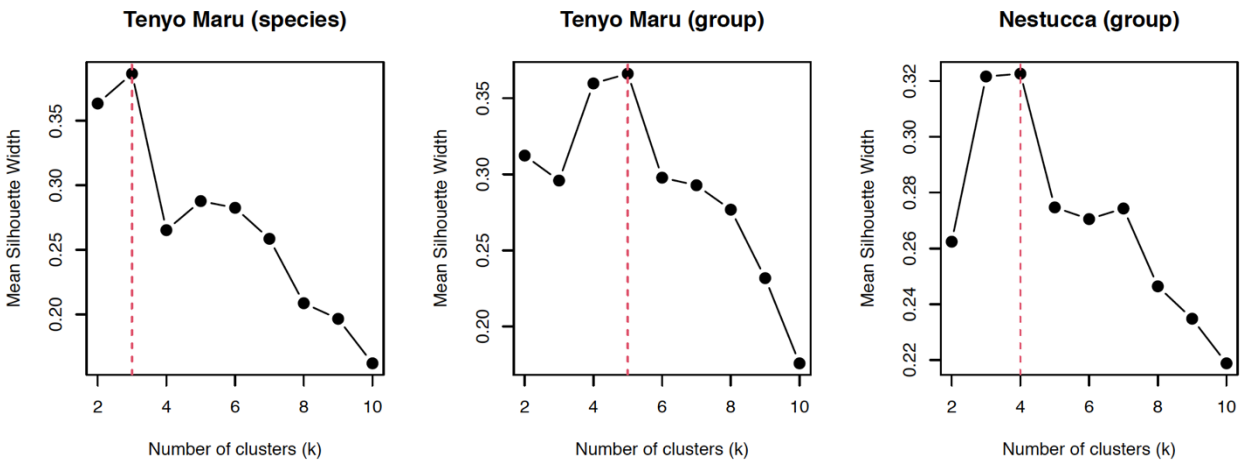


Figure S3. Mean Silhouette widths plotted for varying number of clusters as a means of identifying an optimal number of clusters for each dataset. Vertical dashed lines indicate the number of clusters which maximized mean silhouette width for each dataset.

The degree of clustering varied among clusters, based on examination of the mean silhouette widths for each cluster resulting from the optimal number of partitions (**Table S8**). The majority of clusters had average silhouette widths of 0.3-0.5, indicating moderate cluster membership, although cluster #2 in the *Tenyo Maru* species-level analyses (2011-2013) and cluster #3 in the *Nestucca* analyses (2006-2008, 2011, 2017) had lower average silhouette width, indicating that these clusters were less well defined and consisted of groups of relatively dissimilar years (**Table S8**). In addition, individual or several years assigned to particular clusters had lower silhouette widths than other members of that cluster (i.e., *Nestucca* dataset: 2004; *Tenyo Maru*: 2016, 2018; **Figure S4**), indicative of weak cluster membership of those years (**Figure S4**).

Table S8. Silhouette width statistics for the optimum number of clusters identified based on globally averaged silhouette width

			Silhouette width	
Dataset	**Cluster**	**Member years**	**Mean**	**Range**
Tenyo	**1**	***spill***, 2002-2007, 2009, 2014-2015, 2017	0.49	0.27 – 0.58
Maru	2	2008, 2010, 2016, 2018	0.09	0.02 – 0.19
(species)	3	2011-2013	0.42	0.34 – 0.48
	1	***spill***, 2004, 2015, 2017	0.52	0.42 – 0.60
Tenyo	2	2002, 2003, 2005-2007, 2009, 2014	0.31	0.28 – 0.40
Maru	3	2008, 2010, 2018	0.38	0.22 – 0.48
(group)	4	2011-2013	0.41	0.38 – 0.44
	5	2016	/	/
	1	***spill***	/	/
Nestucca	2	2002-2005, 2009, 2010, 2012, 2013, 2016	0.37	0.06 – 0.48
(group)	3	2006-2008, 2011, 2017	0.23	0.15 – 0.29
	4	2014, 2015, 2018	0.43	0.38 – 0.46

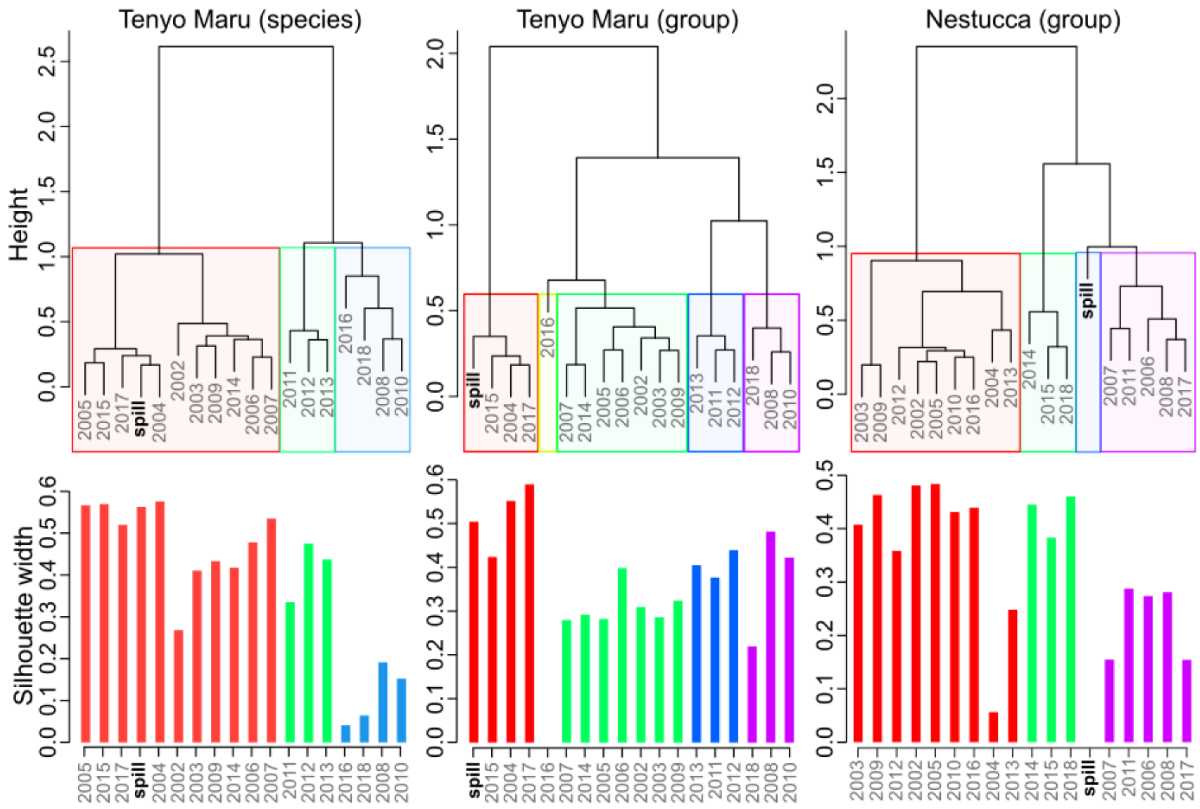


Figure S4. Dendrograms depicting the results of cluster analysis applied to the *Tenyo Maru* (species and group) and *Nestucca* (group) datasets. Overlaid colored rectangles indicate clusters identified as optimal based on average silhouette width. Lower panels show individual year silhouette widths, resulting from application of the optimal number of partitions to each dataset, as an indication of the degree of membership for each year within its assigned cluster.

Section 8. Sensitivity of results to the inclusion of dominant taxa

We investigated how our results are influenced by dominant taxa given that proportions for a particular taxon are not independent from the abundance(s) of other taxa. At the highest level we were interested in whether a taxon was over-represented or under-represented in a particular spill dataset relative to the corresponding baseline dataset. As such, we performed a series of taxon removals basing our decisions on the relative representation, which we define as the ratio of oil spill proportion to the average proportion exhibited in the corresponding baseline dataset:

$$Relative\ representation = \frac{P_{spill}}{\overline{P_{base}}}$$

Sensitivity analyses proceeded via the following steps:

1. Starting with all taxa, proportional abundances were calculated for the spill dataset and baseline dataset, of which the latter were converted into a mean and 95% CI based on bootstrap resampling
2. Relative representation was calculated for each taxon as well as its 95% CI
3. Relative representation per taxa was coded based on whether the value indicated over/under-representation in the spill as follows
    a. -1 : mean RR < 1 & 95% CI did not overlap with 1, indicative of under-represented in the spill relative to the baseline
    b. 0: 95% CI for RR overlapped with 1, indicating that taxon proportional abundance was not significantly different in the baseline from the spill dataset
    c. +1 : mean RR > 1 & 95% CI did not overlap with 1, indicative of over-represented in the spill relative to the baseline

4. For each taxon, we calculated the distance to equivalence, D_{equiv}, as the perpendicular distance between the point defined by (P_{spill}, $\overline{P_{base}}$) to the line of equivalence (i.e., $y = x$),
5. We then removed the taxon with the highest current value of D_{equiv} reasoning that the results may be most influenced by taxa that have particularly extreme values for P_{spill} or $\overline{P_{base}}$ relative to each other
6. After taxon removal, we repeated steps 1-4 to create a different set of values for RR and examined how our results changed after the removal of potentially influential species. In particular, we noted the number of instances where taxa changed status from under-represented – indeterminate – over-represented relative to the previous step after a particular taxon was removed
7. This process was repeated up to and including the removal of five species.

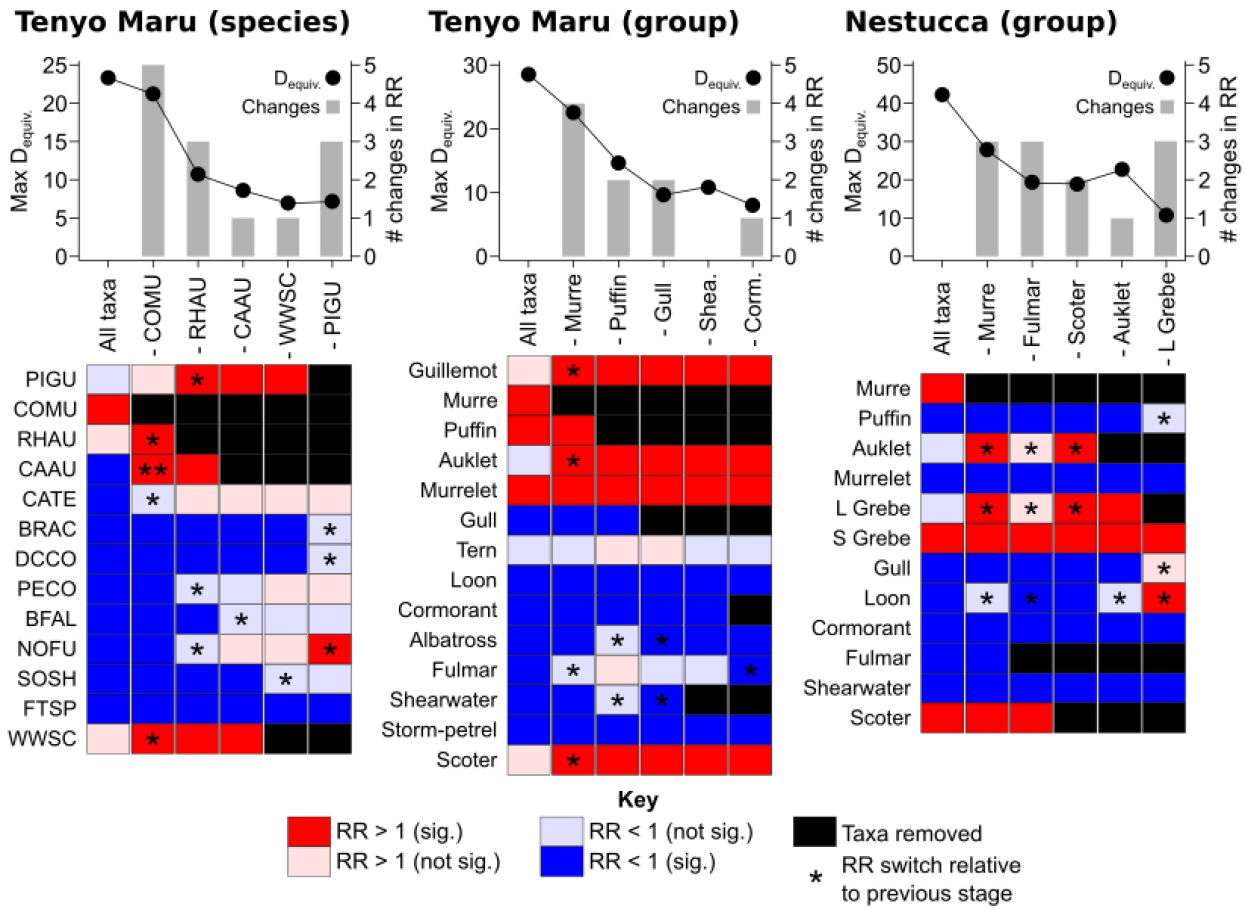


Figure S5. Summary of changes in relative representation when dominant taxa were removed. The upper panels show the maximum distance to equivalence at each stage, as well as the number of changes in status (either under-represented, indeterminate, or over-represented) following that taxon removal. The lower panels illustrate the status of each taxon at each stage, with changes highlighted by *.

8.1 Sensitivity analyses for Tenyo Maru analyses (species level)

The removal of common murres and rhinoceros auklets resulted in numerous changes (5, 3, respectively) to relative representation among remaining species (**Figure S5**). The subsequent removal of Cassin's auklets resulted in only one change to black-footed albatross from under-represented to indeterminate (**Figure S5**). Susceptibility metrics for up to three removals are shown in **Figure S6**.

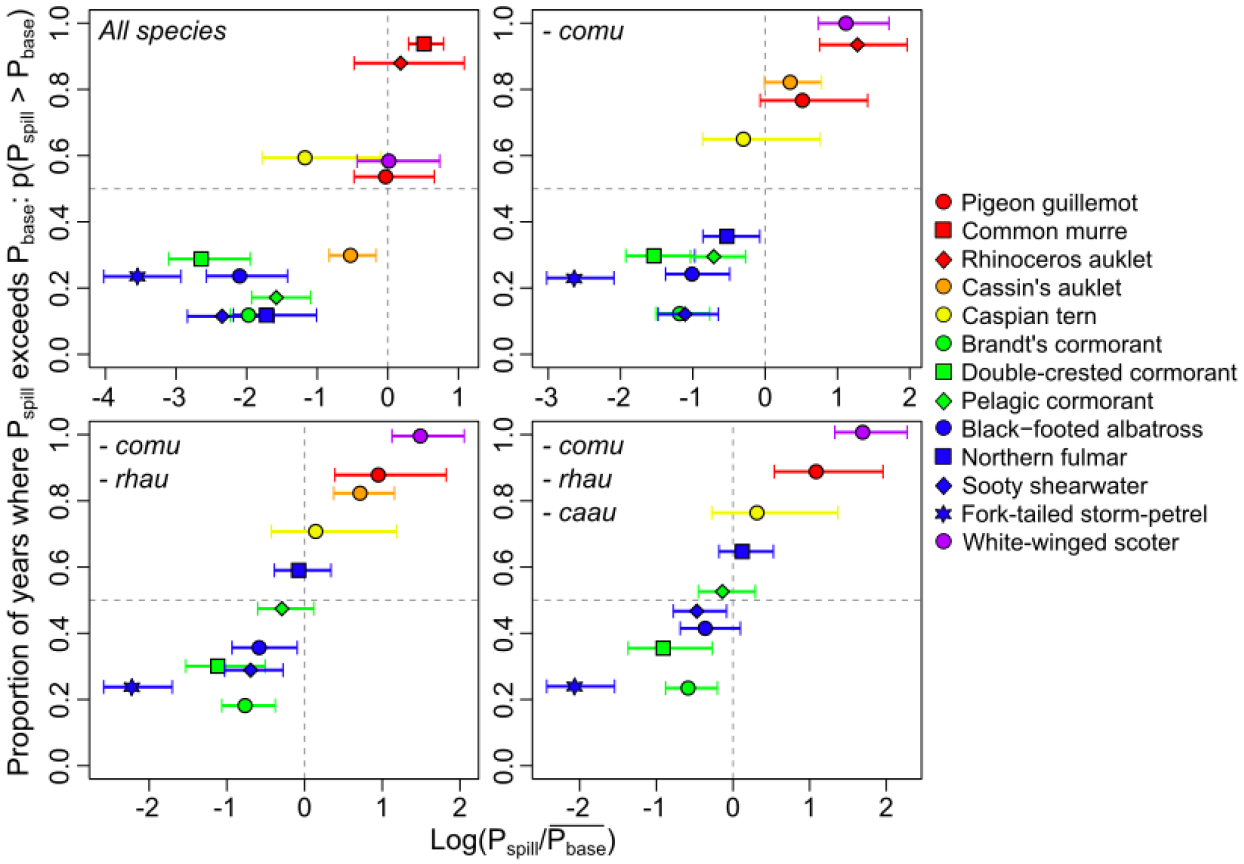


Figure S6. Bi-plots of average susceptibility ($log\left(\frac{P_{spill}}{\overline{P_{base}}}\right)$) and reliability ($p(P_{spill} > P_{base})$) metrics for the *Tenyo Maru* spill at the species level, with alternate panels showing metrics calculated after sequential taxon removals.

8.2 Sensitivity analyses for Tenyo Maru analyses (group level)

Removal of murres, puffins and gulls all resulted in changes to relative representation, albeit with the most changes (4) following the removal of murres (Figure S5). However, the subsequent removal of shearwaters resulted in no changes to relative representation among species. Susceptibility metrics for up to three removals are shown in **Figure S7**.

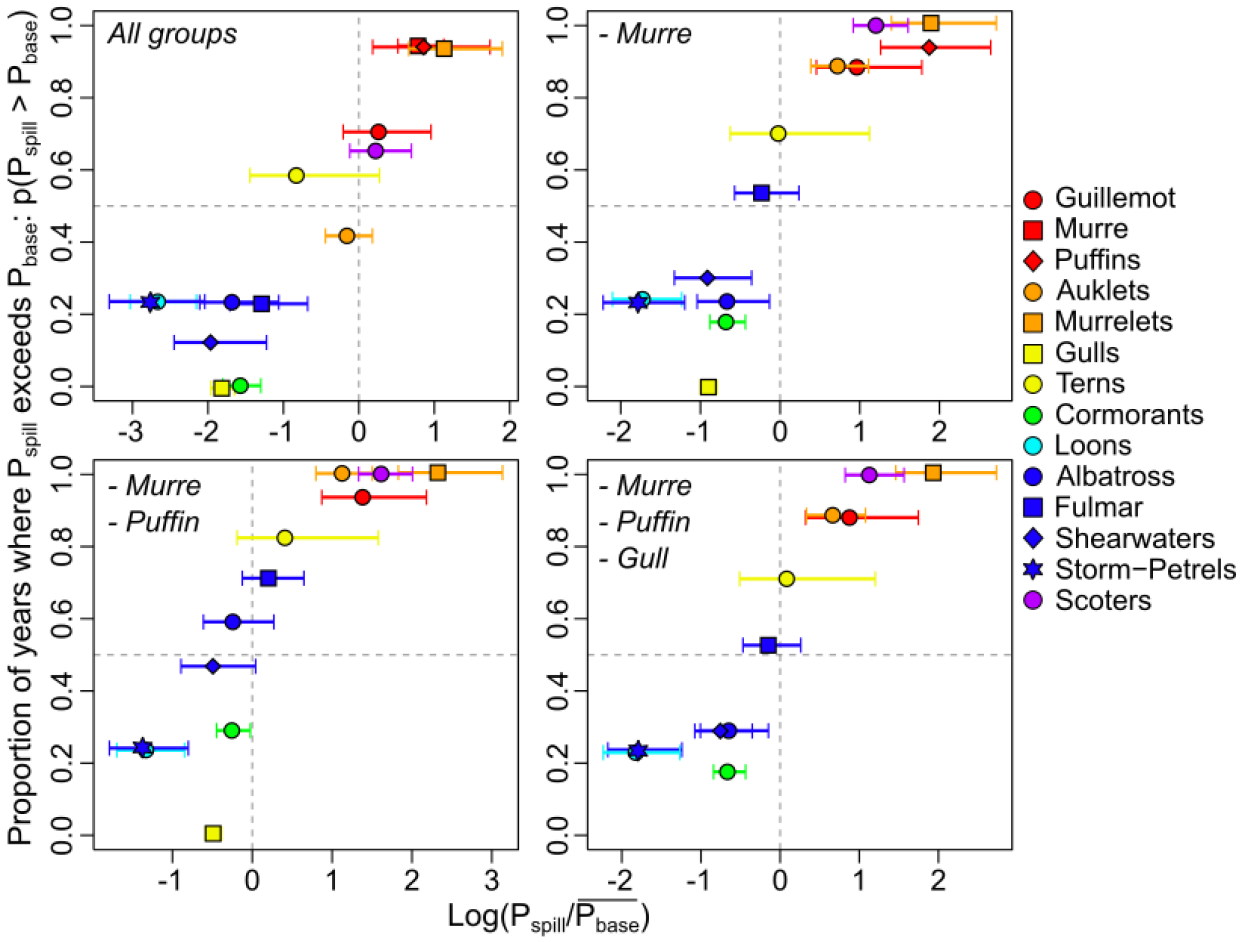


Figure S7. Bi-plots of average susceptibility ($log\left(\frac{P_{spill}}{P_{base}}\right)$) and reliability ($p(P_{spill} > P_{base})$) metrics for the *Tenyo Maru* spill at the group level, with alternate panels showing metrics calculated after sequential taxon removals.

8.3 Sensitivity analyses for Nestucca analyses

Removal of murres, fulmars, and scoters resulted in changes to relative representation for remaining species (**Figure S5**). However, the subsequent removal of auklets only resulted in a minor change in relative representation for loons (relative representation went from under-represented to indeterminate: **Figure S5**).

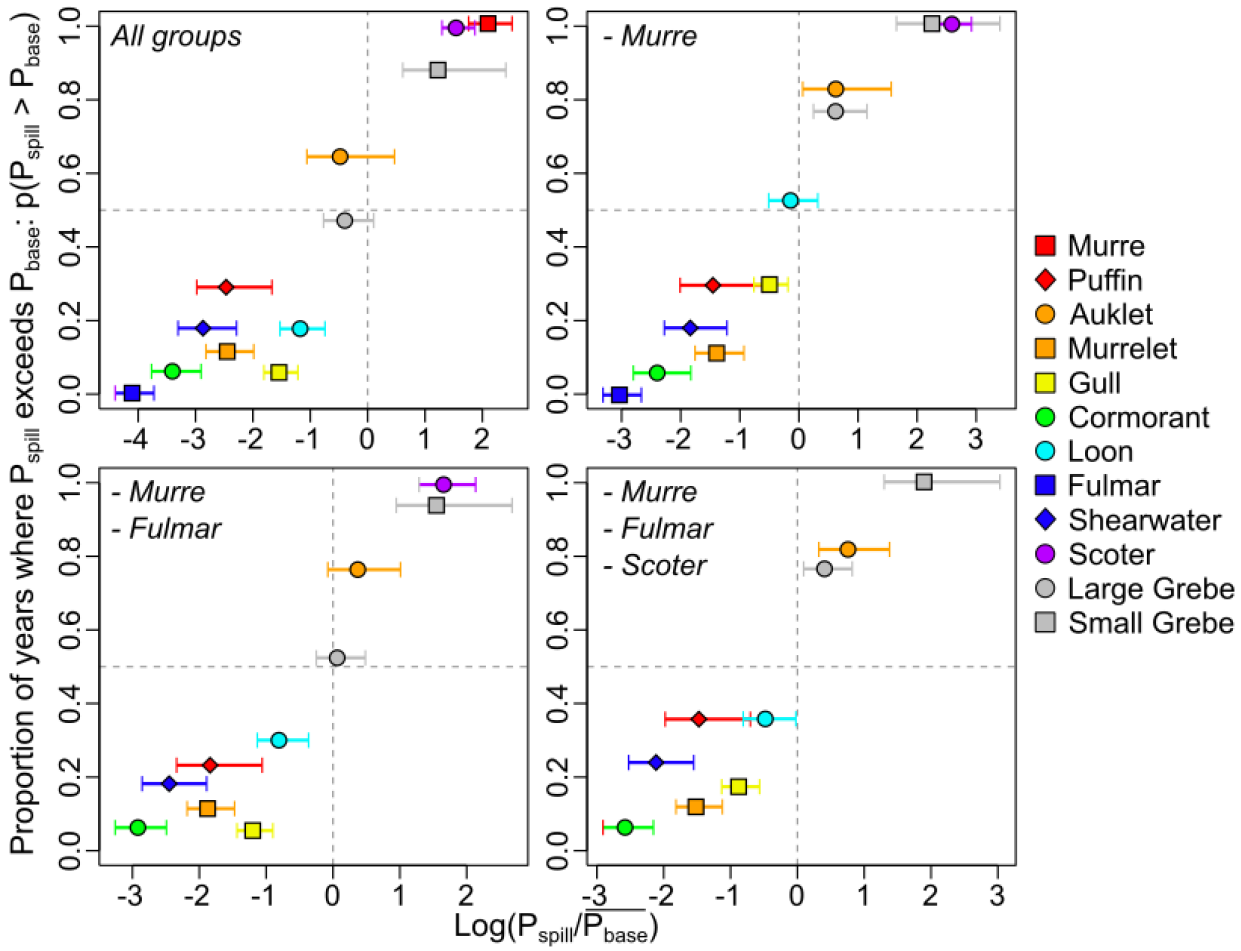


Figure S8. Bi-plots of average susceptibility ($log\left(\frac{P_{spill}}{P_{base}}\right)$) and reliability ($p(P_{spill} > P_{base})$) metrics for the *Nestucca* spill, with alternate panels showing metrics calculated after sequential taxa removals.

Section 9. Sensitivity of results to the inclusion of outlier years

Our hierarchical clustering identified the 2016 baseline year to be a potential outlier (i.e., singled out from other baseline years) in the *Tenyo Maru* group level analyses. We examined how much the inclusion of this year affected values of relative representation and reliability by calculating values for these metrics including 2016 and excluding 2016.

The only group whose values changed meaningfully was puffins, with an increase in relative representation and reliability (**Figure S9**), likely due to the abnormally high proportion of puffins observed in 2016 as a result of a rhinoceros auklet mortality event in that year (Knowles et al. 2019). However, the inclusion of the 2016 data does not affect our conclusions as in both cases (inclusive/exclusive of 2016) we would conclude that puffins were over-represented in the spill with a high degree of confidence.

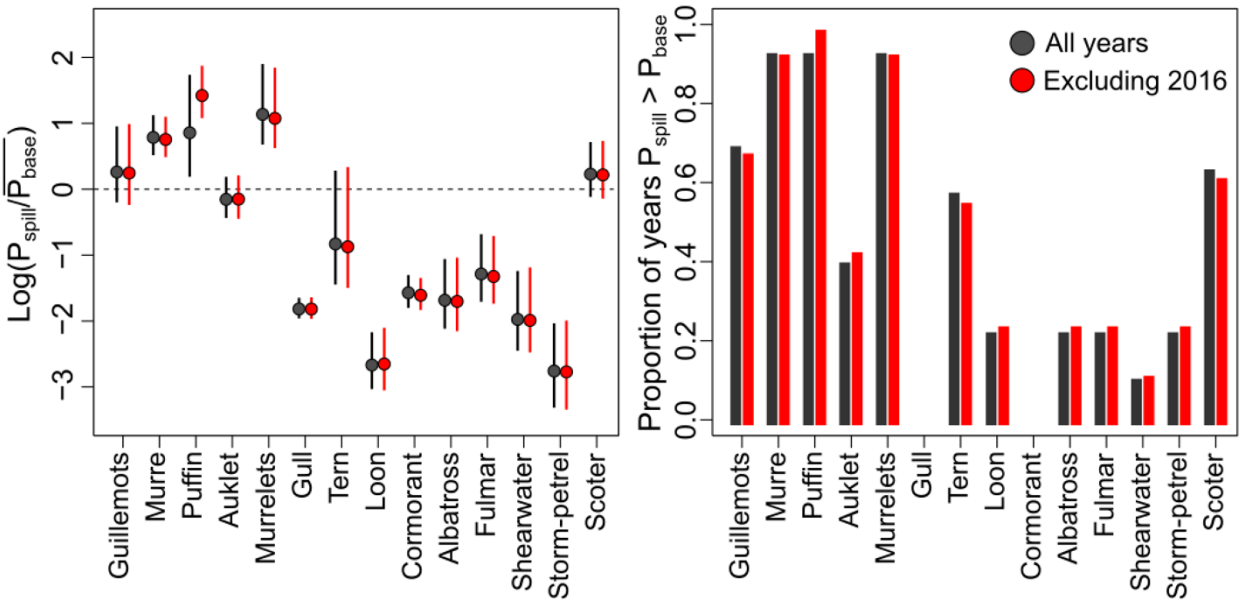


Figure S9. Average susceptibility ($log\left(\frac{P_{spill}}{P_{base}}\right)$) and reliability ($p(P_{spill} > P_{base})$) metrics for the *Tenyo Maru* spill (group), calculating including 2016, and excluding 2016.

References

Anderson, M.J., 2014. Permutational multivariate analysis of variance (PERMANOVA). In: Wiley Statsref: Statistics Reference Online. pp. 1–15.

Knowles, S., Bodenstein, B.L., Berlowski-Zier, B.M., Thomas, S.M., Pearson, S.F., Lorch, J.M., 2019. Detection of bisgaard taxon 40 in rhinoceros auklets (Cerorhinca monocerata) with pneumonia and septicemia from a mortality event in Washington, USA. J. Wildl. Dis. 55 (1), 246–249.

Chapter 2. Elevated seabird mortality associated with cumulative environmental and natural history stressors

ABSTRACT

Adult seabird mortality is influenced by a combination of stressors, from environmental (e.g., marine heat waves) to natural history (e.g., breeding, molting). We investigated how environmental and natural history stressors (specifically, molt) can combine to influence mortality in common murres (*Uria aalge*). To do this, we used data from a beached bird program in the northern California Current System (nCCS) over the years 2001-2022. To explore how molt varies across years, we first created models for each year of the proportion of dead birds found in molt over the day of the year. We then created models that paired the beached bird data with environmental data to investigate how various environmental stressors relate to the proportion of birds that die in molt during the molting period each year. We included both stressors that might impact mortality in the short-term, such as storms, and stressors that might impact mortality in the long-term, including spring transition date (the estimated date when the nCCS switches to predominantly "upwelling" or bringing nutrients to the ocean surface), nearshore foraging habitat quality, and the amount of ocean upwelling taking place prior to molt. We also included a measure of the timing of mortality: the mode in encounter rate (birds found per km surveyed). We found that the timing of adult mortality, spring transition date, and cold-water habitat availability were the strongest predictors of the proportion of birds that died in molt. However, the direction of effect of spring transition was opposite to expectation, plausibly due to delayed spring transition causing mortality prior to the molt period. This suggests that

environmental stressors like cold-water habitat availability can combine with molt to "push" molting birds over the edge to death. However, the timing of environmental stressors and mortality are key—in years with an early peak in death, the proportion of birds that died in molt was lower, likely because mortality occurred predominantly before birds molted. Both timing and intensity of environmental stressors appear to be crucial to molting murre mortality.

INTRODUCTION

As long-lived organisms spending most of their lives on the open ocean, marine birds are adapted to withstand variation in environmental conditions ranging from short-term, intense events such as storms, to large scale environmental forcing exemplified by decadal plus events such as the El Niño Southern Oscillation (Oro 2014). As a short-term stressor, storms can increase the energetic cost and strain of flying (Elliott *et al.* 2014, Shepard *et al.* 2019), impair thermoregulation (Grunst *et al.* 2023), and make it more difficult to find prey (Taylor 1983, Barreau *et al.* 2021, Clairbaux *et al.* 2021). This can lead to changes in population vital rates including decreased reproductive output (Aebishcer 1993, Descamps *et al.* 2015, Newell *et al.* 2015) and increased adult mortality (Anker-Nilssen *et al.* 2017, Tavares *et al.* 2020, Clairbaux *et al.* 2021).

In the longer term (weeks to months or longer), influences on seabirds can be more indirect, primarily through shifts in trophic energy pathways (Robinson et al. 2018, Osborne et al. 2020, Fayet et al. 2021). These can be regularly occurring annual events, such as upwelling bringing nutrient-rich water to the surface (e.g., Parrish & Zador 2003, Peery *et al.* 2009, Johns *et al.* 2022), or irregularly occurring ones, such as periods of high or low sea surface temperature

(SST) or air temperature relative to the climatological averages (e.g., Glencross *et al.* 2021, Tate *et al.* 2021, Schoen *et al.* 2022). High air temperature and/or SST has been associated with increased time spent behaviorally thermoregulating and less time spent at the nest site (Olin *et al.* 2023), as well as a reduction in high quality prey availability (Jones *et al.* 2018). Delayed upwelling in the northern California Current System (nCCS) has been linked to increased time spent foraging (Ronconi & Burger 2008). Long-term, indirect stressors like delayed upwelling or periods of extreme temperature can influence seabird demography through reduced reproductive output (Gjerdrum *et al.* 2003, Sydeman *et al.* 2006, Sandvik *et al.* 2007, Fromant *et al.* 2021) and increased adult mortality (Parrish *et al.* 2007, Jones *et al.* 2018, Piatt *et al.* 2020, Jones *et al.* 2023).

Environmental stress may have a greater impact on marine birds during times when they are already under natural history stress, such as breeding or migration. For example, Osborne *et al.* (2020) found that breeding black-legged kittiwakes (*Rissa tridactyla*) increased foraging range to adequately provision young during a marine heatwave. Migrating shearwaters (*Manx* and *Calonectris* spp.) have been observed responding to wind strength by taking longer but likely less turbulent migration routes (González-Solís *et al.* 2009).

Flight feather molt, or the systematic replacement of flight feathers, is another period of natural history stress for seabirds. Energetic costs have been observed to increase for molting birds (e.g., Dietz *et al.* 1992, Portugal *et al.* 2007, Cyr *et al.* 2008), due to the costs of regrowing feathers (Lindström *et al.* 1993) and/or decreased foraging ability from impaired movement (Bridge 2004). For birds in non-tropical areas, flight feather molt typically occurs immediately after breeding and migration, at a time when post-breeding adults are in relatively poor body condition

due to the stress of incubation and chick-rearing (Jenni & Winkler 2020). Molting typically occurs before the onset of winter (Jenni & Winkler 2020), potentially to ensure that feathers are in good condition before the onset of storms.

Synchronous molt, a type of flight feather molt in which all flight feathers are dropped over a short period of days, is found in a diverse subset of marine birds, including ducks (*Anatidae*) (Anderson *et al.* 2020), loons (*Gaviidae*) (Russell 2020), pelicans (*Pelicanidae*) (Shields 2020), alcids (*Alcidae*) (Bridge 2006), and diving petrels (*Procellariformes*) (Bridge 2006). Synchronous molt has been hypothesized by Guillemette *et al.* (2007) to be more costly per unit of molting time than a strategy of molting more slowly. For birds that undergo synchronous molt, it may be advantageous to minimize the time spent in molt by undergoing molt more quickly (Jenni & Winkler 2020), even if this method of molt is more costly per unit of time. In their study of synchronously molting common eiders (*Somateria mollissima*), Guillemette *et al.* (2007) found that daily metabolic rate increased by 9% overall during molt. Synchronously molting birds may be more heavily impacted by environmental stressors than non-synchronously molting birds (because of the heavy energetic costs of synchronous molt), making them particularly apt for investigating the impacts of molt as a natural history stressor.

We used a synchronously molting seabird species, common murres (*Uria aalge*, hereafter murres) to explore whether and how molt might interact with environmental conditions known to negatively impact marine birds in temperate eastern boundary current systems. Adult murres replace flight feathers annually in late summer or early fall following breeding (Thompson *et al.* 1998). With ~2.8 million breeding individuals, murres are an abundant, deep-diving, piscivorous Alcid found year-round along the western coast of the United States (Ainley *et al.* 2021).

Substantial mortality events of adult murres have been linked to a range of environmental stressors, including storms (Bailey & Davenport 1972) and high SST (Piatt *et al.* 2020, Romano *et al.* 2020).

Both short-term storm events, as well as departures from long-term climatology mediated through changes in physical conditions murres experience, appear to exacerbate the chance that molting murres will suffer energetic and even mortal consequences. Holdgate (1971) attributed a die-off of 12,000 seabirds in the Irish Sea, mostly murres, to interactions between periods of high wind, disease, and molt. During El Niño years, male murres have been observed molting three times slower than during non- El Niño years (Thompson 1998). Following a mass mortality event of murres in California in 2015 associated with the northeast Pacific marine heatwave, Gibble *et al.* (2018) found that five out of eight murres they necropsied were in molt. These findings suggest that environmental stressors may interact with molt to push molting murres "over the edge" to death at a higher rate than non-molting murres.

In this paper, we investigated the effects of environmental stressors and molt on murres in the nCCS over the years 2001-2022. Because little is known about year-over-year variation in molt phenology, we first examined the degree to which the annual molt cycle was conserved. We then addressed the hypothesis that in more environmentally stressful years (i.e., in which upwelling is weak or late, and SST and storminess are high), a higher proportion of beached murres found during the molting period would be in molt because of the cumulative impacts of multiple stressors (i.e., a bird may be more likely to die if they are impacted by both the stress of molt and the environment than the environment alone).

METHODS

Beached Bird Data

We used data from beached bird surveys to explore the interactions between mortality, molt, and environmental stressors. Beached bird surveys have helped to forge a more comprehensive (over space and time) dataset on seabirds, by collecting information such as measurements, sex, age, and/or body condition (Wiese & Elmslie 2006) for individual carcasses. This data has been used to explore both the impact of short-term environmental events (e.g., oil spills- Camphuysen & Heubeck 2001, harmful algal blooms- Jones *et al.* 2017, marine heatwaves- Jones *et al.* 2023) and long-term stressors (e.g., chronic oiling- Roletto *et al.* 2003, fishery bycatch- Hamel *et al.* 2009, marine plastic - Baes *et al.* 2024).

Since 1999, the Coastal Observation and Seabird Survey Team (COASST), a citizen science program at the University of Washington, has conducted beached bird surveys. Coverage has expanded from the southern outer coast of Washington state to ~450 sites throughout the nCCS and Alaska. At present, there are ~800 active participants collecting data. COASST participants are trained to conduct effort-standardized monthly surveys (with no requirement to survey on a particular day within participants or across beaches), collecting data on all bird carcasses found. Carcasses are identified to the lowest possible taxa and plumage using a field key (Parrish *et al.* 2017). Participants send photographs, foot type, measurements (wing chord, tarsus, and culmen), and field identification to the lowest possible taxon to COASST. Expert verifiers use these data to independently determine species identity.

We extracted COASST data for carcasses verified as murres from 2001 through 2022 (22 years), as earlier years had fewer than 30 murres each (1999: n = 7, 2000: n = 29). In this study, we included carcasses found on the outer coast of Washington south to Humboldt County, California (approximately 40.5 N to 48.5° N, Figure 1). This corresponds to the five COASST "regions" of Northern and Southern Washington, Northern and Southern Oregon, and Humboldt.

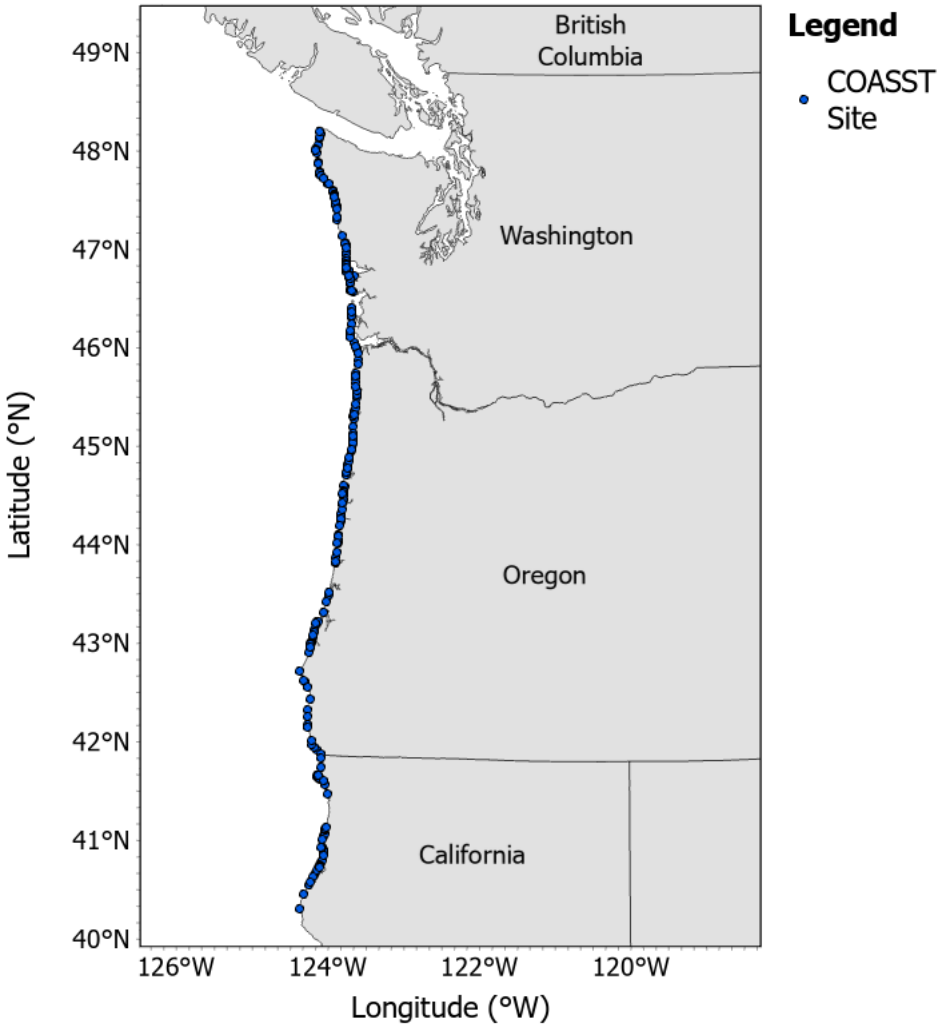


Figure 1: Location of COASST monitoring sites included in this analysis.

We restricted our dataset to carcasses that were either classified as adults during verification (based on morphometric and/or plumage characteristics; n = 10,654) or were unknown age-status following verification (n = 2,739) but had a bill length of ≥ 36 mm (n = 262) as a proxy for adulthood. This measurement threshold was determined based on verified adults with bill

measurements in the COASST dataset (99.1% of which had a bill length ≥ 36 mm). Thus, our total sample size of (presumed) adult murre carcasses was n = 10,916.

Carcasses were classified as either molting or non-molting based on verifier notes and/or wing chord measurement as follows. Adult murres tagged by the verifier as molting (n = 241) or not molting (n = 2,169) were classified accordingly. Birds without a wing chord measurement and with no molt status from the verifier (n = 1,204) were dropped from the analysis. Of the remaining carcasses (n = 7,302), those with a wing chord of ≥ 17 cm were classified as not molting. We chose this measurement threshold by plotting the verifier molt status and wing chord length for all birds with both a verifier molt status and a wing chord length (n = 2,143) (SM Figure 1). We searched for a wing chord length threshold that would maximize agreement for molt status between the verifier molt status and molt status based on wing chord length. At a threshold of 17 cm, agreement for molt status reached 98.8% (SM Table 1). Based on this threshold, n = 811 additional birds were classified as molting, and n = 6,491 were classified as not molting. Out of the 1,052 carcasses classified as in molt overall, 97.9% were found between August and November, which we refer to as the "molt window".

Molt Phenology

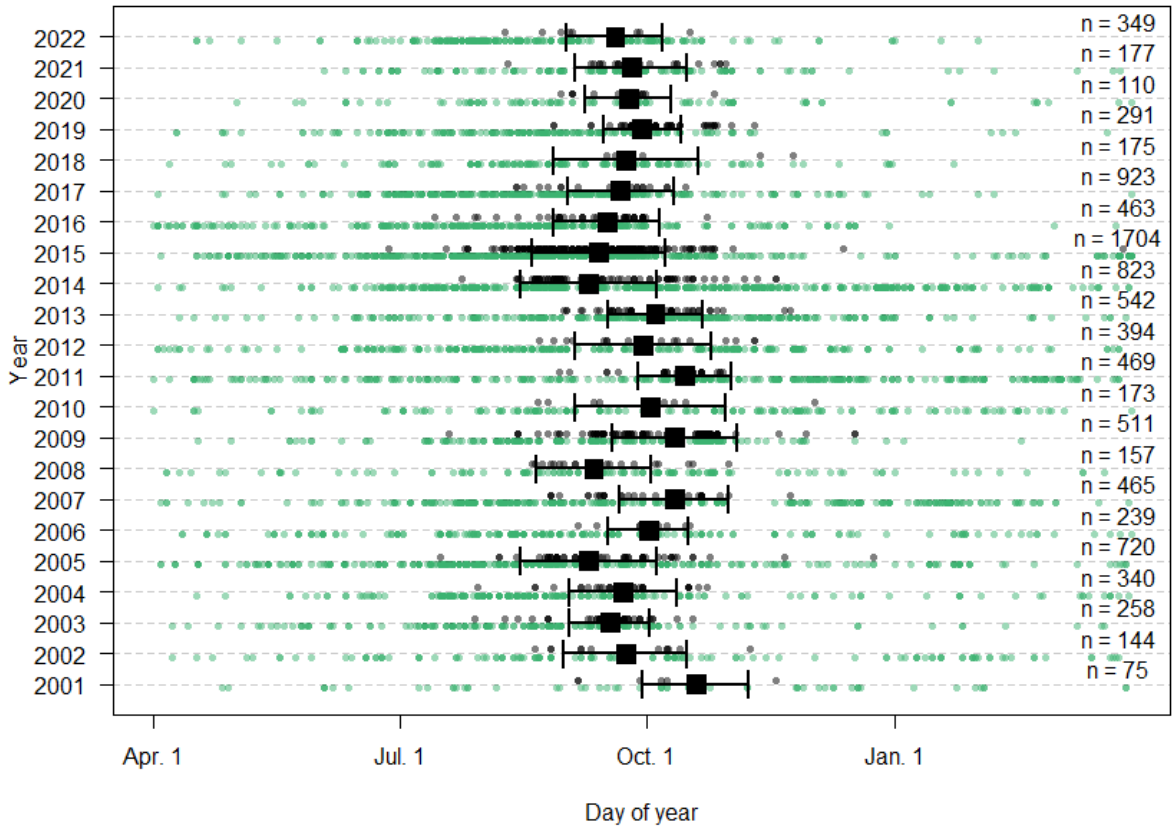
To explore patterns in the annual timing and intensity of molt, we modelled the proportion of birds in molt as a non-linear function of day of year for each year in the dataset (n = 22). We did so via fitting Binomial (logit link) Generalized Additive Mixed Models (GAMMs) with day of year, *d*, as the independent variable and number of carcasses in molt, *m*, as the dependent variable, modelled as a binomial variable according to the number of carcasses found on a particular survey, *n*.

$$logit(\widehat{p_i}) = \beta_0 + s(d_i) + e_i$$

$$m_i \sim B(n_i, \widehat{p_i})$$

where $\widehat{p_i}$ is the estimated proportion of birds in molt on survey i, β_0 is the intercept, e_i is the random effect for survey i, and $s(d_i)$ is a smooth function of the Julian day of year upon which the survey was performed. GAMMs were used to allow for (assumed) non-linearity in the relationship between molt proportion and day of year.

To account for the annual nature of the data, we used cyclic cubic regression splines (Wood 2017). We included a random effect of survey to account for overdispersion relative to the binomial distribution that arose due to similarity (i.e., likely non-independence) in molt status of carcasses found on the same survey. To ensure that molt proportions approached zero at the "beginning" and "ending" of each year (days 1 and 365) to match the phenology of molting (Figure 2), we renumbered days such that day 1 equaled April 1st. Models were fitted, including smoothing parameter optimization, using restricted maximum likelihood to increase mean-square error performance while minimizing undersmoothing (Wood 2011).



Legend

● = Non-molt

● = Molt

■ = Molt med.

⊢ = Molt sd

Figure 2: Individual birds identified as in molt or not in molt ("Non-molt") across each year and day of year when the bird was found (where a "molt year" extends from Apr. 1 until 31 Mar of the following year). Sample sizes for each year are displayed ("n =" value). For each year, the median find day of the year for molting birds ("Molt med.") and standard deviation ("Molt sd") is also displayed. Values for individual birds are semi-transparent, more saturation indicates more individuals were found on that day.

For each year in our dataset, we extracted features from the fitted model representing the timing or intensity of molt:

- peak height (maximum proportion of carcasses found in molt),
- peak location (day of year when the molt proportion was highest),
- onset (first day of the year on which the proportion of carcasses in molt exceeded 50% of the year's peak height).

We initially considered duration (total days elapsed where molt proportion exceeded 50% of the year's peak height) as an additional metric. However, low carcass sample size (both molting and non-molting) at the end of each calendar year (across all years, <5% of murres were found in December) increased model uncertainty and led to unreliable estimates of duration.

To quantify model feature uncertainty, we used simulations from the GAMM posterior covariance matrix and basis function set (Simpson 2023). We simulated alternate realizations of (smooth) model parameters from a multivariate normal distribution, with mean vector and variance-covariance matrix as estimated for the fitted model. Each simulated set of model parameters provided a 'smooth' non-linear function that was plausible given the data and uncertainty surrounding the fitted function. For each generated non-linear function (n = 5000), we extracted model features as defined above, resulting in a distribution of values for each. We subsequently calculated the mean and 95% confidence interval for each feature as the mean and 2.5/97.5 percentiles of the resulting distributions for each feature. While this approach provides an approximation of the 95% confidence interval, due to smaller sample sizes outside of the core molt season, the simulation approach also resulted in ~25% (n = 1,235) of simulated non-linear functions where the peak in molt proportion occurred prior to/after the first/last observations of

molting birds within a given year (i.e., the model was not sufficiently constrained by the data). As a result, when processing simulated values, we excluded those simulations where the peak value occurred outside of the interval when molting birds were observed that year to ensure that resulting values were representative of the data.

Environmental Stressor Analyses

To test our hypothesis about the relationship between environmental variables and the proportion of birds that die in molt during the molt window, we used Generalized Linear Mixed Models (GLMMs). While the GAMMs were useful for exploring the general shape of molt from year to year, it would be ill-advised to input the point estimates from the GAMMS as dependent variables in GLMMs without accounting for variation, which would be computationally difficult. Therefore, we returned to the molt window and extracted a more direct measure of molt intensity from the raw data to use as our dependent variable. This was the proportion of all carcasses found in our molt window (August-November) that were in molt ("Prop. Molt"). Prop. Molt is strongly, positively correlated with our GAMM variable peak height, suggesting it is an adequate proxy ($R^2 = 0.69$; SM Figure 2).

In the Pacific Northwest, seasonal upwelling sets up the quantity and quality of prey available to shelf-associated species (Bograd *et al.* 2009, Checkley & Barth 2009). We retrieved upwelling data from the National Oceanic and Atmospheric Association Southwest Fisheries Science Center (https://oceanview.pfeg.noaa.gov/products/upwelling/intro). Specifically, we used the Biologically Effective Upwelling Transport Index (or "BEUTI"), which estimates the amount of nitrate moving into the surface mixed layer (in $\mu mol\ m^{-1}\ s^{-1}$) using the product of the estimated vertical upwelling strength and the nitrate concentration at the bottom of the mixed layer (Jacox

et al. 2018). Jacox *et al.* (2018) designed BEUTI to be a more meaningful measure of productivity than upwelling indices that do not take nutrient movement into account. We included data from 42 to 48°N, as this spatial extent most closely approximates the COASST data (40 N to 48.5°N); and calculated cumulative (summed) upwelling over May through August. This time window encapsulates upwelling during the breeding season that immediately precedes the start of molt.

The phenology of seasonal upwelling in the Pacific Northwest is marked by a transition from predominantly downwelling to predominantly upwelling. Known as the "spring transition", this physical shift usually occurs between early March and early June (NOAA Fisheries 2024) and is marked by increased cool, nutrient-rich water at the ocean surface. Delayed upwelling (i.e., later spring transition) has been associated with population changes in shelf-associated upper trophic organisms (e.g., decreased survival in Oregon coho (*Oncorhynchus kisutch*)- Logerwell *et al.* 2003, decreased catch rate of nekton in the nCCS- Brodeur *et al.* 2006), and subsequent increases in beached bird deposition, including murres (Parrish *et al.* 2007). We retrieved spring transition date ("ST Date"), or the estimated calendar date on which spring transition occurs, from Columbia Basin Research (https://www.cbr.washington.edu/status/trans). We used ST Date calculated with the "CBR Mean" method, as this was the only method for which a value was available for all years in our study (2001-2022). CBR Mean calculates daily smoothed average upwelling deviation indices across three sites along the Pacific Northwest coast (42°N 125°W, 45°N 125°W, and 48°N 125°W), and takes the minima of these daily values in spring as the Julian day corresponding to the ST Date (Van Holmes 2007).

Sea surface temperature (SST) has been linked to decreases in seabird prey quantity and/or quality (e.g., Schrimpf *et al.* 2012, Jones *et al.* 2018) as well as seabird mortality (e.g., Jones *et al.* 2023). Therefore, we obtained daily SST anomaly (SSTa) maps data from the National Oceanic and Atmospheric Administration OI V2 high-resolution data repository (0.25° resolution, https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html) for the nearshore (within 150 km of the coast) waters within our study area (~40.5 N to 48.5° N). We used the SST data to calculate the Habitat Compression Index (HCI; Schroeder *et al.* 2022) to explore cold-water habitat available to murres. HCI was calculated as the mean proportional area where SST was >1° C above the 30-year climatology (1982-2011); therefore a higher HCI would have less cold-water habitat available. Because we hypothesized HCI to have a long-term, indirect impact on murres, we considered both the molt months and the months leading up to molt. We therefore calculated two versions of HCI: over the months leading up to molt (May-August), and during the molt months (August-November).

High storminess (i.e., periods of strong wind and potentially precipitation) could increase seabird mortality via impairment of the ability to fly, dive, and forage (Fort *et al.* 2009), and has been linked to increased Cassin's auklet (*Ptychoramphus aleuticus*) carcass deposition in the NE Pacific (Jones *et al.* 2018). As a proxy for storminess, we used significant wave height (H_{sig}) data, calculated as the mean wave height of the highest third of waves in a particular area, as these are thought to have the most "significant" impact (e.g., cause the most erosion) on the nearby coast (National Oceanic and Atmospheric Administration 2016). We retrieved H_{sig} data from the National Data Buoy Center (www.ndbc.noaa.gov/) for buoys closest to the geographic extent of the COASST dataset, and which had data available throughout our study years: Cape

Flattery, WA (Station 46087: 48°29'36" N 124°43'38" W) and Columbia River Bar (Station 46029: 46°9'48" N 124°29'12" W). Following the methods of Jones *et al.* (2017), we created an average index of H_{sig} across both buoys. We calculated storminess for two time periods: inclusive of the molt months (August-November) and lagged one month prior to the molt month time window (July-October).

To investigate the degree to which our environmental variables were related to each other, we examined cross correlations. Not surprisingly, different time windows for the same index were correlated (between HCI May-August and August-November, $R^2 = 0.51$; for storminess July-October and August-November, $R^2 = 0.72$, SM Figure 3). Only three other cross-correlations showed any strength. Upwelling was correlated with HCI May-August ($R^2 = 0.14$), storminess July-October ($R^2 = 0.28$), and storminess August-November ($R^2 = 0.25$). All other variable pairs had R^2 values equal to or less than 10%, suggesting very weak to no relationship. Overall, this suggests that the environmental variables we selected were relatively independent of each other.

Because timing of mortality might influence the proportion of birds that die in molt (e.g., if conditions are stressful enough early in the year, birds might die before ever reaching molt), we included in our models an additional variable related to encounter rate (ER- the number of birds encountered per kilometer of beach surveyed). We used ER Mode, or the mode in day of year when birds were encountered, weighted by ER of surveys per day. We also considered ER of adult common murres ("ER adult") but chose ER Mode as it had a higher, albeit weak, correlation with proportion of birds that die in molt (for ER Mode, $R^2 = 0.09$, for ER adult, $R^2 = 0.00$).

We created GLMMs containing scaled values representing beaching phenology and the environmental stressors as independent variables, and Prop. Molt as the dependent variable. To explore which of our variably lagged predictor versions was most appropriate, we created four "model sets" containing: ER Mode, ST Date, upwelling, one of the two HCI timeframes (either May-August or August-November) and one of the two storminess timeframes (either August-November or July-October).

Our response data was the number of birds, Y, identified as being in molt out of the total sample, n, within each survey. Therefore, we used binomial GLMMs with a logit link function, such that fixed effects modelled changes in the proportion of molting birds, p. To account for variation in the molt proportion among years, surveys, and COASST regions, we added random effects for each of these variables.

The models took the form of:

$$Y_i \sim B(n_i, p_i)$$

$$logit(p_i) = \beta_0 + \beta_w w_i + \beta_u u_i + \beta_s s_i + \beta_h h_i + \beta_b b_i + \alpha_{1,y_i} + \alpha_{2,r_i} + \alpha_{3,i}$$

Where subscript i denotes survey, w, u, s, h, and e indicate covariate values for wave height, upwelling, spring-transition date, HCI, and beaching phenology, respectively, β represents model coefficients for each of these covariates, plus the intercept (β_0), and $\alpha_{1\text{-}3}$ are random effects for year (y_i), region (r_i) and survey (i), respectively. Random effects were assumed to follow a normal distribution on the scale of the link function:

$$\alpha_1 \sim N(0, \sigma_y)$$

We evaluated models by first picking the top model set, and then evaluating the models within that set. Because n is small (at 22 years), we used AICc to select the top model set and the top model (Hurvich & Tsai 1989). For model sets, AICc was summed across all models within the set to choose the top model set with the lowest AICc. AICc weights (a) were used to find the conditional probability of each model being the best model within the set (Wagenmakers & Farrell 2004). We also calculated the weight of each predictor across all sets by summing the weights across all models containing it. We then used these values to calculate parameter statistics, including the weight of all models including a specific parameter (Σa), the model-averaged regression coefficient, ($\bar{\beta} = \Sigma_i a_i \beta_i / \Sigma_i a_i$), and the 95% confidence interval of regression coefficients. We conducted model validation to check model fit and performance following the methods of Zuur (2009). We plotted the best model using the jtools package (Long 2022).

RESULTS

Phenology of murre mortality and molt

Across the years included in our study (2001-2022), the majority of adult murre carcasses were found between the months of July and October (79% of adult murres, n = 9,502; Figure 2). The modal date in encounter rate, or ER Mode, fell between Julian days 235 and 282 across all years (a 47-day spread, equivalent to 23 August- 8 October); however, 50% of the modal points were within a much narrower window of only 24 days centered on Julian day 257 (~14 September). In two of the sampled years, carcass deposition started earlier than usual, with >25% of birds dying prior to July (in 2002: 36.8%, in 2011: 28.1%).

Carcasses in molt were encountered about a month after the onset of the July-October mortality window (Figure 2). In July, only 1.0% of carcasses found were in molt, whereas by August, just over 7.0% of carcasses were molting birds. The median in molting carcass find date ranged from molt day 163 (in 2005 and 2014) to day 203 (in 2001) or 10 September to 20 October, for a range of 40 days. However, the proportion of birds in molt each year varied by an order of magnitude, ranging from a low of 2.3% in 2017 (n = 923 birds overall) to a high of 26.2% in 2009 (n = 511 birds overall).

Molt Phenology Models

We used GAMMs to understand the variability of molt phenology and intensity from year to year. Although there was variation in the shape of the curves describing the proportion of carcasses in molt, all years were strongly unimodal, with the possible exception of 2018 (SM Figure 6). For all years, it was possible to extract peak features (Figure 3). Not surprisingly, all years had a data shortage at the "tails" as murres had either mostly not begun molt yet (early) or had mostly completed molt (late). This caused high uncertainty relative to the middle of the GAMMs (SM Figure 6).

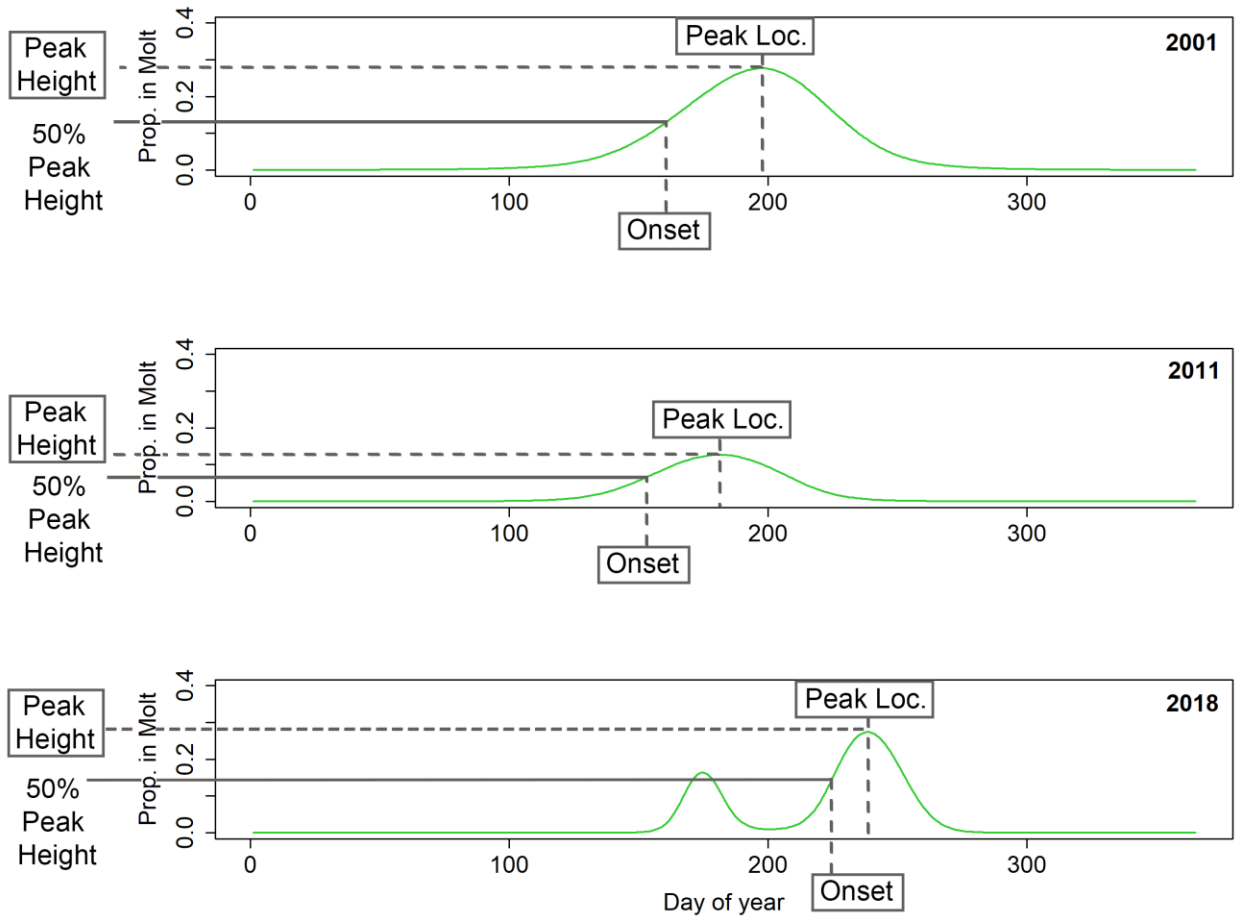


Figure 3: GAMM point estimates for three example years (2001, 2011, and 2018). A graphical depiction of peak feature extraction is shown for peak height, peak location ("Peak Loc."), and onset. For day of year, day 1 = April 1st.

There was a high degree of variability in peak height (Figure 4a), or the highest estimated proportion of carcasses found in molt in each year, with a 53- percentage point spread between 2019 (0.61) and 2017 (0.08). Peak location was also variable, with an 81-day spread between the earliest (2008; molt day 158 or 4 September) and latest (2018; molt day 239 or 24 November) estimates (Figure 4b). However, 50% of the modal points were constrained to only 12 days centered on molt day 181 (28 September). Both variables had 2-3 years with relatively high

uncertainty, with confidence intervals stretching ~60 percentage points for peak height in the most variable years (for 2001: 58.9, for 2018: 57.1) and ~60-days for peak location in the most variable years (for 2001: 69.5, for 2002: 64.1, for 2018: 74.2). As 2001 is the year with the lowest sample size (at n =75), the high uncertainty in this year is likely due, at least in part, to this.

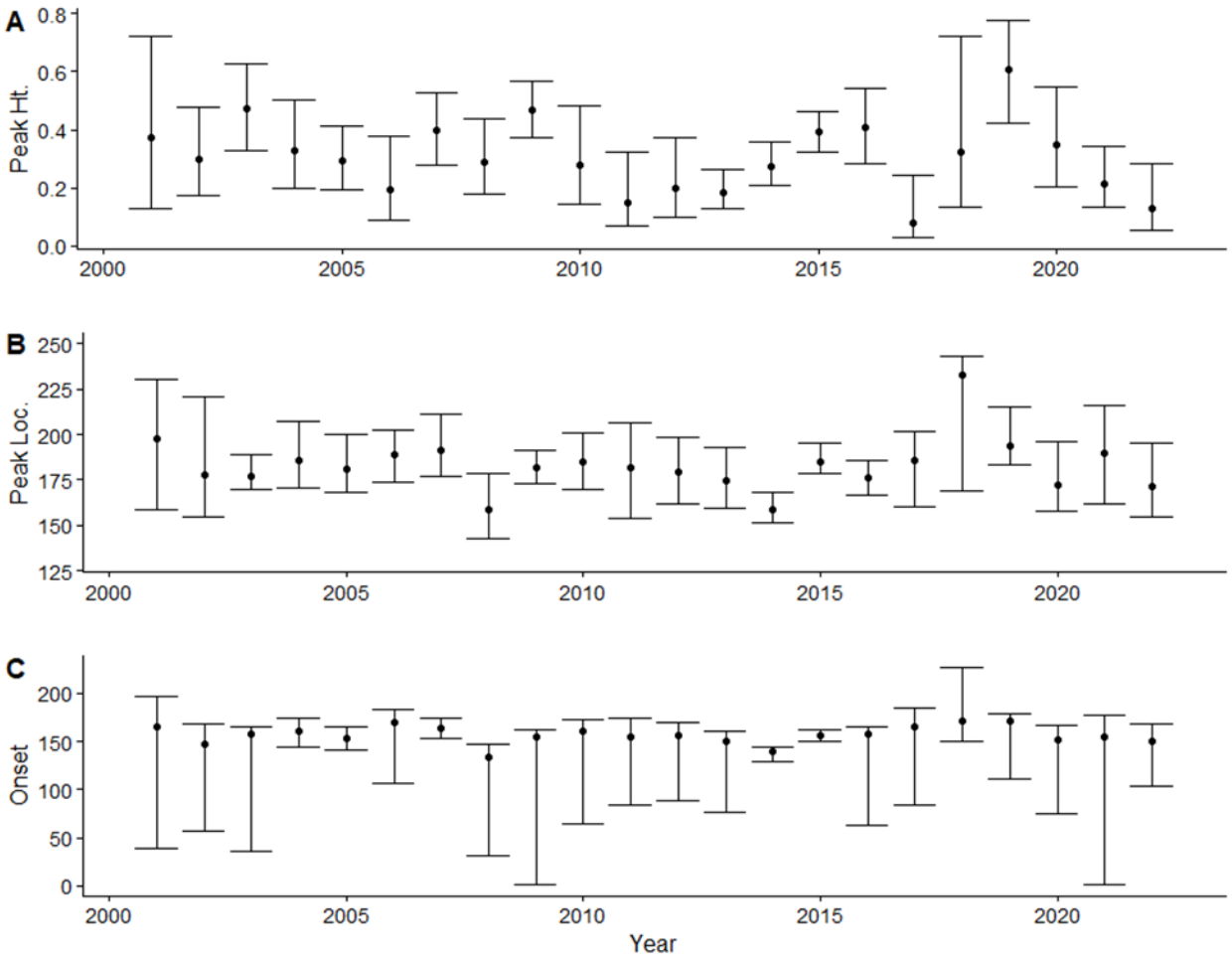


Figure 4: The point estimate and 95% confidence interval for each feature extracted from the GAMMs (peak height, peak location, and onset) from 2001-2022. For peak location and onset, day 1 corresponds to April 1st.

By contrast, onset was much more constrained, with only a 38-day range among years with the earliest (2008; molt day 134 or 12 August) and latest (2019, molt day 172 or 19 September) point estimates (Figure 4c). 50% of modal points fell within a 12-day range centered on molt day 156 (corresponding to 2 September). Unlike the other two peak features, onset often exhibited asymmetrically high uncertainty (Figure 4c) in the early season, likely because of low sample

size. Even though we attempted to control for de facto zeroes defining the transition from no molting to molting by defining onset as the first day of the year on which the proportion of carcasses in molt exceeded 50% of the year's peak height, there were still far fewer murres in molt at the onset relative to the peak. However, this also suggests that murres have some individual variation in when they begin to molt.

There did not appear to be any relationship between the intensity of the molt signal (i.e., peak height) and its timing, either peak location ($R^2 = 0.03$; SM Fig 7) or onset ($R^2 = 0.06$). However, the timing metrics (peak location and onset) were strongly positively correlated ($R^2 = 0.65$), suggesting that the shape of the annual frequency distribution of molt is conserved across years. In sum, while the correlation between onset and peak location suggests that once the population starts to molt that physiological event is fairly constrained in time, there is not necessarily any relationship between when molt starts (or peaks) and the tendency for adult murres to die while molting.

Environmental Stressor Analyses

We used GLMMs to explore the relationship between the proportion of carcasses found in molt during the molt window (August-November) and a suite of environmental stressors proxying the quality of the foraging habitat as well as shorter-term stressors such as storms. To account for impacts of year-to-year changes in mortality phenology, we also included the peak in carcass encounter rate in our models.

To build our models, we first examined the relative fits of environmental variables gathered over lagged (pre-molt) or non-lagged (during molt) time windows (for HCI and storminess) across

four model sets composed of cumulative upwelling, HCI, storminess, and ST Date, along with the phenology variable of ER mode. We used summed AICc across all models within each set to attempt to select the set that explains the most variation in Prop. Molt using the fewest number of variables. The model set with May-August HCI and August-to-November storminess had the highest support/lowest AICc (SM Table 2). Our final variable set therefore included: ER Mode, May-August HCI, August-to-November storminess, ST Date, and cumulative upwelling. To determine whether unusual years could be driving our results, we explored removal of one year with a low sample size (2001, at n = 75) and two years that had known mass mortality events in our system: 2009, when a harmful algal bloom led to a mass mortality event which included murres (Jones *et al.* 2017), and 2015, when the onset of the northeast Pacific marine heatwave led to mass starvation of Cassin's Auklets (Jones *et al.* 2018). However, removal of these potentially impactful years changed our estimates of model-averaged β by less than 0.1 (SM Tables 3-6), so we chose to keep all years in our GLMMs.

From our chosen model set we calculated variable-specific importance as the sum of the Akaike weight of all models within the set that included the specific parameter (Σa). Three of our variables were consistently represented in top models and had a relatively high impact on the proportion of adult murre carcasses found in molt (Table 1): modal day of encounter rate of murre carcasses (ER Mode); foraging habitat quality during and following the breeding season, proxied by the mean proportional area where SST was >1° C above the 30-year climatology during the breeding season, which we refer to as habitat compression index (HCI May-August); and spring transition date (ST Date), or the estimated Julian day when the spring minima of smoothed cumulative upwelling deviation occurs (Van Holmes 2007).

Table 1: Regression coefficients, model statistics, and parameter statistics for the Generalized Linear Mixed Models for the relationship between proportion of dead murres in molt per year and environmental variables. Model statistics include AICc, ΔAICc = AICc - min(AICc), and model Akaike weight (*a*). Parameter statistics include summed Akaike weight of all models featuring that parameter (Σ*a*), model-averaged regression coefficients ($\bar{\beta} = \Sigma_i a_i \beta_i / \Sigma_i a_i$), and model-averaged confidence intervals (model-averaged CI). Int = intercept, AN Storm = August-November Storm, ER = mode day of year on which birds were found, MA HCI = May-August Heat Compression Index, ST Date = Spring transition date, and Upwell = May-August cumulative upwelling. Models are listed in descending order of weight. Only models with weights >= 0.01 are shown.

Regression Coefficients							**Model Statistics**		
Predictors included	Int	AN Storm	ER	MA HCI	ST Date	Upwell	AICc	Δ AICc	Weight
ER, ST Date, MA HCI	-2.10	NA	0.24	0.30	-0.31	NA	3400.77	0.00	0.30
ER, AN Storm, Upwell, ST Date, MA HCI	-2.10	0.03	0.12	0.17	-0.28	0.00	3401.49	0.72	0.21
ER, ST Date, Upwell, MA HCI	-2.10	NA	0.24	0.32	-0.29	0.07	3402.12	1.34	0.15
ER, ST Date, AN Storm, MA HCI	-2.11	0.06	0.26	0.31	-0.30	NA	3402.33	1.55	0.14
ST Date, MA HCI	-2.12	NA	NA	0.20	-0.35	NA	3404.04	3.26	0.06
ST Date, Upwell, MA HCI	-2.12	NA	NA	0.22	-0.33	0.07	3405.51	4.74	0.03
ER, AN Storm, Upwell, MA HCI	-2.10	0.27	0.36	0.45	NA	0.28	3406.01	5.23	0.02
ST Date, AN Storm, MA HCI	-2.12	0.00	NA	0.20	-0.35	NA	3406.05	5.27	0.02
ST Date	-2.15	NA	NA	NA	-0.34	NA	3406.50	5.73	0.02
AN Storm, Upwell, ST Date, MA HCI	-2.13	0.07	NA	0.24	-0.31	0.11	3407.12	6.35	0.01
ER, ST Date	-2.14	NA	0.09	NA	-0.32	NA	3407.85	7.07	0.01
ST Date, AN Storm	-2.14	-0.01	NA	NA	-0.34	NA	3408.51	7.73	0.01
ST Date, Upwell	-2.15	NA	NA	NA	-0.34	0.00	3408.51	7.74	0.01
Parameter Statistics									
Σ weight		0.42	0.85	0.95	0.97	0.44			
Model-averaged β		0.05	0.21	0.27	-0.30	0.05			
Model-averaged CI (2.5%, 97.5%)		0.00, 0.34	0.00, 0.42	0.00, 0.56	-0.48, 0.00	-0.04, 0.27			

We found that in years when the majority of murre mortality occurred later (i.e., ER Mode is higher), the proportion of those carcasses in molt increased relative to years with earlier

mortality peaks (mean proportion = 0.12 for years with a mortality peak before August, versus 0.17 for years with a mortality peak in August or later; Table 1, Figure 5, SM Figure 5). This makes intuitive sense – in years in which the majority of murres that die do so before molt (molt season begins in August), molting is a non-issue. Early mortality years also tend to be years with higher adult ER (between ER Mode and adult ER, $R^2 = 0.13$, SM Figure 9).

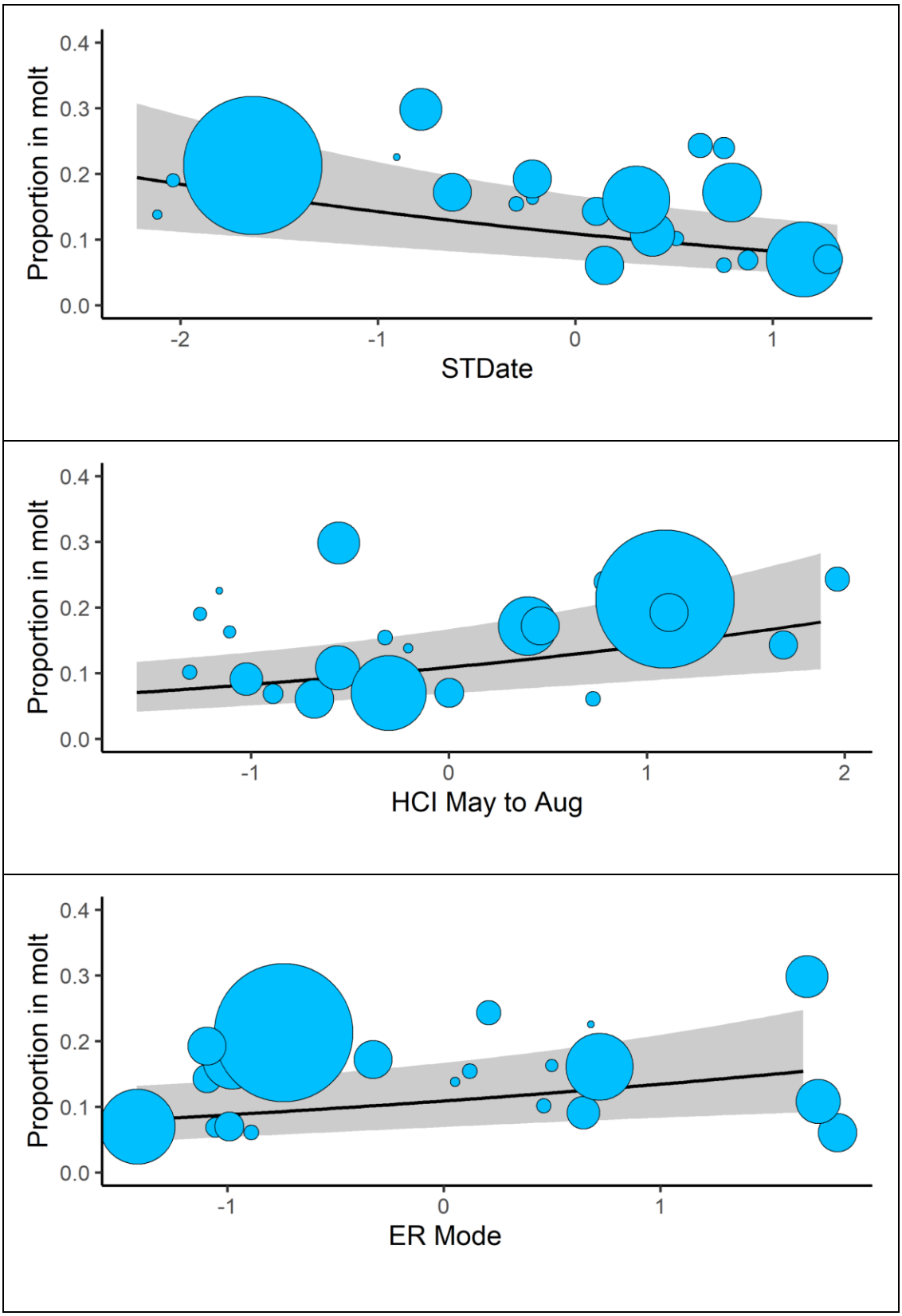

Proportion in molt
0.4
0.3
0.2
0.1
0.0
-2
-1
0
1
STDate
Proportion in molt
0.4
0.3
0.2
0.1
0.0
-1
0
1
2
HCI May to Aug
Proportion in molt
0.4
0.3
0.2
0.1
0.0
-1
0
1
ER Mode

Figure 5: Fitted values from the top model with point estimate and 95% confidence interval for spring transition date ("STDate"), HCI over May to August ("HCI May to Aug"), and the mode date on which birds were encountered ("ER Mode"). The data and fitted values are scaled. The size of each dot is proportional to the sample size of the year.

Our modeling also indicated that as high quality (i.e., colder water) foraging habitat became increasingly compressed in the nearshore space (i.e., higher HCI values), the proportion of murre carcasses in molt increased (Figure 5), rising at a model-averaged β of 0.27. This finding supports our hypothesis that environmental stress should exacerbate the role of molt in provoking mortality. However, earlier ST Dates were also associated with a higher proportion of carcasses in molt, with a model-averaged β of -0.30. We had theorized that earlier transition years would create more productivity in the system (e.g., Logerwell *et al.* 2003, Brodeur *et al.* 2006), thereby decreasing the proportion of birds that die while molting. Thus, this relationship does not appear to support our hypothesis.

To understand these apparently conflicting results, we re-examined them in the broader context of mortality patterns. Years with the highest environmental stress typified by high HCI, low cumulative upwelling, and a later ST Date (Table 1, Figure 5) were also years in which more murres died (i.e., ER of adults is high, SM Figures 7-9), and died earlier (i.e., ER Mode is early, Table 1, Figure 5, SM Figure 4, SM Figure 9). In these years, molt is an inconsequential mortality factor. In years when environmental stressors were lower overall, and ST date tended to be earlier, more murres survived to molt. We posit that in these years there is a secondary impact of HCI: warmer summers tended to increase the chance of mortality in molting murres.

Finally, our other two variables: storminess proxied by significant wave height over the August-November window, and May-August cumulative upwelling, were variably present in top models,

approximately half as important measured as cumulative weight, and had a relatively low impact ($\beta = 0.05$ in both cases; Table 1). Cumulative upwelling in particular appeared to have little association with yearly variation in Prop. Molt, with a model-averaged 95% confidence interval that spanned zero (2.5%: -0.04, 97.5%: 0.27). The impact of fall storms, while mild, was in the predicted direction, with a higher proportion of carcasses found in molt in stormier years.

DISCUSSION

General

Our analyses demonstrate that while natural history stressors, specifically flight feather molt, can interact with environmental stressors to affect murre mortality, timing is key. Specifically, we found that years with a mortality peak before the beginning of August (corresponding to a Julian date of ~250) tended to have a slightly lower proportion (0.12) of carcasses in molt within our molt time window (August-November), whereas when the mortality peak fell in August or later, the molt proportion was higher (0.17; Table 1, Figure 5, SM Figure 5). Years with an early mortality peak also tended to have a later system transition to a predominantly upwelling state, or ST Date (Bograd *et al.* 2009). Later ST dates corresponded to a lower proportion of carcasses that died in molt during the molt window. This is consistent with an interpretation that in years with particularly high environmental stress early in the breeding season, murres reach a mortality threshold before they reach molt. In other words, molt becomes an inconsequential factor. However, we also found that foraging habitat quality, which we proxied by calculating the mean proportional area within 150 km of the coast >1° C above the 30-year climatology (or HCI),

appeared to interact with molt to influence mortality. When foraging habitat quality was poorer during the breeding season and prior to molt, the proportion of carcasses in molt during the molt window was increased. In sum, it appears that while molt is stressful (i.e., it can interact with HCI to increase the likelihood of dying), the molt signal can be obscured by high environmental stress early in the year.

Molt Phenology

We used GAMMs to explore the timing of molt in murres, using records of beached birds collected by a citizen science program. Although our estimates of the calendar window of molt dynamics may be biased later by the delays between molt onset, mortality, carcass deposition, and carcass detection (e.g., Wiese *et al.* 2001, Munilla *et al.* 2011, Martin *et al.* 2019), because COASST conducts effort-controlled surveys on a monthly basis in a way that is consistent from year-to-year (Parrish *et al.* 2017), we have no reason to believe that carcass deposition would be systematically biased from year to year.

Our estimate of the interannual range in onset of molt in murres, at ~38 days, was somewhat higher than estimates from other synchronously molting birds. While few studies have estimated range in onset of molt in synchronously molting waterbirds at a population level, Flockhart (2010) studied 5 species of synchronously molting warblers (*Dendroica*, *Setophaga*, *Seiurus*, and *Wilsonia* spp.) from 2000-2007 and found that all species initiated molt within a 6-day period ($n \approx 450$ across all species). Other estimates for synchronously molting birds were derived from individual birds in a single year. For example, molt onset in one year (2006) was estimated for 15 common eiders using length of stay at known molting grounds as a metric, and found to range over 19 days (Savard *et al.* 2011). Overall, it may be that synchronously molting birds have some

variation across species in ability to control when to undergo molt, such that murres have more individual variation in when to start molt. However, it is important to note that our dataset spans 811 molting birds over 22 years, so it may be better able to pick up variation that exists across years.

We also observed greater variation in the number of days spanned by peak location than onset across years (peak location: 81 days, onset: 38 days). This supports the greater variation in molt duration, versus onset, observed for other synchronously molting birds at both the individual, intra-annual level (e.g., Savard *et al.* 2011) and the population, interannual level (e.g., Flockhart 2010). For instance, Darby *et al.* (2022) found a range in onset of flight feather molt in Atlantic puffins (*Fratercula arctica*) in 2010 (n = 4) of 20 days, compared to a 30 day range in duration. Higher variance in molt duration may be a de facto response to environmental conditions, at least in murres. Thompson (1998) compared molt in 12 murres during an El Niño year (1993) with 42 murres in a non-El Niño year (1996) and found a tripling in molt duration during the El Niño year. Murres have been found to change their diet in response to shifts in prey availability (Ainley *et al.* 1996), and Thompson (1998) suggested the extended molt during El Niño could be due to nutritional constraint (i.e., without sufficient high-quality prey, birds were unable to regrow their flight feathers as fast as in higher prey availability years).

In summary, our modeling of the annual timing of synchronous molting suggests that molt is fairly constrained in timing of onset and peak location most years, but that peak location (and likely, duration) in particular can be influenced by the timing and intensity of environmental stressors, as suggested by Thompson (1998). This is in line with findings for other synchronously

molting birds, although we found higher variability for both metrics of molt timing than other studies did, which could be explained by our interannual and population-level approach.

Influence of Environmental Stressors

Numerous studies have linked anomalously high sea surface temperature to elevated seabird mortality (Jones *et al.* 2018, Glencross *et al.* 2021, Jones *et al.* 2023), and specifically to elevated murre mortality (Piatt *et al.* 2020, Schoen *et al.* 2022). Our results suggest that in years with high environmental stress, murres tend to die earlier and in greater numbers, diluting the impact of molt on mortality (i.e., a lower Prop. Molt). However, environmental stressors like foraging habitat quality can still play a role in post breeding season mortality, leading to relative high numbers of carcasses in molt during the molt window. Our findings are also very much in line with intra-annual reports of high numbers of molting murre deaths during environmentally stressful time periods (Holdgate 1971, Jones *et al.* 2017). During a 2015 die-off in California, necropsies on a subset of murres (n = 35) indicated the birds died from starvation due to a warm water event along with a harmful algal bloom (Gibble *et al.* 2018). Of the adult birds necropsied (n = 5), 63% were undergoing molt. Gibble et al. (2018) hypothesized that prey shortages occurring during molt meant that murres had difficulty moving to new prey patches when food was low. Taken together with this past research, our findings suggest that environmental stressors, such as low habitat foraging quality, can combine with the natural history stressor of molt to "push molting birds over the edge" sooner than non-molting birds.

Overall, it appears that in years that are highly environmentally stressful early in the year, more murres die early on, and fewer birds get the chance to reach molt. A year with a high proportion of birds dying in molt is likely to be environmentally stressful, but less stressful than years when

many birds died before reaching molt. By itself, the proportion of birds in molt is a better signal of the timing than the intensity of stressors. Use of a long-term, fine-grained beached bird dataset is necessary to be able to account for stressor timing and pick up the increased mortality associated with interactions between environmental and natural history stressors.

ACKNOWLEDGEMENTS

Thank you to the many citizen scientists who contributed data to this work, and to Jennifer Ruesink and Abigail Swann for critical feedback.

REFERENCES

Aebischer, N. J. (1993). Immediate and delayed effects of a gale in late spring on the breeding of the shag Phalacrocorax aristotelis. *Ibis*, *135*(3), 225-232.

Ainley, D. G., Nettleship, D. N., and Storey, A.E. (2021). Common Murre (*Uria aalge*), version 2.0. In Birds of the World (S. M. Billerman, P. G. Rodewald, and B. K. Keeney, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.offcampus.lib.washington.edu/10.2173/bow.commur.02

Ainley, D. G., Spear, L. B., Allen, S. G., & Ribic, C. A. (1996). Temporal and spatial patterns in the diet of the common murre in California waters. The Condor, 98(4), 691-705.

Anderson, E. M., R. D. Dickson, E. K. Lok, E. C. Palm, J.-P. L. Savard, D. Bordage, and A. Reed (2020). Surf Scoter (*Melanitta perspicillata*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.offcampus.lib.washington.edu/10.2173/bow.sursco.01

Anker-Nilssen, T., Harris, M. P., Kleven, O., & Langset, M. (2017). Status, origin, and population level impacts of Atlantic Puffins killed in a mass mortality event in southwest Norway early 2016. *Seabird*, *30*, 1-14.

Baes, L., Santiago, C. D., Roman, L., dos Santos Costa, P. C., Pugliesi, É., & Reigada, C. (2024). Beached seabirds as plastic biomonitors in Brazil from the Beach Monitoring Project of the Santos Basin (PMP-BS). *Marine Pollution Bulletin*, *199*, 115847.

Bailey, E. P., & Davenport, G. H. (1972). Die-off of common murres on the Alaska Peninsula and Unimak Island. The Condor, 74(2), 215-219.

Barreau, E., Kato, A., Chiaradia, A., & Ropert-Coudert, Y. (2021). The consequences of chaos: Foraging activity of a marine predator remains impacted several days after the end of a storm. *Plos one*, *16*(7), e0254269.

Birkhead, T. R., & Taylor, A. M. (1977). Moult of the guillemot Uria aalge. Ibis, 119(1), 80-85.

Bograd, S. J., Schroeder, I., Sarkar, N., Qiu, X., Sydeman, W. J., & Schwing, F. B. (2009). Phenology of coastal upwelling in the California Current. *Geophysical Research Letters*, *36*(1).

Bridge, E. S. (2004). The effects of intense wing molt on diving in alcids and potential influences on the evolution of molt patterns. Journal of Experimental Biology, 207(17), 3003-3014.

Bridge, E. S. (2006). Influences of morphology and behavior on wing-molt strategies in seabirds. *Mar Ornithol*, *34*, 7-19.

Brodeur, R. D., Ralston, S., Emmett, R. L., Trudel, M., Auth, T. D., & Phillips, A. J. (2006). Anomalous pelagic nekton abundance, distribution, and apparent recruitment in the northern California Current in 2004 and 2005. *Geophysical Research Letters*, *33*(22).

Camphuysen, C. J., & Heubeck, M. (2001). Marine oil pollution and beached bird surveys: the development of a sensitive monitoring instrument. *Environmental pollution*, *112*(3), 443-461.

Checkley Jr, D. M., & Barth, J. A. (2009). Patterns and processes in the California Current System. *Progress in Oceanography*, *83*(1-4), 49-64.

Clairbaux, M., Mathewson, P., Porter, W., Fort, J., Strøm, H., Moe, B., ... & Grémillet, D. (2021). North Atlantic winter cyclones starve seabirds. *Current Biology*, *31*(17), 3964-3971.

Darby, J. H., Harris, M. P., Wanless, S., Quinn, J. L., Bråthen, V. S., Fayet, A. L., ... & Jessopp, M. J. (2022). A new biologging approach reveals unique flightless molt strategies of Atlantic puffins. Ecology and Evolution, 12(12), e9579.

Dawson, A., Perrins, C. M., Sharp, P. J., Wheeler, D., & Groves, S. (2009). The involvement of prolactin in avian molt: the effects of gender and breeding success on the timing of molt in Mute swans (Cygnus olor). *General and comparative endocrinology*, *161*(2), 267-270.

Descamps, S., Tarroux, A., Varpe, Ø., Yoccoz, N. G., Tveraa, T., & Lorentsen, S. H. (2015). Demographic effects of extreme weather events: snow storms, breeding success, and population growth rate in a long-lived Antarctic seabird. Ecology and Evolution, 5(2), 314-325.

Dietz, M. W., Daan, S., & Masman, D. (1992). Energy requirements for molt in the kestrel Falco tinnunculus. *Physiological zoology*, *65*(6), 1217-1235.

Elliott, K. H., Chivers, L. S., Bessey, L., Gaston, A. J., Hatch, S. A., Kato, A., ... & Hare, J. F. (2014). Windscapes shape seabird instantaneous energy costs but adult behavior buffers impact on offspring. *Movement Ecology*, *2*, 1-15.

Fayet, A. L., Clucas, G. V., Anker-Nilssen, T., Syposz, M., & Hansen, E. S. (2021). Local prey shortages drive foraging costs and breeding success in a declining seabird, the Atlantic puffin. Journal of Animal Ecology, 90(5), 1152-1164.

Flockhart, D. T. (2010). Timing of events on the breeding grounds for five species of sympatric warblers. *Journal of Field Ornithology*, *81*(4), 373-382.

Fort, J., Porter, W. P., & Grémillet, D. (2009). Thermodynamic modelling predicts energetic bottleneck for seabirds wintering in the northwest Atlantic. *Journal of Experimental Biology*, *212*(15), 2483-2490.

Fromant, A., Delord, K., Bost, C. A., Eizenberg, Y. H., Botha, J. A., Cherel, Y., ... & Arnould, J. P. (2021). Impact of extreme environmental conditions: foraging behaviour and trophic ecology responses of a diving seabird, the common diving petrel. *Progress in Oceanography*, *198*, 102676.

Gibble, C., Duerr, R., Bodenstein, B., Lindquist, K., Lindsey, J., Beck, J., ... & Kudela, R. (2018). Investigation of a largescale Common Murre (Uria aalge) mortality event in California, USA, in 2015. *Journal of wildlife diseases*, *54*(3), 569-574.

Gjerdrum, C., Vallée, A. M., St. Clair, C. C., Bertram, D. F., Ryder, J. L., & Blackburn, G. S. (2003). Tufted puffin reproduction reveals ocean climate variability. *Proceedings of the National Academy of Sciences*, *100*(16), 9377-9382.

Glencross, J. S., Lavers, J. L., & Woehler, E. J. (2021). Breeding success of short-tailed shearwaters following extreme environmental conditions. *Marine Ecology Progress Series*, *672*, 193-203.

González-Solís, J., Felicísimo, A., Fox, J. W., Afanasyev, V., Kolbeinsson, Y., & Muñoz, J. (2009). Influence of sea surface winds on shearwater migration detours. Marine Ecology Progress Series, 391, 221-230.

Grunst, M. L., Grunst, A. S., Grémillet, D., Kato, A., Gentès, S., & Fort, J. (2023). Keystone seabird may face thermoregulatory challenges in a warming Arctic. *Scientific Reports*, *13*(1), 16733.

Guillemette, M., Pelletier, D., Grandbois, J. M., & Butler, P. J. (2007). Flightlessness and the energetic cost of wing molt in a large sea duck. *Ecology*, *88*(11), 2936-2945.

Hamel, N. J., Burger, A. E., Charleton, K., Davidson, P., Lee, S., Bertram, D. F., & Parrish, J. K. (2009). Symposium Paper: Bycatch and beached birds: assessing mortality impacts in coastal net fisheries using marine bird strandings. *Marine Ornithology*, *37*, 41-60.

Heubeck, M., Meek, E. R., Mellor, R. M., & Wilson, M. (2009). Numbers of Atlantic Puffins Fratercula arctica found on beached bird surveys in Orkney and Shetland over a 30-year period. *Seabird*, *22*, 19-35.

Holdgate, M. W. (1971). The sea bird wreck in the Irish Sea Autumn 1969. The seabird Wreck of 1969 in the Irish Sea: A Report of the investigation conducted under the auspices of the Natural Environment Research Council.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Jacox, M. G., Edwards, C. A., Hazen, E. L., & Bograd, S. J. (2018). Coastal upwelling revisited: Ekman, Bakun, and improved upwelling indices for the US West Coast. *Journal of Geophysical Research: Oceans*, *123*(10), 7332-7350.

Jenni, L., & Winkler, R. (2020). The biology of moult in birds. Bloomsbury Publishing.

Johns, M. E., Warzybok, P., Jahncke, J., Doak, P., Lindberg, M., & Breed, G. A. (2022). Episodes of high recruitment buffer against climate-driven mass mortality events in a North Pacific seabird population. *Journal of Animal Ecology*, *91*(2), 345-355.

Jones, T., Parrish, J. K., Lindsey, J., Wright, C., Burgess, H. K., Dolliver, J., ... & Lindquist, K. (2023). Marine bird mass mortality events as an indicator of the impacts of ocean warming. *Marine Ecology Progress Series*.

Jones, T., Parrish, J. K., Peterson, W. T., Bjorkstedt, E. P., Bond, N. A., Ballance, L. T., ... & Harvey, J. (2018). Massive mortality of a planktivorous seabird in response to a marine heatwave. *Geophysical Research Letters*, *45*(7), 3193-3202.

Jones, T., Parrish, J. K., Punt, A. E., Trainer, V. L., Kudela, R., Lang, J., ... & Hickey, B. (2017). Mass mortality of marine birds in the Northeast Pacific caused by Akashiwo sanguinea. *Marine Ecology Progress Series*, *579*, 111-127.

Lindström, Å., Visser, G. H., & Daan, S. (1993). The energetic cost of feather synthesis is proportional to basal metabolic rate. Physiological Zoology, 66(4), 490-510.

Logerwell, E. A., Mantua, N., Lawson, P. W., Francis, R. C., & Agostini, V. N. (2003). Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (Oncorhynchus kisutch) marine survival. Fisheries Oceanography, 12(6), 554-568.

Long, J. A. (2020). jtools: Analysis and presentation of social scientific data. R package version, 2(0).

Martin, N., Varela, V. W., Dwyer, F. J., Tuttle, P., Ford, R. G., & Casey, J. (2019). Evaluation of the fate of carcasses and dummies deployed in the nearshore and offshore waters of the northern Gulf of Mexico. Environmental monitoring and assessment, 191, 1-13.

Munilla, I., Arcos, J. M., Oro, D., Álvarez, D., Leyenda, P. M., & Velando, A. (2011). Mass mortality of seabirds in the aftermath of the Prestige oil spill. Ecosphere, 2(7), 1-14.

National Oceanic and Atmospheric Administration. (2016, September 12). Significant wave height. National Weather Service. https://www.weather.gov/mfl/waves#:~:text=Significant%20wave%20height%20is%20an,most%20erosion%20on%20a%20beach

NOAA Fisheries. (2024, February 2). Local physical indicators. https://www.fisheries.noaa.gov/west-coast/science-data/local-physical-indicators

Newell, M., Wanless, S., Harris, M. P., & Daunt, F. (2015). Effects of an extreme weather event on seabird breeding success at a North Sea colony. *Marine Ecology Progress Series*, *532*, 257-268.

Olin, A. B., Dück, L., Berglund, P. A., Karlsson, E., Bohm, M., Olsson, O., & Hentati-Sundberg, J. (2023). Breeding failures and reduced nest attendance in response to heat stress in a high-latitude seabird. *Marine Ecology Progress Series*.

Oro, D. (2014). Seabirds and climate: knowledge, pitfalls, and opportunities. Frontiers in Ecology and Evolution, 2, 79.

Osborne, O. E., Hara, P. D., Whelan, S., Zandbergen, P., Hatch, S. A., & Elliott, K. H. (2020). Breeding seabirds increase foraging range in response to an extreme marine heatwave. *Marine Ecology Progress Series*, *646*, 161-173.

Parrish, J. K., Bond, N., Nevins, H., Mantua, N., Loeffel, R., Peterson, W. T., & Harvey, J. T. (2007). Beached birds and physical forcing in the California Current System. *Marine Ecology Progress Series*, *352*, 275-288.

Parrish, J. K., Litle, K., Dolliver, J., Hass, T., Burgess, H. K., Frost, E., ... & Jones, T. (2017). Defining the baseline and tracking change in seabird populations: The Coastal Observation and Seabird Survey Team (COASST). In Citizen science for coastal and marine conservation (pp. 19-38). Routledge.

Parrish, J. K., & Zador, S. G. (2003). Seabirds as indicators: an exploratory analysis of physical forcing in the Pacific Northwest coastal environment. *Estuaries*, *26*, 1044-1057.

Peery, M. Z., Newman, S. H., Storlazzi, C. D., & Beissinger, S. R. (2009). Meeting reproductive demands in a dynamic upwelling system: foraging strategies of a pursuit-diving seabird, the marbled murrelet. The Condor, 111(1), 120-134.

Piatt, J. F., Parrish, J. K., Renner, H. M., Schoen, S. K., Jones, T. T., Arimitsu, M. L., ... & Sydeman, W. J. (2020). Extreme mortality and reproductive failure of common murres resulting from the northeast Pacific marine heatwave of 2014-2016. *PloS one*, *15*(1), e0226087.

Portugal, S. J., Green, J. A., & Butler, P. J. (2007). Annual changes in body mass and resting metabolism in captive barnacle geese (Branta leucopsis): the importance of wing moult. *Journal of Experimental Biology*, *210*(8), 1391-1397.

Robinson, B. W., Decicco, L. H., Johnson, J. A., & Ruthrauff, D. R. (2018). Unusual foraging observations associated with seabird die-offs in Alaska. Marine Ornithology, 46, 149-153.

Roletto, J., Mortenson, J., Harrald, I., Hall, J., & Grella, L. (2003). Beached bird surveys and chronic oil pollution in central California. *Marine Ornithology*, *31*, 21-28.

Romano, M. D., Renner, H. M., Kuletz, K. J., Parrish, J. K., Jones, T., Burgess, H. K., ... & Causey, D. (2020). Die–offs, reproductive failure, and changing at–sea abundance of murres in the Bering and Chukchi Seas in 2018. *Deep Sea Research Part II: Topical Studies in Oceanography*, *181*, 104877.

Ronconi, R. A., & Burger, A. E. (2008). Limited foraging flexibility: increased foraging effort by a marine predator does not buffer against scarce prey. *Marine Ecology Progress Series*, *366*, 245-258.

Russell, R. W. (2020). Arctic Loon (*Gavia arctica*), version 1.0. In Birds of the World (S. M. Billerman, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.arcloo.01

Sandvik, H., Coulson, T. I. M., & SÆTHER, B. E. (2007). A latitudinal gradient in climate effects on seabird demography: results from interspecific analyses. *Global Change Biology*, *14*(4), 703-713.

Savard, J. P. L., Lesage, L., Gilliland, S. G., Gilchrist, H. G., & Giroux, J. F. (2011). Molting, staging, and wintering locations of common eiders breeding in the Gyrfalcon Archipelago, Ungava Bay. *Arctic*, 197-206.

Schoen, S. K., Arimitsu, M. L., Marsteller, C. E., & Piatt, J. F. (2022). Lingering impacts of the 2014-2016 northeast Pacific marine heatwave on seabird demography in Cook Inlet, Alaska (USA). *Marine Ecology Progress Series*.

Schrimpf, M. B., Parrish, J. K., & Pearson, S. F. (2012). Trade-offs in prey quality and quantity revealed through the behavioral compensation of breeding seabirds. Marine Ecology Progress Series, 460, 247-259.

Schroeder, I. D., Santora, J. A., Mantua, N., Field, J. C., Wells, B. K., Hazen, E. L., ... & Bograd, S. J. (2022). Habitat compression indices for monitoring ocean conditions and ecosystem impacts within coastal upwelling systems. *Ecological Indicators*, *144*, 109520.

Shepard, E., Cole, E. L., Neate, A., Lempidakis, E., & Ross, A. (2019). Wind prevents cliff-breeding birds from accessing nests through loss of flight control. *Elife*, *8*, e43842.

Shields, M. (2020). Brown Pelican (*Pelecanus occidentalis*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.offcampus.lib.washington.edu/10.2173/bow.brnpel.01

Simpson G (2023). _gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv_. R package version 0.8.1, <https://gavinsimpson.github.io/gratia/>.

Sydeman, W. J., Bradley, R. W., Warzybok, P., Abraham, C. L., Jahncke, J., Hyrenbach, K. D., ... & Ohman, M. D. (2006). Planktivorous auklet Ptychoramphus aleuticus responses to ocean climate, 2005: Unusual atmospheric blocking? *Geophysical Research Letters*, *33*(22).

Tate, H. M., Studholme, K. R., Domalik, A. D., Drever, M. C., Romero, L. M., Gormally, B. M., ... & Crossin, G. T. (2021). Interannual measures of nutritional stress during a marine

heatwave (the Blob) differ between two North Pacific seabird species. Conservation Physiology, 9(1), coab090.

Tavares, D. C., Moura, J. F., Merico, A., & Siciliano, S. (2020). Mortality of seabirds migrating across the tropical Atlantic in relation to oceanographic processes. *Animal Conservation*, *23*(3), 307-319.

Taylor, I. R. (1983). Effect of wind on the foraging behaviour of Common and Sandwich Terns. *Ornis Scandinavica*, 90-96.

Thompson, C. W., Wilson, M. L., Melvin, E. F., & Pierce, D. J. (1998). An unusual sequence of flight-feather molt in common murres and its evolutionary implications. *The Auk*, *115*(3), 653-669.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic bulletin & review, 11, 192-196.

Wiese, F. K., & Elmslie, K. I. M. (2006). Underuse and misuse of data from beached bird surveys. *Marine Ornithology*, *34*, 157-159.

Wiese, F. K., Jones, I. L., & Nettleship, D. (2001). Experimental support for a new drift block design to assess seabird mortality from oil pollution. The Auk, 118(4), 1062-1068.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3-36.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and hall/CRC.

Van Holmes, C. (2007). Mean spring and fall upwelling transition dates off the Oregon and Washington coasts.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R (Vol. 574, p. 574). New York: springer.

Chapter 2- Supplementary Materials

Classifying Molt Status

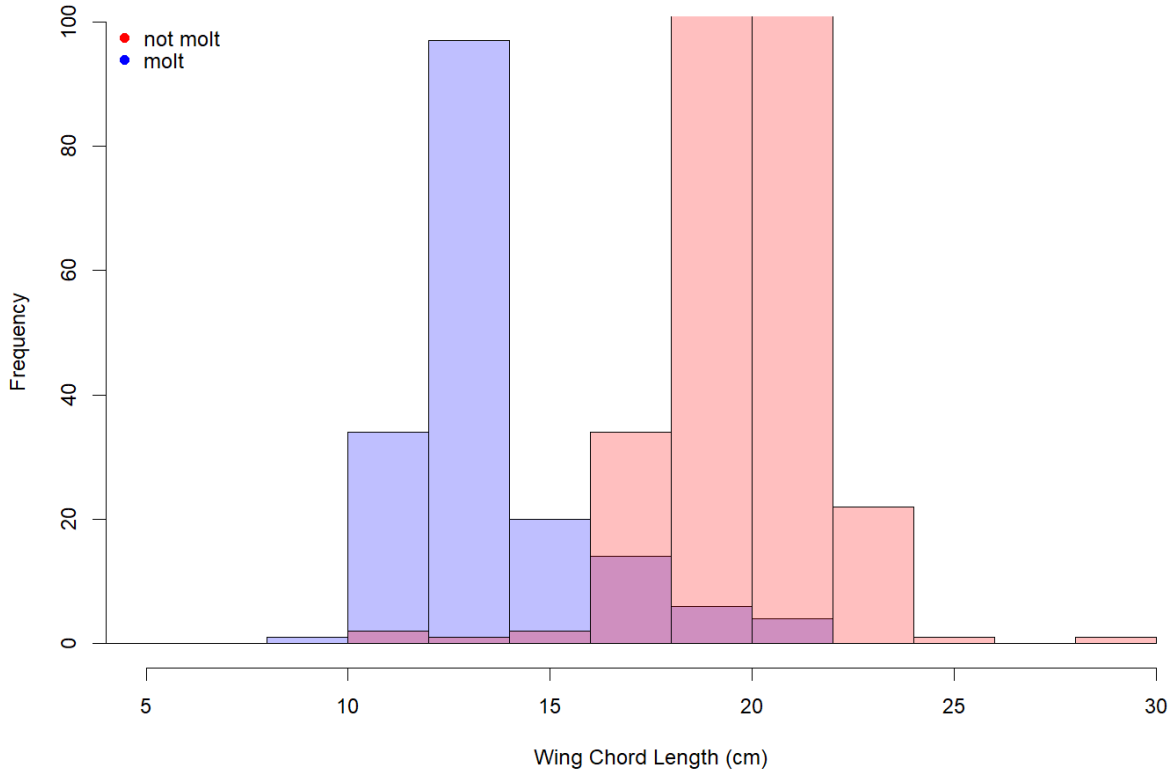


Figure 1: For the 2,143 birds with both a wing chord measurement and a molt status from the verifier ("in molt" or "not molt"), the frequency of wing chord measurement lengths (cm).

Table 1: For birds with both a wing chord measurement and a molt status from the verifier ("in molt" or "not in molt"), this shows the molt status from the verifier and the wing chord measurement of the bird. A wing chord measurement of <17 cm was considered in molt, while ≥17 cm was considered not in molt. Disagreements (the verifier molt status was "in molt" but the wing length suggested "not in molt" by our threshold, or vice versa) are bolded.

	Wing measurement- in molt	Wing measurement - not in molt
Verifier- in molt	156	**21**
Verifier- not in molt	**5**	1961

Using 17 cm as a wing chord length threshold resulted in agreement with the verifier molt status on 2,117 out of 2,143 birds (or 98.8%). Therefore, for our analyses, we classified molt based on

this measurement threshold for the 7,302 carcasses that did not have a molt classification from the verifier. 811 birds were classified in molt based on measurement, and 6,491 were classified as not in molt based on measurement.

Relationship between Proportion Molt and Peak Height

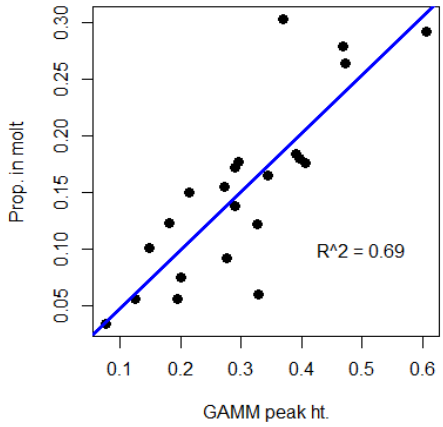
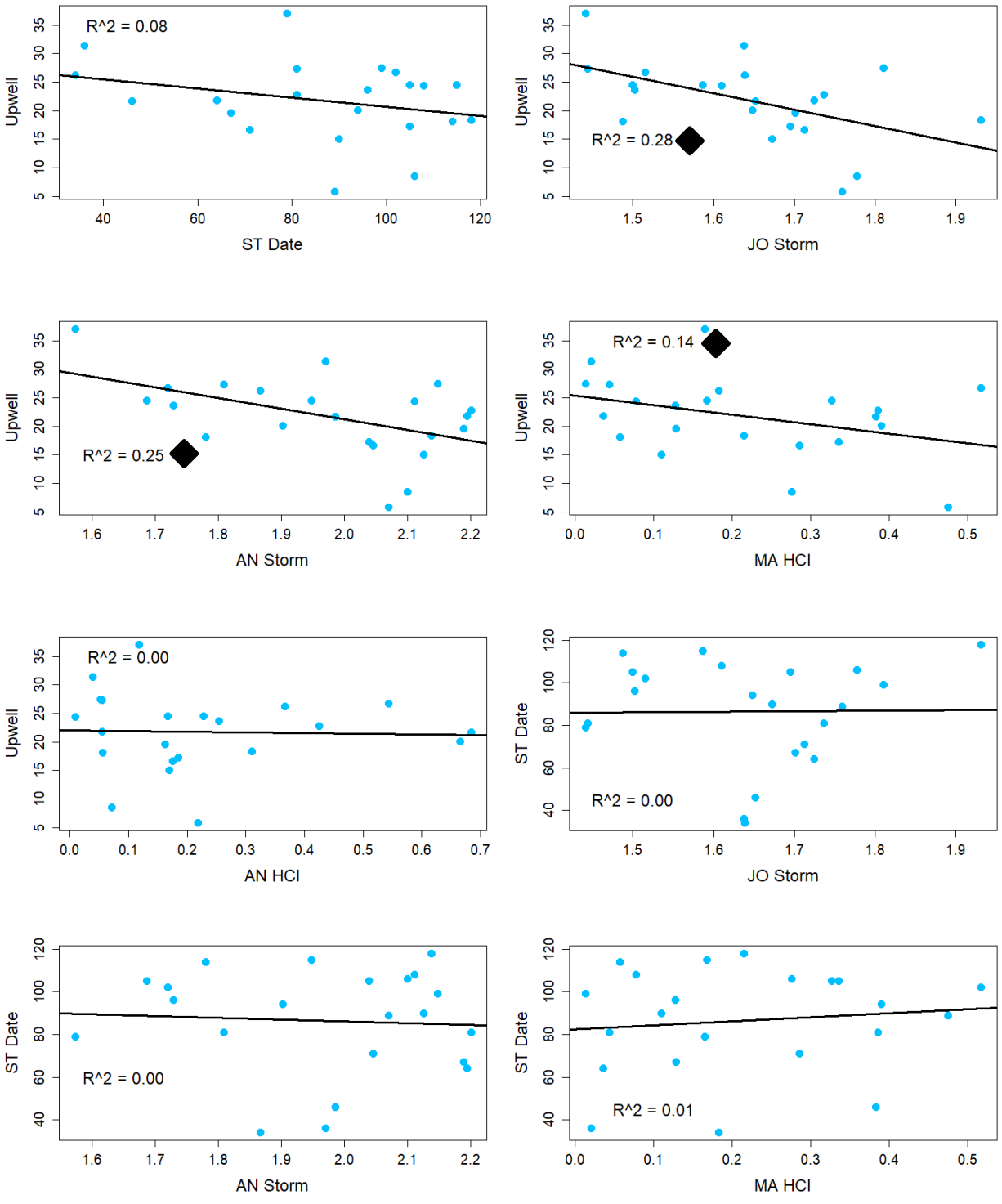


Figure 2: Relationship between proportion of birds in molt between August and November ("Prop. In molt") and the peak height from the Generalized Additive Mixed Models ("GAMM peak ht.).

Relationships Between Environmental Variables, Prop. Molt, and Encounter Rate



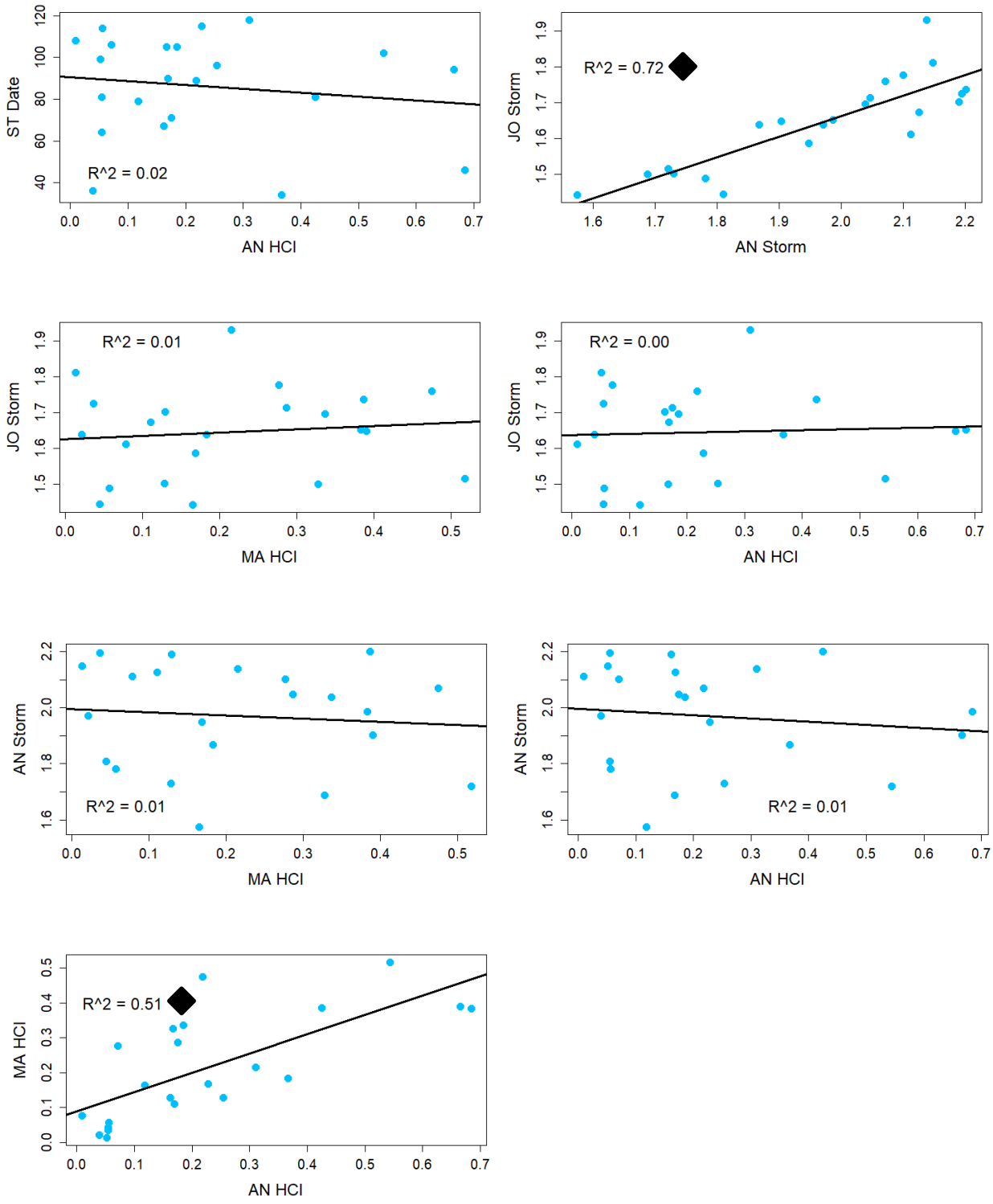


Figure 3: Pairwise comparisons between the environmental variables. R^2 and line of best fit are displayed along with raw data. "Upwell"= upwelling May to August, spring transition date = "ST Date", storminess in July to October = "JO Storm", storminess in August to November =

"AN Storm", HCI in May to August = "MA HCI", and HCI in August to November = "AN HCI". ◆ indicates an $R^2 \geq 0.10$.

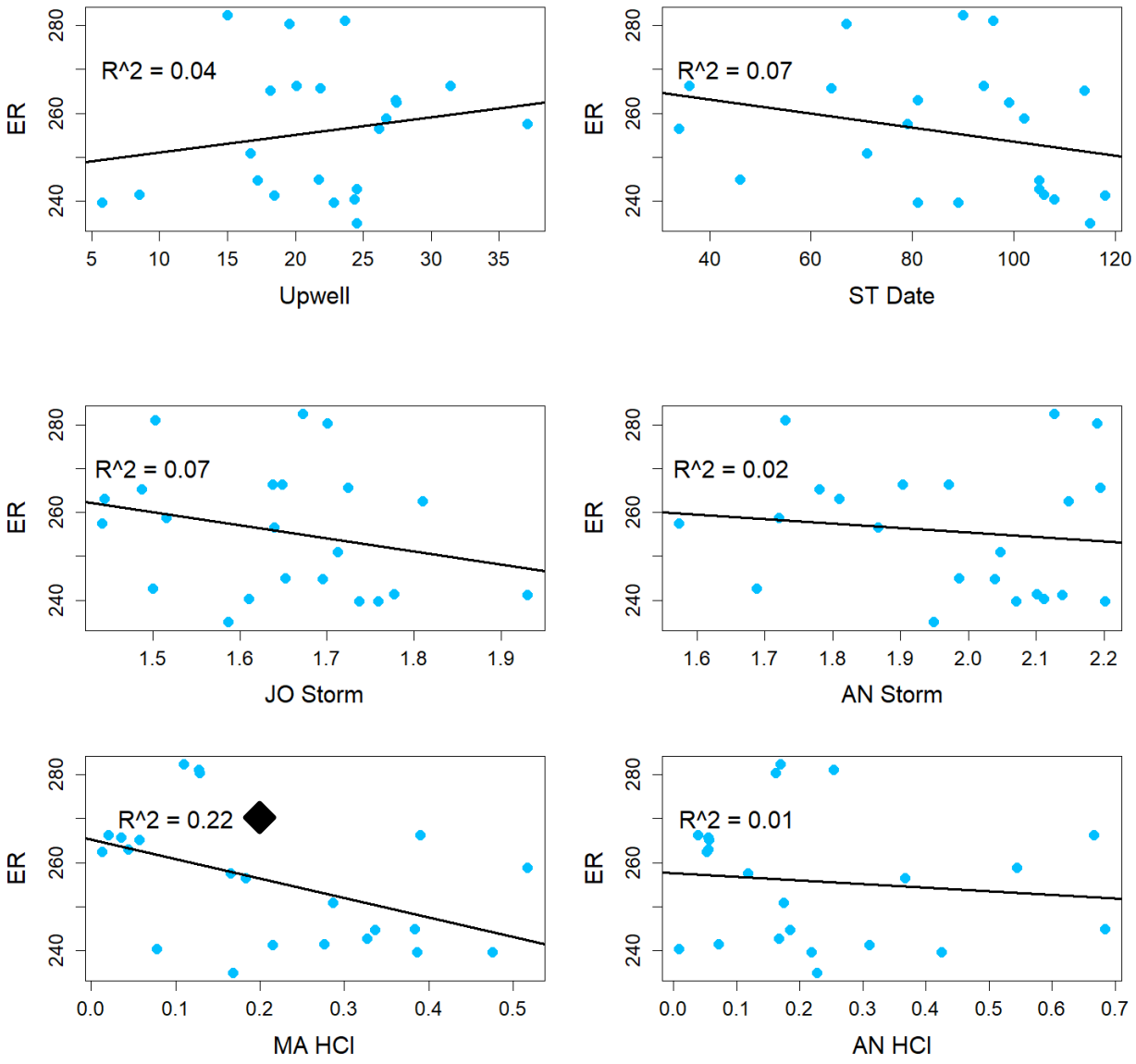


Figure 4: Relationship between encounter rate and the environmental variables. R^2 and line of best fit are displayed along with raw data. "Upwell"= upwelling May to August, spring transition date = "ST Date", storminess in July to October = "JO Storm", storminess in August to November = "AN Storm", HCI in May to August = "MA HCI", HCI in August to November = "AN HCI", and mode in encounter rate = "ER". ◆ indicates an $R^2 \geq 0.10$.

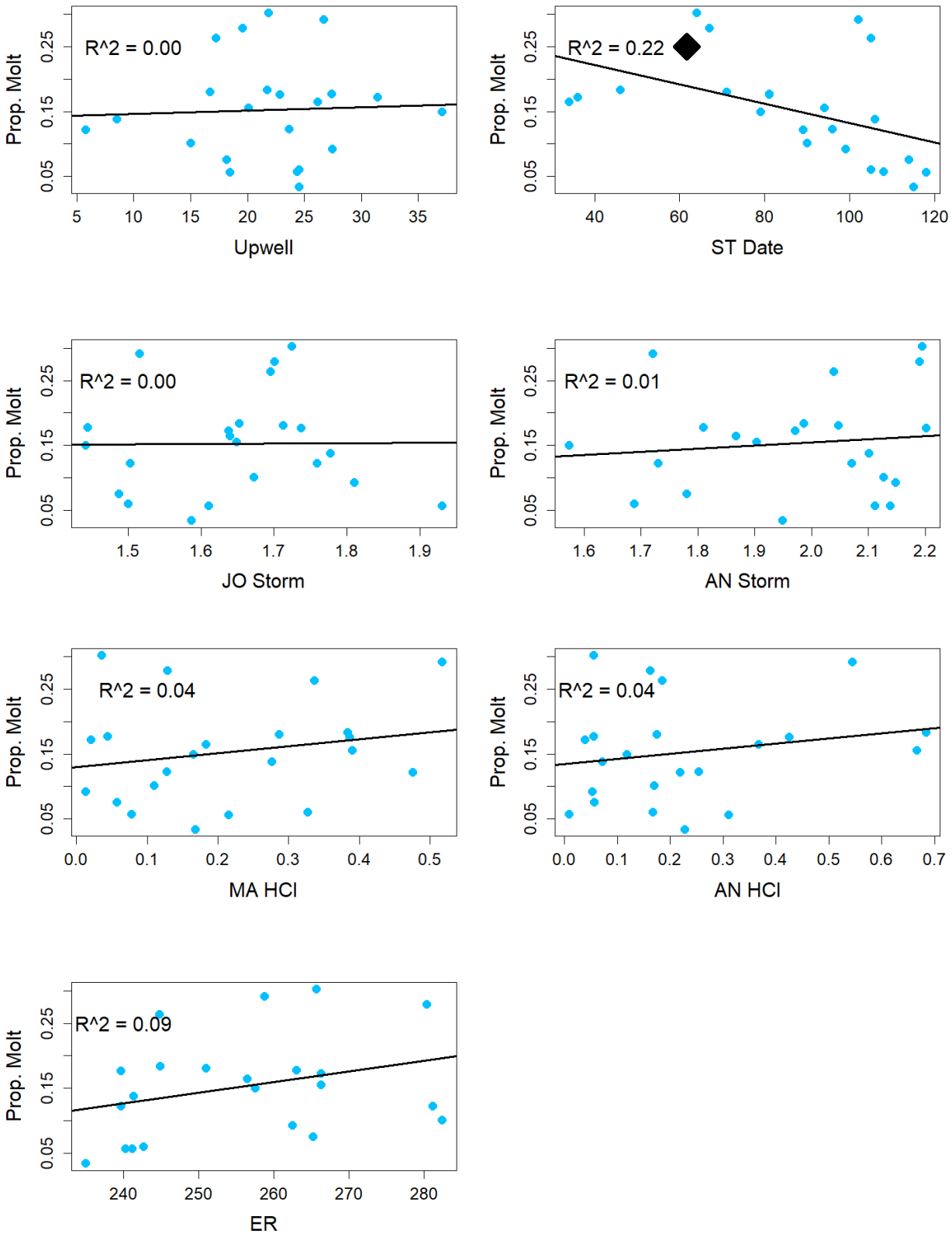
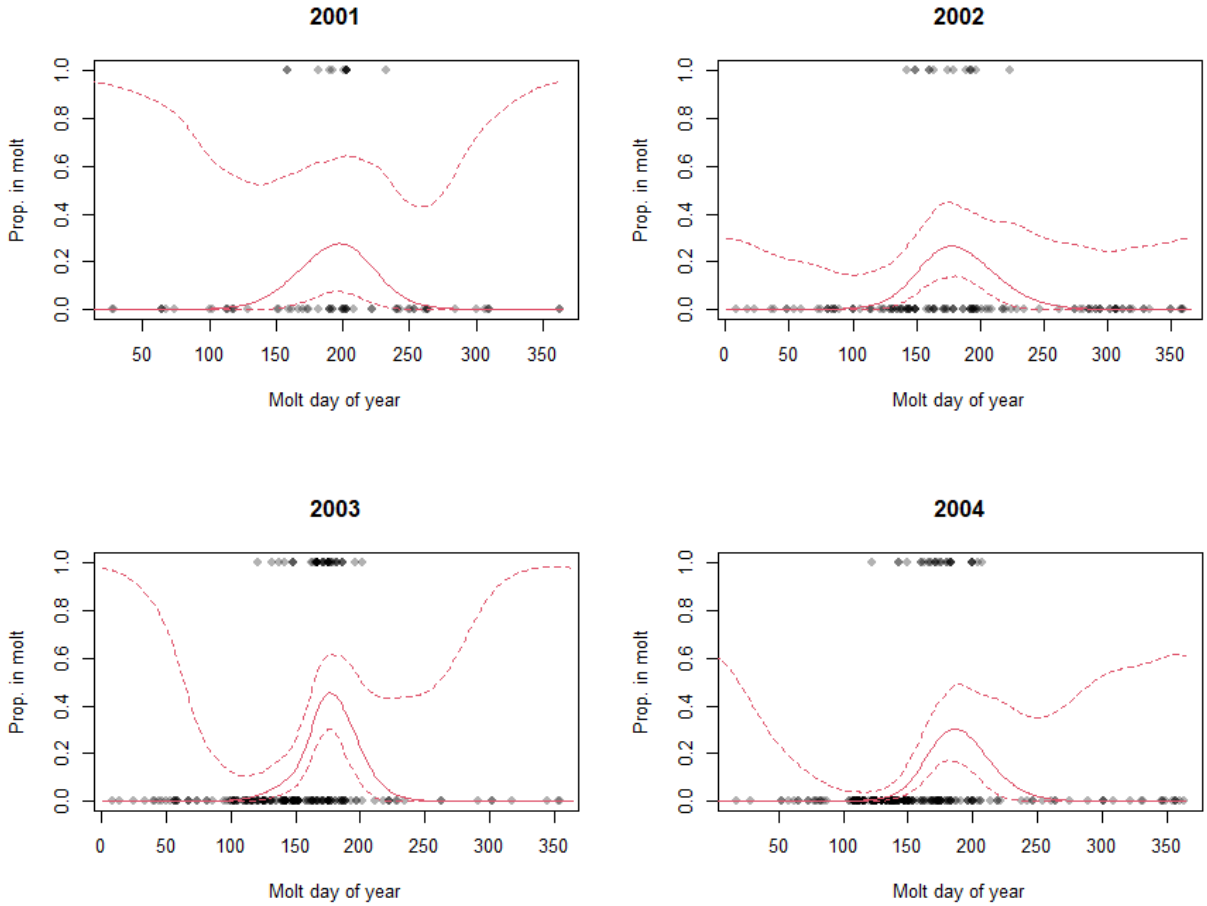
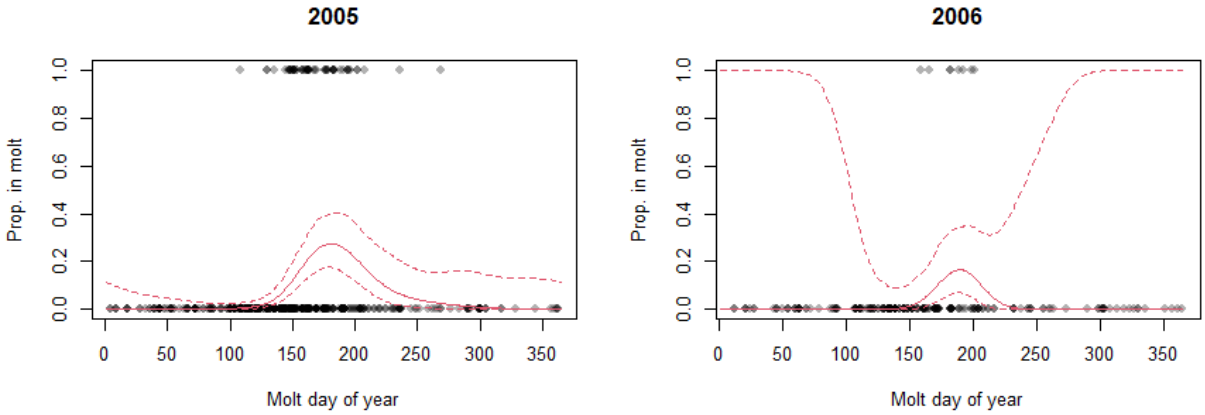


Figure 5: Relationship between Prop. Molt and the environmental variables. R^2 and line of best

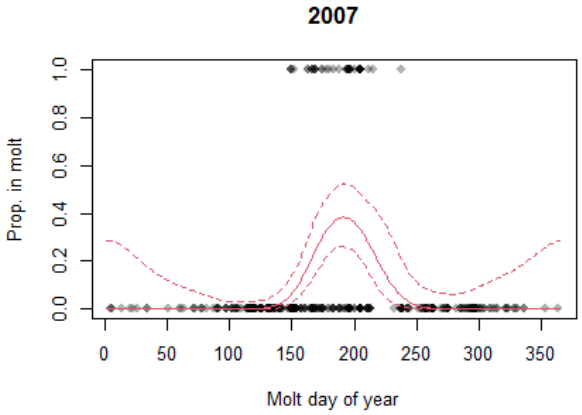
fit are displayed along with raw data. “Upwell”= upwelling May to August, spring transition date = “ST Date”, storminess in July to October = “JO Storm”, storminess in August to November = “AN Storm”, HCI in May to August = “MA HCI”, HCI in August to November = “AN HCI”, and mode in encounter rate = “ER”. ◆ indicates an $R^2 \geq 0.10$.

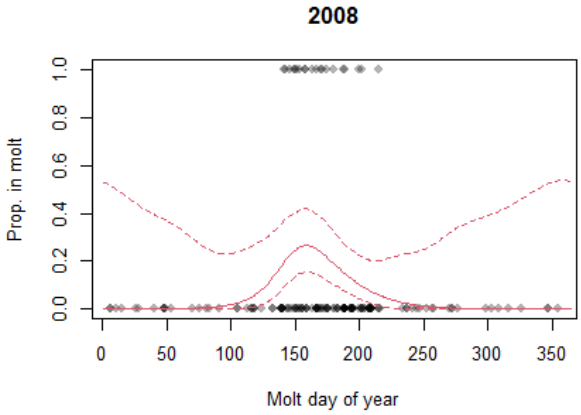
Generalized Additive Mixed Model Outputs

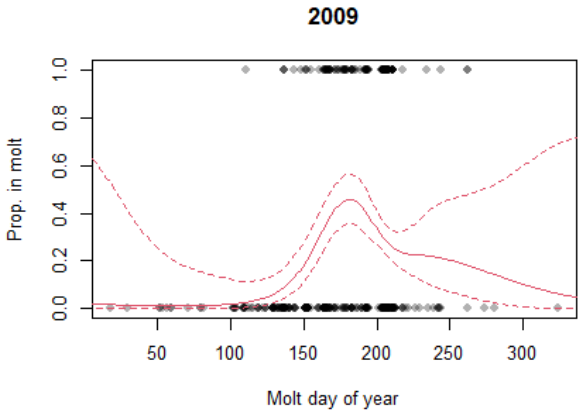


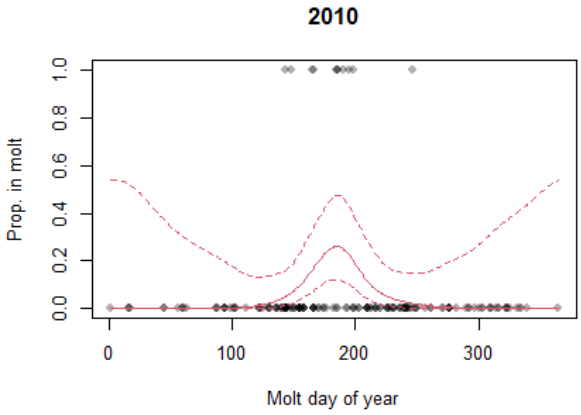

2005
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350
Molt day of year

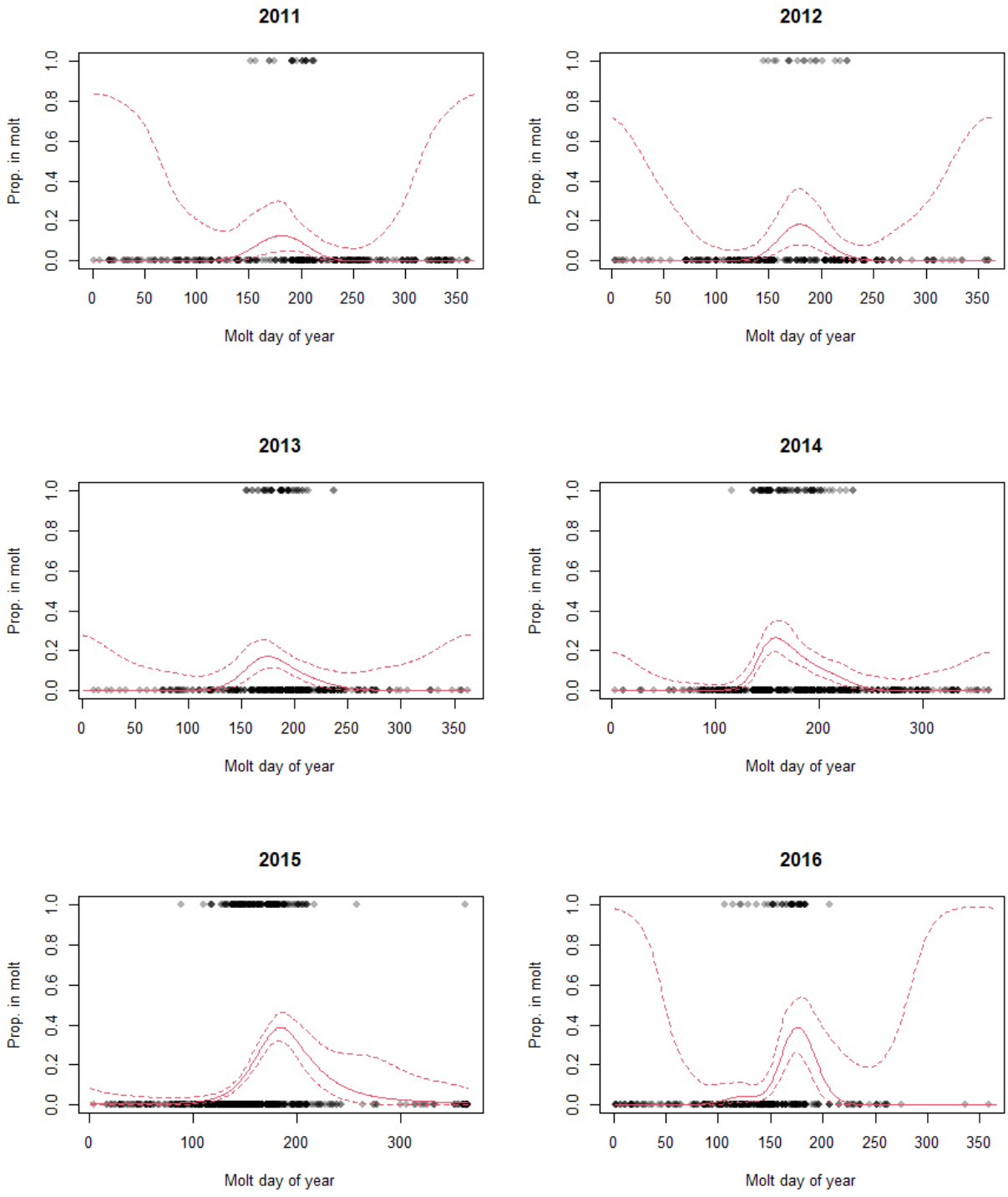
2006
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350
Molt day of year


2007
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350
Molt day of year


2008
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350
Molt day of year


2009
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
50
100
150
200
250
300
Molt day of year


2010
Prop. in molt
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
Molt day of year


2011
2012
2013
2014
2015
2016
Prop. in molt
Molt day of year
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
350

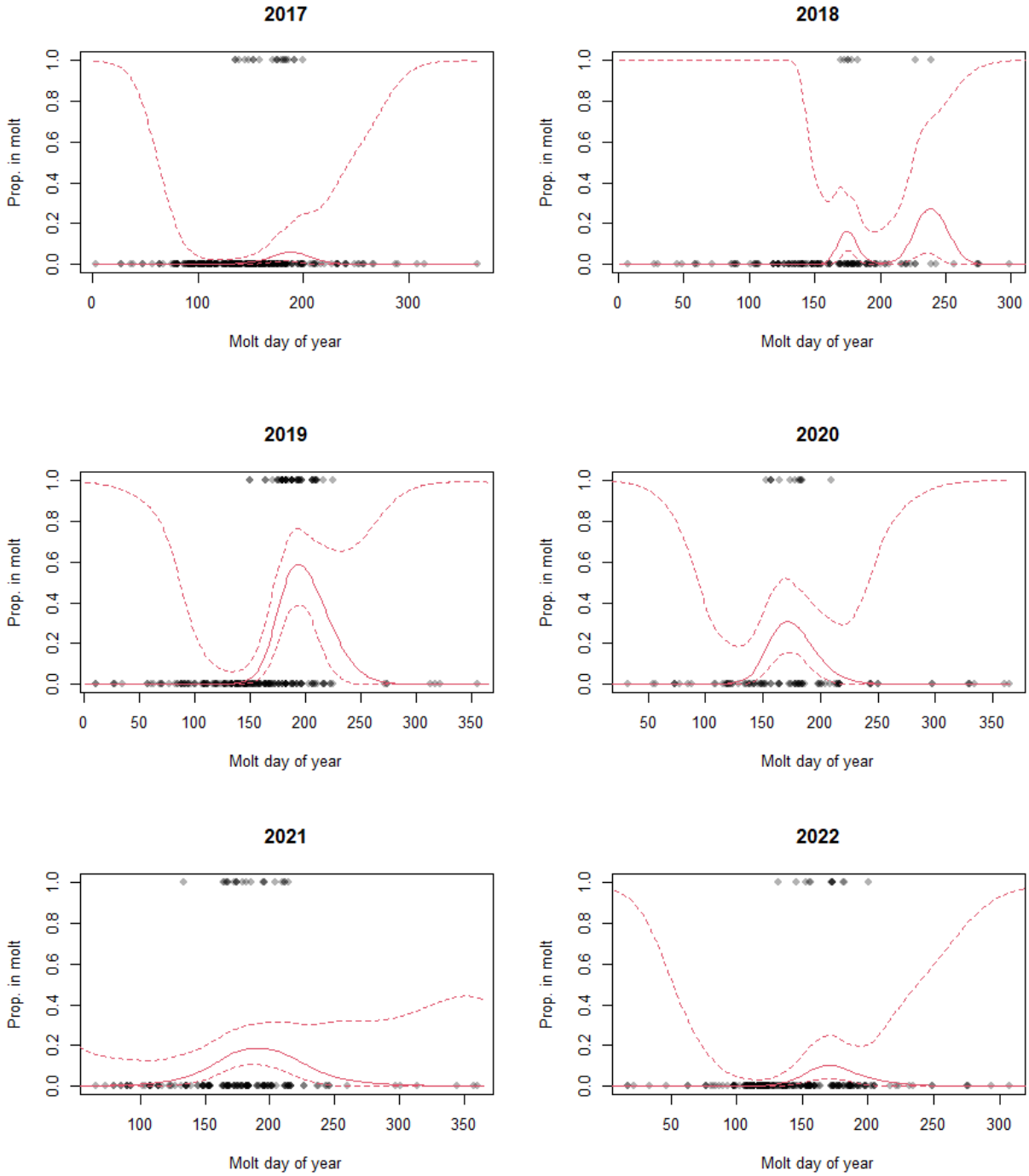


Figure 6: The Generalized Additive Mixed Model outputs for each year (2001-2022). For each year, the point estimate (solid line) and 95% confidence interval (dotted lines) for the proportion in molt over molt day in year (where day 1 = April 1st) are shown. Raw values (individual birds) are shown in grey dots. The darker the grey dot, the more birds were found on that day.

Relationships between different peak features

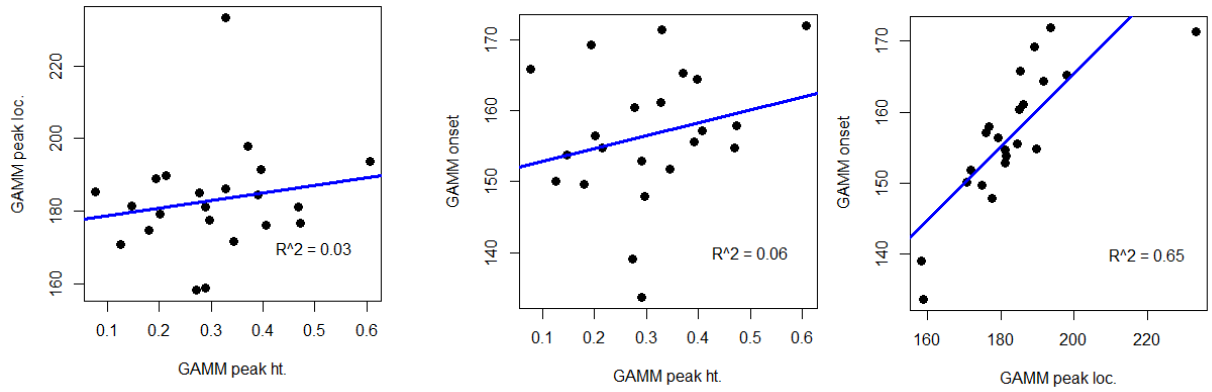


Figure 7: Relationships between the three peak features from the Generalized Additive Mixed Models: peak height ("peak ht."), onset, and peak location ("peak loc."). The R^2 value and line of best fit is shown for each set of features.

Choosing a Model Set

Table 2: Summed AICc for all models within each model set, and Δ AICc between each model set and the model set with the lowest summed AICc.

Model Set	**Summed AICc**	**Δ AICc**
August-November Storminess + May-August HCI	109121.2	0
July-October Storminess + May-August HCI	109128.5	7.4
August-November Storminess + August-November HCI	109186.2	65.0
July- October Storminess + August- November HCI	109188.1	66.9

Removal of years

Table 3: Resultsof removing 2001. Regression coefficients, model statistics, and parameter statistics for the Generalized Linear Mixed Models for the relationship between proportion of dead murres in molt per year and environmental variables. Model statistics include AICc, ΔAICc = AICc - min(AICc), model Akaike weight (a). Parameter statistics include summed Akaike weight of all models featuring that parameter (Σa), and model-averaged regression coefficients ($\bar{\beta} = \Sigma_i a_i \beta_i / \Sigma_i a_i$), and model-averaged confidence intervals (model-averaged CI). Int = intercept, AN Storm = August-November Storm, ER = mode day of year on which birds were found, MA HCI = May-August Habitat Compression Index, ST Date = Spring transition date, and Upwell = May-August cumulative upwelling. Models are listed in descending order of weight. Only models with weights >= 0.01 are shown.

Regression Coefficients							**Model Statistics**		
Predictors included	Int	AN Storm	ER	MA HCI	ST Date	Upwell	AICc	Δ AICc	Weight
ER, ST Date, MA HCI	-2.11	NA	0.24	0.32	-0.30	NA	3355.84	0.00	0.31
ER, AN Storm, Upwell, ST Date, MA HCI	-2.10	0.02	0.12	0.18	-0.27	0.00	3356.85	1.00	0.19
ER, ST Date, Upwell, MA HCI	-2.10	NA	0.24	0.35	-0.27	0.07	3357.02	1.17	0.17
ER, ST Date, AN Storm, MA HCI	-2.10	0.06	0.29	0.36	-0.29	NA	3357.68	1.83	0.12
ST Date, MA HCI	-2.13	NA	NA	0.22	-0.34	NA	3359.21	3.37	0.06
ST Date, Upwell, MA HCI	-2.13	NA	NA	0.24	-0.32	0.08	3360.54	4.70	0.03
ER, AN Storm, Upwell, MA HCI	-2.10	0.24	0.35	0.46	NA	0.27	3360.84	4.99	0.03
ST Date, AN Storm, MA HCI	-2.13	-0.01	NA	0.22	-0.34	NA	3361.21	5.37	0.02
ST Date	-2.15	NA	NA	NA	-0.34	NA	3362.09	6.25	0.01
AN Storm, Upwell, ST Date, MA HCI	-2.13	0.06	NA	0.26	-0.30	0.11	3362.32	6.47	0.01
ER, Upwell, MA HCI	-2.10	NA	0.31	0.39	NA	0.15	3363.34	7.50	0.01

ER, ST Date	-2.14	NA	0.09	NA	-0.32	NA	3363.49	7.64	0.01
ER, MA HCI	-2.10	NA	0.32	0.34	NA	NA	3363.73	7.88	0.01
ST Date, AN Storm	-2.15	-0.02	NA	NA	-0.34	NA	3364.08	8.23	0.01
Parameter Statistics									
Σ weight		0.39	0.85	0.96	0.95	0.45			
Model-averaged β		0.05	0.22	0.30	-0.30	0.06			
Model-averaged CI (2.5%, 97.5%)		0.00, 0.33	0.00, 0.42	0.04, 0.58	-0.46, 0.00	-0.03, 0.27			

Table 4: Results of removing 2009. Regression coefficients, model statistics, and parameter statistics for the Generalized Linear Mixed Models for the relationship between proportion of dead murres in molt per year and environmental variables. Model statistics include AICc, ΔAICc = AICc - min(AICc), model Akaike weight (a). Parameter statistics include summed Akaike weight of all models featuring that parameter (Σa), and model-averaged regression coefficients ($\bar{\beta} = \Sigma_i a_i \beta_i / \Sigma_i a_i$), and model-averaged confidence intervals (model-averaged CI). Int = intercept, AN Storm = August-November Storm, ER = mode day of year on which birds were found, MA HCI = May-August Habitat Compression Index, ST Date = Spring transition date, and Upwell = May-August cumulative upwelling. Models are listed in descending order of weight. Only models with weights >= 0.01 are shown.

Regression Coefficients							**Model Statistics**		
Predictors included	Int	AN Storm	ER	MA HCI	ST Date	Upwell	AICc	Δ AICc	Weight
ER, ST Date, MA HCI	-2.21	NA	0.18	0.32	-0.30	NA	3096.71	0.00	0.24
ER, ST Date, Upwell, MA HCI	-2.21	NA	0.17	0.35	-0.26	0.09	3097.33	0.62	0.17
ST Date, MA HCI	-2.24	NA	NA	0.25	-0.32	NA	3097.83	1.12	0.13
ER, AN Storm, Upwell, ST Date, MA HCI	-2.18	0.00	0.05	0.19	-0.26	0.02	3098.28	1.57	0.11
ST Date, Upwell, MA HCI	-2.24	NA	NA	0.29	-0.28	0.10	3098.29	1.58	0.11
ER, ST Date, AN Storm, MA HCI	-2.21	0.01	0.19	0.32	-0.30	NA	3098.72	2.01	0.09
ST Date, AN Storm, MA HCI	-2.24	-0.05	NA	0.25	-0.32	NA	3099.48	2.76	0.06
AN Storm, Upwell,	-2.22	0.03	NA	0.31	-0.27	0.13	3100.37	3.65	0.04

ST Date, MA HCI									
ST Date	-2.23	NA	NA	NA	-0.32	NA	3102.65	5.93	0.01
ER, AN Storm, Upwell, MA HCI	-2.20	0.22	0.29	0.46	NA	0.27	3103.16	6.45	0.01
Parameter Statistics									
Σ weight		0.31	0.63	0.97	0.97	0.45			
Model-averaged β		0.00	0.16	0.29	-0.29	0.08			
Model-averaged CI (2.5%, 97.5%)		-0.04, 0.28	0.00, 0.36	0.00, 0.55	-0.47, 0.00	-0.03, 0.27			

Table 5: Results of removing 2015. Regression coefficients, model statistics, and parameter statistics for the Generalized Linear Mixed Models for the relationship between proportion of dead murres in molt per year and environmental variables. Model statistics include AICc, ΔAICc = AICc - min(AICc), model Akaike weight (a). Parameter statistics include summed Akaike weight of all models featuring that parameter (Σa), and model-averaged regression coefficients ($\bar{\beta} = \Sigma_i a_i \beta_i / \Sigma_i a_i$), and model-averaged confidence intervals (model-averaged CI). Int = intercept, AN Storm = August-November Storm, ER = mode day of year on which birds were found, MA HCI = May-August Habitat Compression Index, ST Date = Spring transition date, and Upwell = May-August cumulative upwelling. Models are listed in descending order of weight. Only models with weights >= 0.01 are shown.

Regression Coefficients							**Model Statistics**		
Predictors included	Int	AN Storm	ER	MA HCI	ST Date	Upwell	AICc	Δ AICc	Weight
ER, ST Date, MA HCI	-2.02	NA	0.19	0.32	-0.36	NA	2750.60	0.00	0.24
ER, AN Storm, Upwell, ST Date, MA HCI	-1.99	0.00	0.06	0.27	-0.35	0.01	2751.06	0.46	0.19
ER, ST Date, Upwell, MA HCI	-2.02	NA	0.19	0.34	-0.34	0.09	2751.21	0.61	0.18
ST Date, MA HCI	-2.03	NA	NA	0.24	-0.42	NA	2752.35	1.75	0.10
ER, ST Date, AN Storm, MA HCI	-2.03	0.03	0.20	0.32	-0.36	NA	2752.51	1.91	0.09
ST Date, Upwell, MA HCI	-2.02	NA	NA	0.27	-0.39	0.09	2753.14	2.54	0.07
ST Date, AN Storm, MA HCI	-2.03	-0.02	NA	0.24	-0.42	NA	2754.32	3.73	0.04
AN Storm, Upwell, ST Date, MA HCI	-2.03	0.07	NA	0.29	-0.37	0.13	2754.73	4.13	0.03
ER, AN Storm, Upwell, MA HCI	-2.06	0.25	0.34	0.43	NA	0.28	2756.24	5.64	0.01
ST Date	-2.08	NA	NA	NA	-0.36	NA	2756.59	5.99	0.01
ER, ST Date	-2.08	NA	0.06	NA	-0.33	NA	2758.28	7.68	0.01
Parameter Statistics									
Σ weight		0.38	0.74	0.97	0.98	0.50			
Model-averaged β		0.02	0.16	0.30	-0.37	0.07			

Model-averaged CI (2.5%, 97.5%)	-0.02, 0.29	0.00, 0.37	0.00, 0.54	-0.57, 0.00	-0.01, 0.28	

Table 6: Difference between the model-averaged β for models with all years included versus models with individual years removed (2001, 2009, or 2015). Values shown are calculated as value with all models included – value with year removed.

Year Removed	AN Storm	ER	MA HCI	ST Date	Upwell
2001	0.00	-0.01	-0.03	0.00	-0.01
2009	0.05	0.05	-0.02	-0.01	-0.03
2015	0.03	0.05	-0.03	0.07	-0.02

Removing potentially unusual years changed the model-averaged β by less than 0.1, so we chose to maintain all years in the models.

Relationship between Habitat Compression Index (HCI) and encounter rate of adults and of murres overall

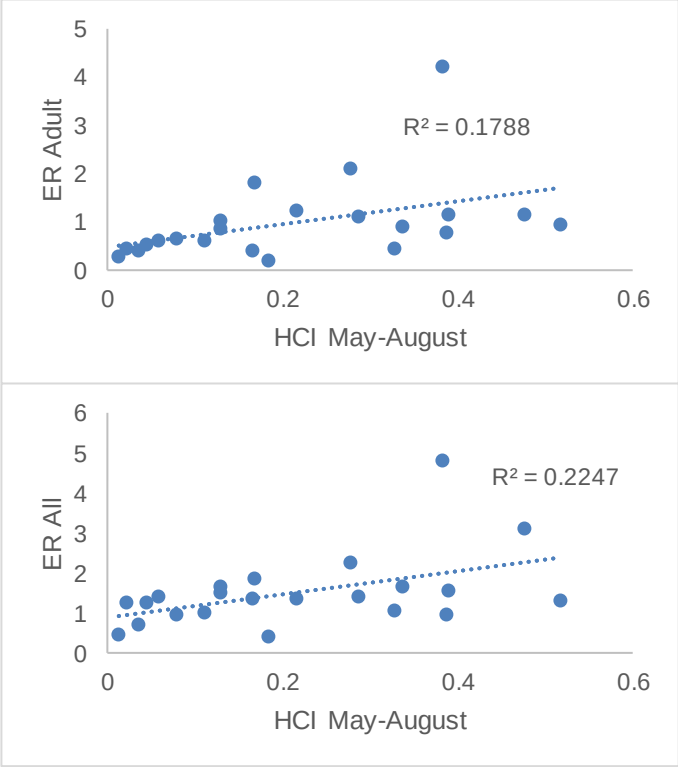


Figure 7: The relationship between Habitat Compression Index in May to August ("HCI May-August") and encounter rate ("ER") of adults and of murres of all ages ("all").

Relationship between proportion of murres that died in molt (Prop. Molt) and encounter rate

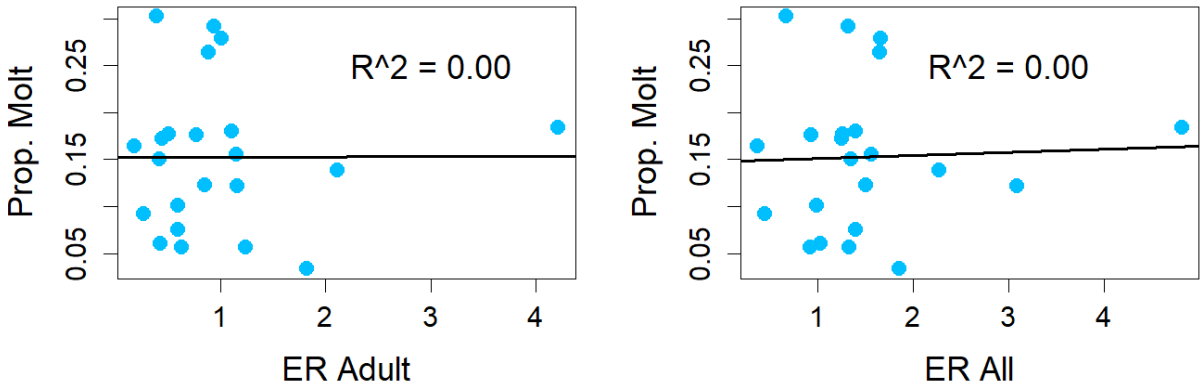
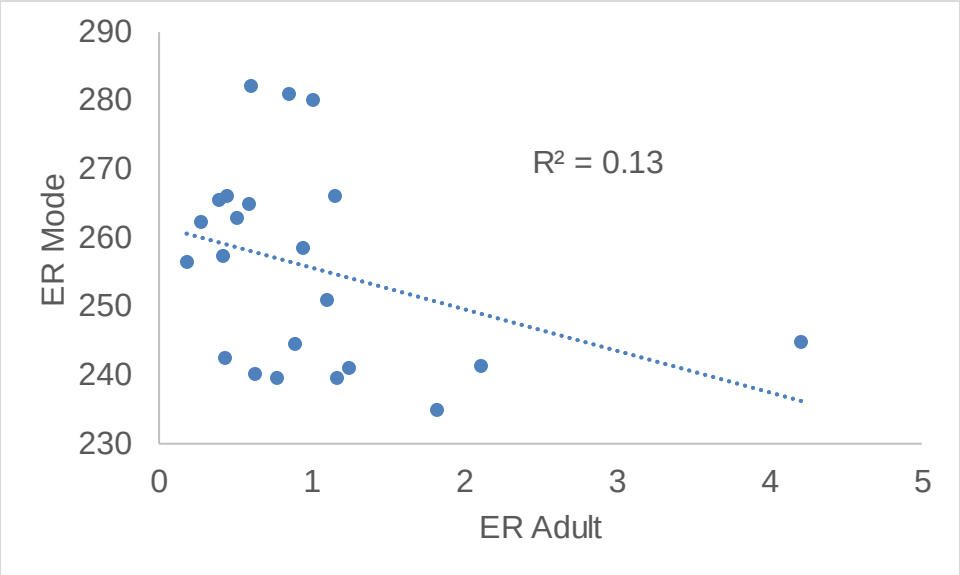


Figure 8: The relationship between the proportion of murres that died in molt (Prop. Molt) and encounter rate (ER) of adults and of all birds.

Relationship between encounter rate of adults, spring transition, and encounter rate timing



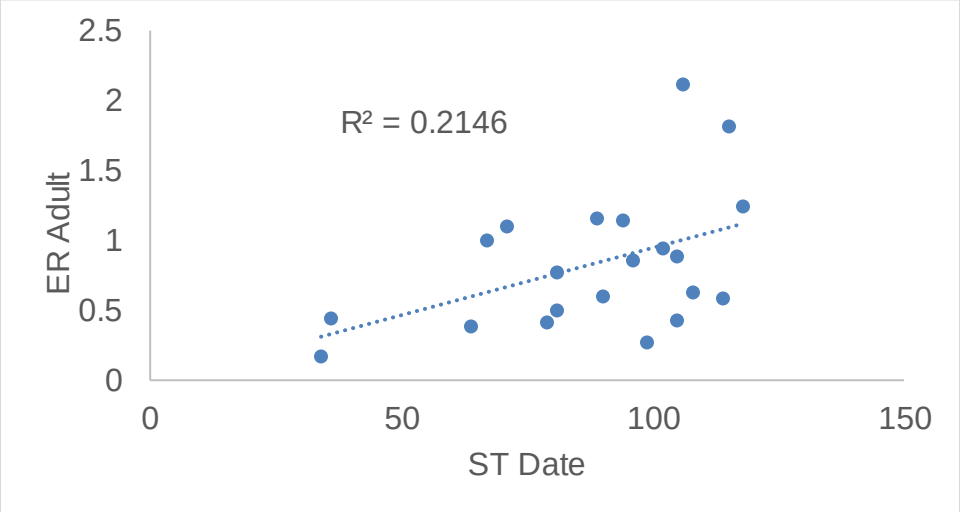


Figure 9: The relationship between the encounter rate of adult murres ("ER Adult"), spring transition date ("ST Date"), and mode in encounter rate day of year ("ER Mode").

Chapter 3. Demographics of Public Participation in Science: A Meta-Analytic Approach

Citation: Waugh, J. K., Lindsey, J. K., Stewart, M. Z., Winter, J. C., & Parrish, J. K. (2023). Demographics of Public Participation in Science: A Meta-Analytic Approach. *Citizen Science: Theory and Practice*, *8*(1), 61. DOI: 10.5334/cstp.610

Abstract

Citizen science, defined here as the voluntary participation of the public in scientific research, has been proposed as a method of increasing diversity in science. However, it is unknown whether citizen science participants are truly more diverse than traditional (i.e., academically trained) scientists. In this paper, we conducted a meta-analysis of citizen science participant demographics (for gender, race/ethnicity, retiree participation, age, and education) across English language peer-reviewed literature in the Web of Science (WOS) database. We collected data encompassing 151,854 unique within-project participants from 157 projects. By pairing a subset of our data confined to North America (n = 21 projects) with geographically compared census data, we found that citizen science participants are disproportionately white and educated, with high variation in gender of participants across projects. The geographically compared results also suggested that participants are primarily adults. We then used linear modeling to explore whether project attributes, including publication year, scientific focus of the project (project focus), and whether the project is online or hands-on (project access), explain variation in demographics (n = 121 projects). We found an increase in female participation over time, while biodiversity projects had higher participation from white and retired people than health projects, and online projects had more female and educated participants than hands-on projects. However, only ~7.5% of

citizen science papers we reviewed reported demographic data, suggesting a need for more representative data. This research suggests that there may be room for increased participation from groups that are currently underrepresented in citizen science.

Introduction

Public participation in science offers participants opportunities to engage through designing a study, collecting data, analyzing data, and/or interpreting results (Wiggins and Crowston 2011; Shirk et al. 2012; Haklay 2013). Such engagement spans from top-down, scientist-driven projects (often referred to as citizen science) to bottom-up, community-driven projects (often referred to as community science). Because there is no globally accepted term encompassing the full range of public participation in science, we default to the term citizen science, which we define as allowing individuals to join or leave at will, where participants are not specifically paid for their work, and where the work results in scientific findings.

In the United States of America (USA) alone, citizen science may now involve tens of millions of participants (Thigpen and Funk 2020), making it a major conduit of scientific information and praxis (Theobald et al. 2015). Increasing participation in citizen science is seen as a promise of diversification in science (Pandya 2012; Bonney et al. 2016; Paleco et al. 2021), particularly through inclusion of genders and races/ethnicities that are underrepresented in the Science, Technology, Engineering, and Mathematics (STEM) work force (Okrent and Brent 2021).

However, limited research to date suggests that citizen science participants tend to be white (NASEM 2018; Blake, Rhanor, and Pajic 2020; Allf et al. 2022) and highly educated (Evans et al. 2005; Blake, Rhanor, and Pajic 2020; Allf et al. 2022), at least in the USA. The pattern is less

clear for gender (Cooper and Smith 2010; NASEM 2018; Paleco et al. 2021). This pattern of underrepresentation may be repeated more broadly, at least in the USA. For instance, Schlachter (2021) reported that white, college-educated women are overrepresented in volunteer efforts, suggesting the problem may reflect trends beyond science and citizen science.

In the past 25 years, publications involving citizen science have increased by orders of magnitude (Follett and Stezov 2015); and across the landscape of public participation in science, equitable inclusion has been recognized as necessary (e.g., Paleco et al. 2021; Cooper et al. 2021). Given these developments, we return to the question of representation in citizen science to examine three primary questions:

- Are citizen science participants demographically different from the communities from which they are drawn?
- Does the structure or focus of the project explain variation in participant demographics?
- Are participant demographic trends changing over time?

In particular, we examined whether the overarching focus of the project (e.g., biodiversity), how the public is engaged (e.g., online versus in person), and where the project is located geographically, influence the demographics of participants. To date, individual studies of demographic trends in citizen science—with the exception of NASEM (2018)—have focused on only a few projects and/or demographic measures, with the result that a more comprehensive picture is elusive. Although NASEM (2018) reported descriptive statistics (including gender, age, race/ethnicity, education, and previous citizen science experience) for 68 projects, they did

not quantitatively examine how project factors co-vary with, or potentially influence, participant demographics, or whether demographics are changing as the social pressures for such change (e.g., Paleco et al. 2021; Cooper et al. 2021) increase.

In this paper, we present a literature meta-analysis that includes peer-reviewed articles between 2000 and June 2022. Our analysis is exploratory rather than hypothesis-driven. However, given the increasing number of calls for more diversity within citizen science (e.g., Paleco et al. 2021; Pateman and West 2023), our study sheds light on whether diversification is being realized. While our results reinforce some of the prior findings, we also find both temporal trends and high variation in participant demographics, some of which are associated with the structure and/or focus of the project. Therefore, we suggest that there are indications of how the citizen science landscape is changing and may become more inclusive.

Methods

Article set

Rather than redo the search produced by NASEM (2018), we incorporated their sample set of publications directly into this study. For consistency, we also used their topic field search terms:

- ("citizen science" or "community science" or "PSSR" or "public participation in scientific research" or "crowd-sourc*") and
- ("evaluation" or "survey" or "motivation").

We extended the NASEM dataset over the period 2017-01-01 to 2021-11-05, focusing our search within Web of Science (WOS: https://webofscience.com), and restricting our search to articles in English.

Our search resulted in 1,474 unique articles. However, articles were filtered out if they did not report on citizen science projects, where citizen science is defined as voluntary participation in scientific research. We further refined our article set to include only those that contained quantifiable demographic data. To this we added one paper published after our literature search: Allf et al. (2022). Within the citizen science publications, demographic data were collected through surveys, interviews, participant lists, and/or program databases. Our final sample set included 134 peer-reviewed papers that collectively described 157 unique projects (Figure 1; Supplemental File 2). Where projects were featured in multiple citations (e.g., two articles featured NestWatch), we used the most recent publication. Across all studies used in the analysis, there were a total of 151,854 unique within-project participants (because participant identity is protected, it was impossible to determine whether a single person was a participant in multiple projects).

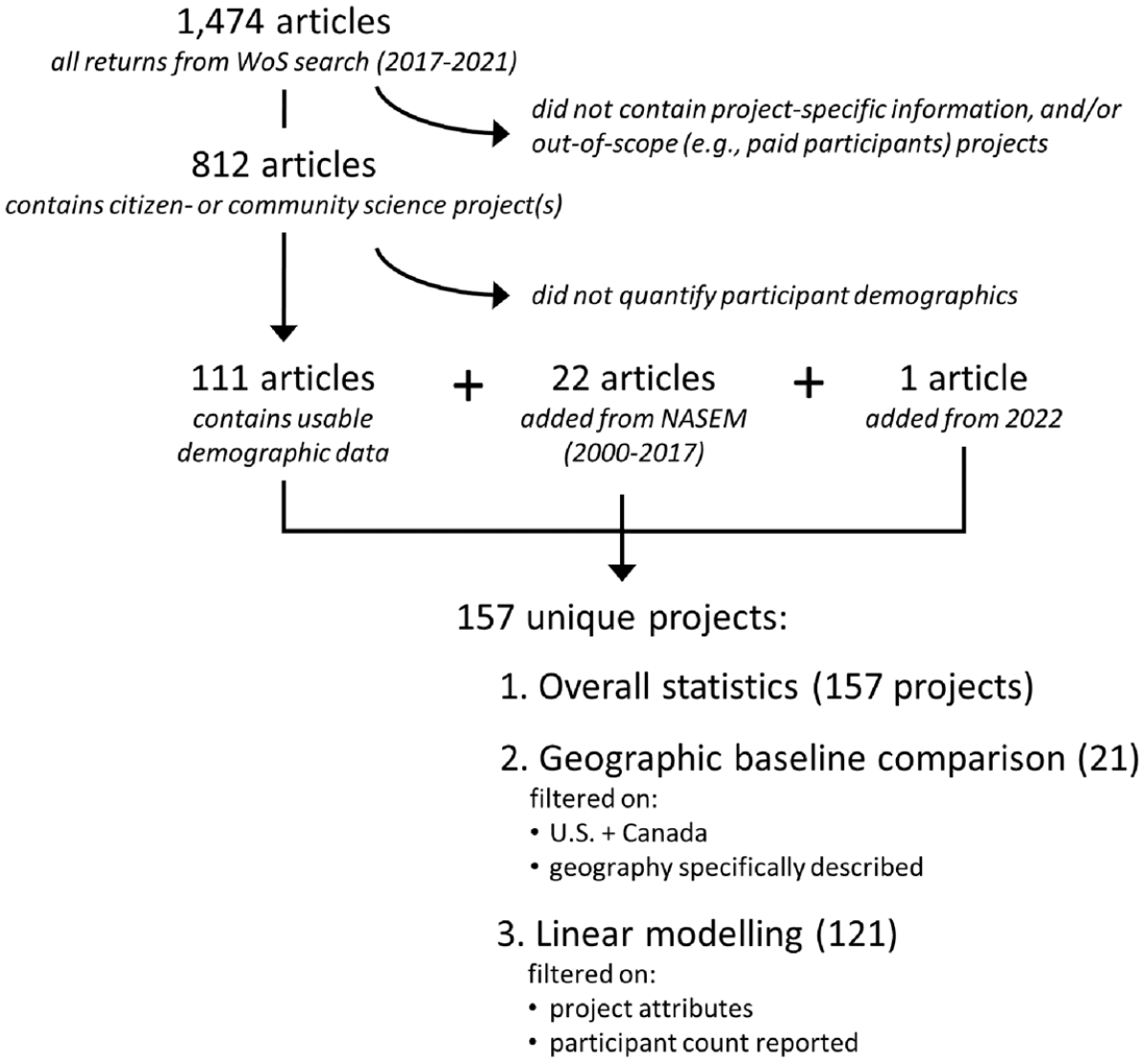


Figure 1. Counts of articles and projects, demonstrating effects of filtering by project attribute, reported statistics, and geographic location on sample size for each analysis.

To determine whether the publication of papers reporting demographic data of citizen science participants is growing proportionately with citizen science publication as a whole, we compared the publication rate of citations included in this paper with that of citations returned by a WOS topic search for "citizen science" over the years 2000–2021. Because of the rapid rise in

publication within the field we used a log-scale, offset by 1.001 to avoid undefined values, and we assessed fit with R^2 values.

Demographic data

We collected data on any demographic variable we came across in the papers. In total, we collected information on 28 variables (Supplemental File 1: Supplemental Table 1), representing 10 broad demographic categories. Where possible, we coalesced slightly divergent formulations into a single category. For example, mean and median age had a correlation coefficient of R^2 = 86.7% (n = 7 projects), allowing us to combine these into a single category: "central tendency" of age.

Twelve projects reported diverse gender categories (including transgender, gender diverse, other, mixed, non-conforming, and/or non-binary). As these participants made up a small percentage of each participating population (0.1–1%; n = 12), and most studies (n = 117) only reported gender as binary male/female, we maintained gender as binary within our analysis and reported % female. For papers that only presented gender data as % male (n = 14), we calculated % female as % female = 1–% male, assuming that people that were not male were overwhelmingly female. This decision increased our gender sample size to 143 projects.

To estimate retiree data from projects that did not explicitly report percent retiree but did provide age-binned participation, we assumed that 65% of participants age 60+ (i.e., 60 or over) and 90% of participants age 65+ were retirees. When projects reported both the number of people 60+ and the number of people 65+, we used the latter. We assumed that all ages within a reported age bin had an equal probability of occurrence (e.g., if 40% of the participating population was age 55–

64, each year represents 4% of the total participant population), and calculated percent of the population aged 60+ or 65+ accordingly. We verified our estimation assumptions with the 10 papers reporting information on percent retirees as well as age-binned participation (R^2 = 94.9%; Supplemental File 1: Supplemental Figure 1). Including our estimation of retirees increased the number of projects for which we had retiree data from 14 to 65.

Finally, we constrained all analyses to variables that were reported consistently across projects and for which we had data from 15+ projects. This decision reduced our dataset to five demographic variables (Supplemental File 1: Supplemental Table 2), including: *gender* (presented as % female), *race/ethnicity* (recorded variously as the proportion of participants in one or more of the USA census categories for under-represented groups, for minorities, and for the majority group, and presented as % white not including Hispanic), *age* (collected as a central tendency measure, and presented as a grand mean and standard deviation), *estimated retiree participation* (presented as % estimated retirees), and *education* (presented as % with a graduate or professional degree because data for these degrees were reported more consistently across papers than data for associate's/ bachelor's degrees).

Project attribute data

We collected information on six project attributes as possible explanatory factors of project demographics. Project information was drawn primarily from the publication(s), with additional information from project websites as needed and available. For categorical coded variables (e.g., project focus), two coders independently categorized each project. Disagreements (~6.5% of the total) were resolved by a third coder followed by consensus through discussion.

Sample size: number of participants for which demographic data were reported. If an approximation was given, e.g., “about 90,” we used that number.

Project location: reported geographic area over which participants were recruited. In some cases, this area was national (e.g., the USA) whereas in others it was more local (e.g., Petersburg, Virginia).

Geo-scheme: We also used project location to group projects according to the United Nations sub-region geo-scheme (https://unstats.un.org/unsd/methodology/m49/). This broad geographic scale ensured we had multiple projects in each area category and allowed inclusion of location as a variable in our models. Projects that took place over multiple geo-schemes (21 projects, or 13.4%) were included in one “mixed” geo-scheme. These projects were primarily online, worldwide projects (17 projects, or 81.0% of the mixed category), such as Foldit (Curtis 2018).

Project access: whether participants accessed the project and collected/analyzed data solely online (e.g., CosmoQuest- Gugliucci et al. 2021; 21 projects or 13.4%) or performed some hands-on activity (e.g., Audubon’s Christmas Bird Count, Allf et al. 2022; 136 projects or 86.6%).

Project focus: disciplinary focus of the project categorized as: physical science, health, biodiversity, other, or unknown (descriptions in Supplemental File 1: Supplemental Table 3).

Project year: Because the actual year of data collection could be determined for only 67.5% of projects (n = 106), whereas the publication year could be determined for all projects, we used publication year as a proxy for data collection year. We made the simplifying assumptions that

publication year is lagged by some unknown quantity relative to the time when the data were collected (for the 106 projects for which data collection year was known, average (mean) time to publication was 2.7 years, standard deviation (SD) = 1.5 years); and that projects with different attributes (e.g., foci) did not vary systematically in publication lag.

To ground our meta-analysis in the fast-growing community of citizen science, we used SciStarter, which "hosts one of the largest online, searchable catalogues of citizen science projects" (Allf et al. 2022), by comparing the frequency distribution across project foci. Because SciStarter does not categorize project foci the same way we did, we mapped SciStarter categories to our foci (Supplemental File 1: Supplemental Table 4).

Geographic comparison

For projects located in the USA and/or Canada and for which the paper gave an explicit "recruitment" geography, we used project location as the basis to collect demographics on the population reported in census data, which we refer to as the "geographic comparison." We restricted our sample set to the USA and Canada because we could readily access federal census data (USA: https://www.census.gov/; Canada: https://www12.statcan.gc.ca/census-recensement/index-eng.cfm). In cases where census data were included in the study, we used these data (n = 5 projects). For all other projects we created a theoretical "participant pool" of all residents within the project location(s). While our pool is not necessarily the target audience for a particular project, it is not unreasonable to assume that most projects in the area would be accessible to people within the participant pool. We collected census data from the nearest reported year to publication year as follows: age (median), gender (% female), estimated retiree

participation (% estimated retiree), education (% with graduate or professional degree), and race/ethnicity (USA: % white one race. Canada: % not a visible minority). When projects took place across the USA and Canada, we calculated averages weighted by population size.

Analysis

Geographic baseline

To determine whether citizen science participants are representative of their relevant geographies, we compared the reported participant demographic data with the census-derived geographic comparison representing the theoretical pool of participants. Gender, race/ethnicity, estimated retiree participation, and education data were collected as a percent, so we calculated percent difference (expected - observed). Age data were collected as a numerical value, so we calculated the residual (residual = expected – observed).

Linear modeling

We used linear modeling to explore the extent to which variation in demographics are explainable by project attributes. Because of the paucity of published projects in the early part of our sample set (Supplemental File 1: Supplemental Figure 2; 120 out of the 157 projects [76.4%] were published between 2017–2021), we restricted our modeling analyses to data published in the years 2011–2021. Furthermore, we combined the 2011–2016 data into one "year" of data, as the total number of projects in this period was small (n = 11), and doing so did not affect which models were selected as the top model (Supplemental File 1: Supplemental Table 5). This decision increased our sample size to 131 projects. Ten projects (6.4% of the 157 projects) were subsequently dropped because of: "unknown" (2 projects) or "other" (8 projects) focus, leaving

121 projects for our modeling effort (77.1% of the 157 projects, with 107,884 unique within-project participants).

Age central tendency and percent estimated retirees were highly correlated (R^2 = 91.5%; n = 4 projects, Supplemental File 1: Supplemental Figure 3). Because age data were often presented as an average without a measure of variation (n = 14), age binning was variable across publications and projects (excepting the tendency to report bins with a break at age 60 and/or age 65), and age central tendency data (n = 14) were not as prevalent as percent estimated retiree (n = 65), we elected to use percent estimated retiree in our models.

We modeled the relationship between each demographic variable and the fixed effects of time (represented by publication year), project access, and project focus. As our response data consisted of the number of participants, Y, identified as being in a demographic group out of a total sample, n, within each program, we modeled our data using quasi-binomial Generalized Linear Models (GLMs; Zuur et al. 2009) with a logit link function, specifying that fixed effects altered the probability [of a demographic group], p, within a program. We used quasi-binomial GLMs instead of binomial GLMs as initial examination of our data indicated that our response exhibited overdispersion relative to a standard binomial distribution. Geo-scheme was added as a random effect, to account for potential variation in demographics due to location.

Models took the general form of:

$$Y_i \sim QuasiBinomial\ (n_i, p_i)$$

$$logit\ (p_i) = \beta_0 + \beta_{accessi} + \beta_{focusi} + \beta_{yeari} + \alpha_{geo\text{-}schemei}$$

Where β_{access} is the categorical effect of project access, β_{focus} is the categorical effect of project focus, β_{year} is the trend coefficient through time, and $\alpha_{geo\text{-}scheme}$ is the random effect of geo-scheme for observation i, which is modelled as a normal distribution on the scale of the link function

$$\alpha_{geo\text{-}schemei} \sim N(0, \sigma_{geo}).$$

Because studies with lower participant sample sizes may be more likely to exhibit biased demographics by chance (Bolthausen and Wüthrich 2013), we weighted each project in the model by participant count. To ensure projects with large participant counts did not "swamp" smaller projects, we capped the maximum weight (i.e., participant count). We explored caps of 75, 100, and 125. Because this range of potential caps did not affect which models were selected (Supplemental File 1: Supplemental Tables 6 and 7), we used 100 as our maximum weight cap.

With quasi-binomial GLMs, the Akaike Information Criterion is not defined; therefore, we selected variables using the drop1 function in R (Zuur et al. 2009). Drop1 sequentially deletes each term from the model, recalculating model deviance without that term. These deviance estimates follow an F distribution and can be used to find an F-statistic and associated p-value as an indication of variable importance. For each demographic variable, we started with the full model with all explanatory variables included and no interaction terms. If all variables had a p-value < 0.05, we used that as the best model. If any variables had a p-value ≥ 0.05, we removed the variable with the largest P-value, reassessing and repeating until all remaining variables had a p-value < 0.05. We validated the final model by plotting the deviance residuals against predicted

response values and independent variables to check for violations of model assumptions (Zuur et al. 2009).

All modeled analyses were conducted in R (version 4.2.1, www.r-project.org, accessed 6 Aug 2022). To create the models, we used the package lme4 (Bates et al. 2015). We used the confint function to calculate confidence intervals for each predictor (base R; with level = 0.95, using a profile method and 1000 simulations). To display the relationship between the predictors and the demographic variables, we used the package jtools (Long 2019).

Results

Among the citations returned by our search terms, publications containing citizen science participant demographic information increased exponentially over time, exceeding the rate of publications returned in a WOS topic search for "citizen science" (Figure 2). This suggests that studies reporting demographic data (~7.5% of all citations reviewed as part of this paper) are on the rise. However, the percentage of papers reporting participant demographics is still very small; for 2021, ~2.2%.

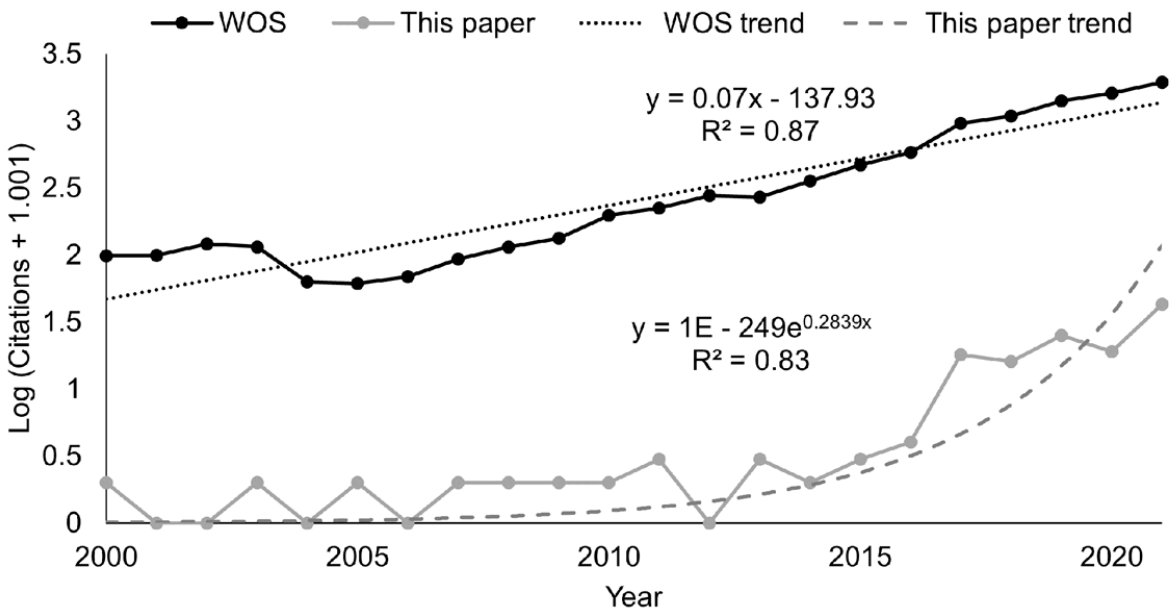


Figure 2. Log (citations, offset by 1.001 to avoid undefined values) for citations returned by a Web of Science topic search for "citizen science" ("WOS") and citations included in this paper over the years (2000–2021) ("This paper"). Equations and R2 values for lines of best fit ("trend") are also displayed.

To understand the degree to which we had captured citizen science, we compared project foci from this study to those found in SciStarter (1,599 projects, data collected on 25 Sept 2022; Supplemental File 1: Supplemental Table 4). Both datasets were similarly, and minorly, online (SciStarter 11.9%; this study 13.2%). Biodiversity was dominant for both our project set (57.7% of projects with a known focus) and in SciStarter (54.9%), followed by projects centered on human health (SciStarter 26.5%; this study 27.6%), and physical science (SciStarter 5.6%; this study 8.3%).

Geographic comparison

To assess project reach into the relevant geographic communities, we used the subset of projects for which geographic location was specifically reported, restricting our analysis to the USA and

Canada, and using federal census data to create a theoretical pool of participants within each project's geographic reach. Although this resulted in a small sample set (n = 21 projects), it does provide a cleaner comparison than wider (e.g., USA-wide) geographies, which may be less/not relevant to smaller projects.

Within this subset, participants in about half the projects were more female than would be expected from their geographic comparison (10/17, or 58.8%, n = 17; Figure 3*a*), and the mean for these projects was slightly above no difference, with high variability (participant count weighted mean: 2.4%, unweighted: 3.5%, unweighted SD = 18.8%). Participants in all projects were more white than their geographic comparison (n = 6, Figure 3*a*), with a mean 29.3% (unweighted) to 30.5% (weighted) above no difference (unweighted SD = 13.2%), suggesting a statistical and/or social bias against non-white citizen science participants, albeit with a very small sample size. All but one project (4/5 or 80%, n = 5; Figure 3*a*) had more highly educated participants (i.e., with a graduate or professional degree) than would be expected from the geographic comparison, leading to a mean at 23.0% (unweighted) to 33.9% (weighted) (unweighted SD = 15.8%) above no difference.

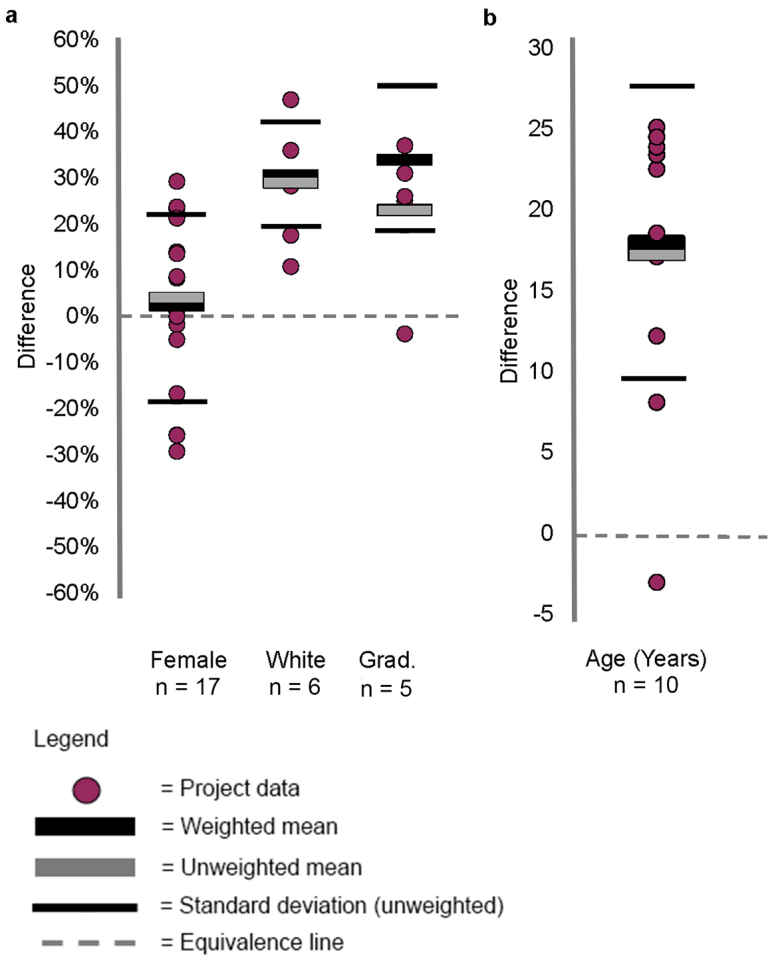


Figure 3. Differences between citizen science and geographic comparison (census) demographics, reported as percent difference (a: female [% female], white [% white], grad. [% with graduate/professional degree]) or as the residual (b: age [years]). Values on the line at zero indicate no difference between citizen science and the geographic comparison, while values above the line at zero indicate project participants were more female, white, educated (a), or older (b) than the geographic comparison (and the converse is true for values below the line).

Participants from all but one project were older than would be expected from the geographic comparison (9/10 or 90%; Figure 3*b*) with a mean 17.2 (unweighted) to 18.5 (weighted) years above no difference (unweighted SD = 9.1). Because census data includes all residents in the geographic pool, including youth, this result can be interpreted to suggest that citizen science

projects primarily target adults. Further evidence supporting this interpretation includes that some papers stated that they limited participation in their demographic surveys to people either 16 and over (n = 3) or 18 and over (n = 19) due to constraints of including children as research subjects. Estimated retiree data were excluded owing to small sample size.

Modeling results

General

To explore potential within-project drivers of demographic biases (statistical and social), we modeled the influence of project attributes on demographics using quasi-binominal models. This analysis was restricted to projects with identifiable attributes and participant count (n = 121 projects). Selected models for each demographic variable contained 1–2 of our predictor variables (Table 1), and all predictor variables were selected in at least one model. Here we report demographic statistics over all projects, as well as model-specific results.

Table 1: Results of quasi-binomial models of the proportion of individuals in citizen science projects who fell into different demographic categories (gender- % female, race/ethnicity- % white, education- % with a graduate or professional degree, and estimated retiree participation- % estimated retiree). For each coefficient, variable estimates, standard error (SE), the lower (2.5% CI) and upper (97.5% CI) confidence intervals and P-values for each variable in the selected models are shown. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Significance is assessed against hands-on for project access and biodiversity for focus.

Coefficient	Estimate	SE	2.5% CI	97.5% CI	P-value
Female (N = 115)					
Intercept	-517.18	88.72	-693.08	-344.92	5.46×10^{-8} ***
Year	0.26	0.04	0.17	0.34	5.59×10^{-8} ***
Project Access					1.88×10^{-5} ***
Online	0.62	0.13	0.36	0.89	1.07×10^{-5} ***
White (N = 14)					
Intercept	2.96	0.38	2.29	3.80	8.23×10^{-6} ***
Focus					0.00 **
Health	-1.24	0.38	-2.09	-0.56	0.01 **
Grad (N = 33)					
Intercept	-0.75	0.12	-0.98	-0.53	2.91×10^{-7} ***
Project Access					0.00 ***
Online	0.56	0.12	0.32	0.80	9.05×10^{-5} ***
Retiree (N = 53)					
Intercept	-0.75	0.11	-0.96	-0.54	7.70×10^{-9} ***
Focus					1.66×10^{-12} ***
Health	-1.48	0.14	-1.76	-1.21	2.43×10^{-14} ***
Physical Sciences	-0.97	0.47	-2.00	-0.12	0.0444 *

Gender

In contrast to the results from the geographic comparison (Figure 3*a*), male participants dominated slightly over all projects with (binary) gender information (n = 143 projects, 45.6% female, compared with 51.3% female for USA adults, U.S. Census Bureau 2020; Supplemental File 1: Supplemental Table 2). In our models, gender was best explained by publication year and access type (Table 1). Projects published more recently had a higher proportion of women (Figure 4). Women were also significantly overrepresented in online relative to hands-on projects (Figure 5*a*).

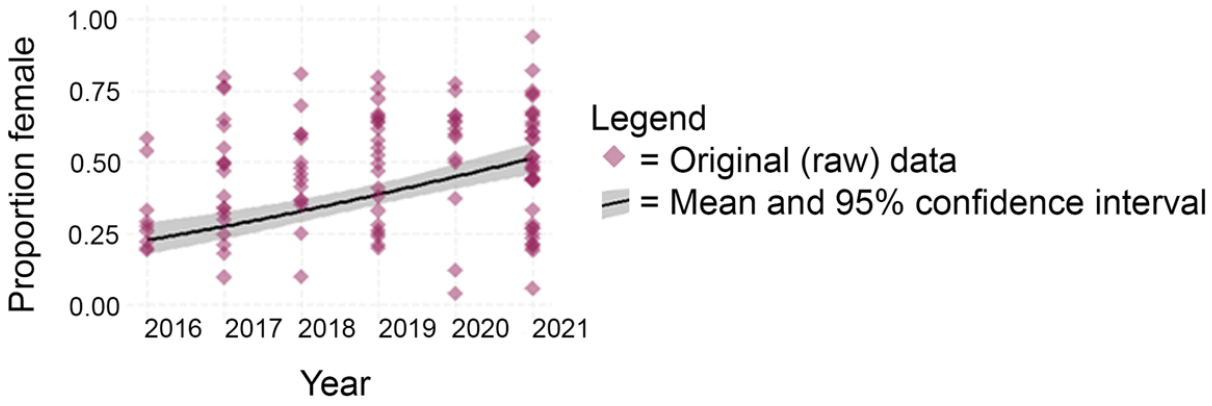


Figure 4. Fitted values (mean and 95% confidence interval) for the year variable from the selected model for proportion of citizen science participants that were female, along with original (raw) data. Predictions are calculated based on varying the year while holding other predictors constant. The year 2016 represents data from 2011–2016.

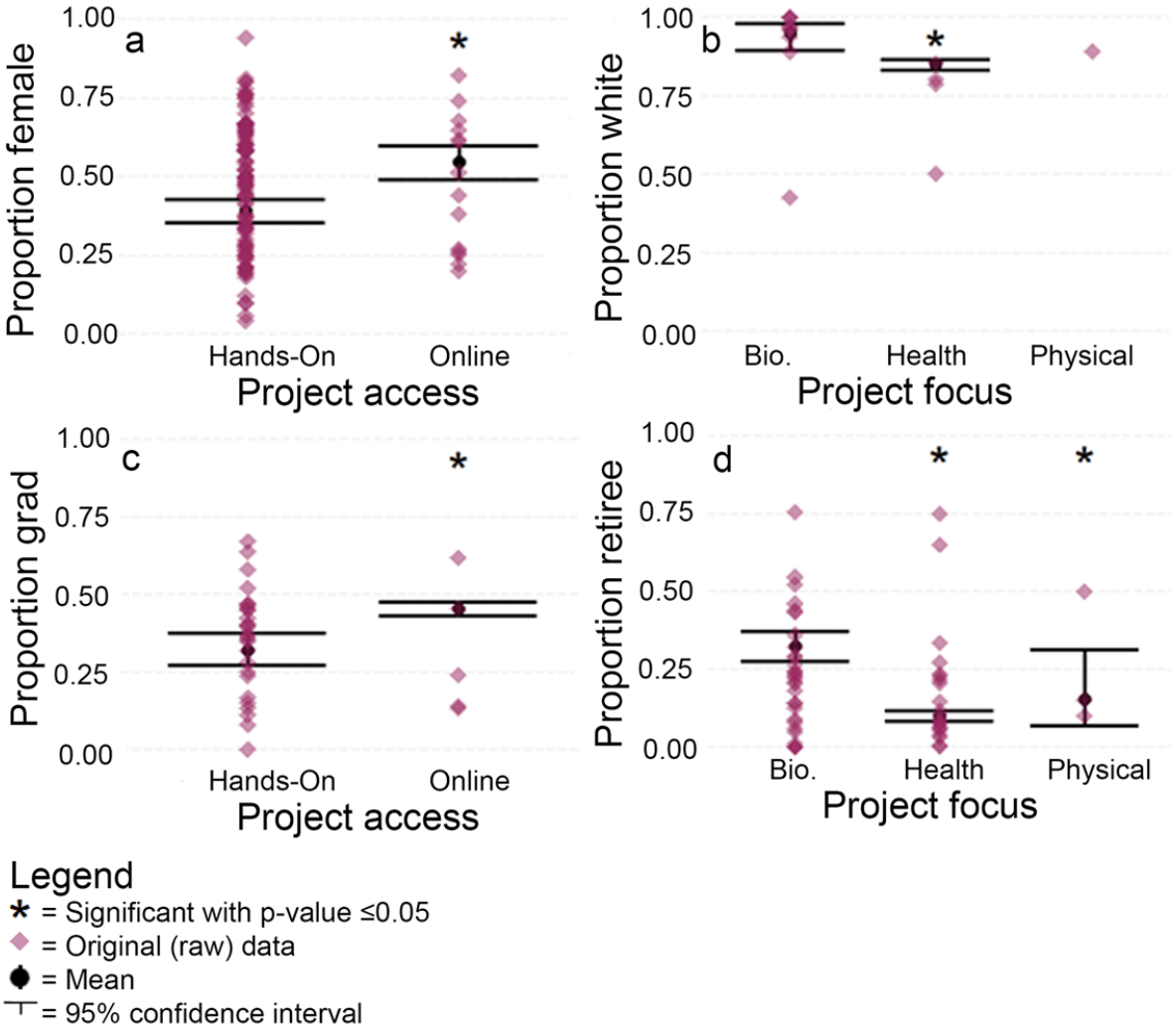


Figure 5. Fitted values (mean and 95% confidence interval) for the selected model for each demographic variable (a = proportion of citizen science participants that were female, b = proportion of participants that were white, c = proportion of participants with a graduate/professional degree, and d = proportion of participants that were estimated to be retired), along with the original/raw data. Predictions are calculated based on varying the predictor shown while holding other predictors constant. For project focus, "Bio." = biodiversity and "Physical" = physical sciences. Significance is assessed against hands-on for project access and biodiversity for focus.

Race/ Ethnicity

White non-Hispanic participation was quite high (88%, compared with USA adults at 64%; Jones 2022), although the reporting sample was low (n = 17 projects, Supplemental File 1: Supplemental Table 2). Only one online project and one physical science project reported participant race/ethnicity data, respectively, which precluded modeling. However, we did include these points in graphic comparisons (i.e., Figure 5*b*).

Given those omissions, race/ethnicity was best explained by project focus. Projects with a health focus had a higher non-white participation (~17%) than those with a biodiversity focus (~8%, Figure 5*b*), but these values are still quite low when compared to the USA adult non-white population (35.9%, Jones 2022). The single physical science project in our sample reported 90% white participants.

Education

The average proportion of the participant population with an advanced degree (graduate or professional) over the entire dataset was substantially higher (n = 40 projects, weighted: 43%, unweighted: 35%, Supplemental File 1: Supplemental Table 2) than for the USA population aged 25+ (USA: 13%, America Counts Staff 2021). Education level was best explained by project access, with online projects reporting a significantly higher proportion of people with advanced degrees than hands-on projects (Figure 5*c*).

Age and estimated retiree participation

The mean weighted age across all participants in all projects for which age was directly reported (n = 52) was 48.1, ~9 years older than for the USA as a whole (38.8; U.S. Census Bureau 2022;

Supplemental File 1: Supplemental Table 2). Participants in projects with a biodiversity focus had the highest mean age, while health projects had the lowest (Table 2).

Table 2: For each demographic variable (gender [% Female], race/ethnicity [% White] education [% with graduate/professional degree] estimated retiree participation [% retiree] and age [years]), the project count, mean (%), and SD (% points) for project focus and access. NA indicates we did not calculate descriptive statistics because only one or all but one of the projects were in the sub-category.

		Count	Mean (%)	SD (% pts)
% Female				
Project Focus	Biodiversity	80	45.7	19.4
	Health	41	52.8	21.5
	Physical Science	12	30.4	11.1
Project Access	Online	19	45	21
	Hands-on	124	45.5	20.8
% White				
Project Focus	Biodiversity	12	92.0	15.9
	Health	4	73.5	15.9
	Physical Science	1	NA	NA
Project Access	Online	1	NA	NA
	Hands-on	16	NA	NA
% Grad Degree				
Project Focus	Biodiversity	26	38.1	15.5
	Health	11	25.6	17.1
	Physical Science	2	NA	NA
Project Access	Online	5	31.7	21.3
	Hands-on	35	35.8	16.3
% Retirees				
Project Focus	Biodiversity	36	23.5	17.8
	Health	21	17.7	19.7
	Physical Science	5	16.2	19.6
Project Access	Online	8	12.2	18.5
	Hands-on	57	22.3	18.7
Age				
Project Focus	Biodiversity	32	50.9	8.5
	Health	11	39.4	9.3
	Physical Science	6	48.4	8.1
Project Access	Online	6	41.4	4.3
	Hands-on	47	49	9.6

Across all projects, estimated retirees averaged 21% of all participants (n = 65 projects, Supplemental File 1: Supplemental Table 2), slightly higher than USA statistics when youths are excluded (17.4% of USA adults in 2019 were categorized as retired; Administration on Aging 2021). Estimated retiree participation was best explained by project focus. Biodiversity projects had a significantly higher proportion of estimated retirees than health projects (Figure 5*d*).

Discussion

To our knowledge, this paper is the most comprehensive meta-analysis of citizen science participant demographics to date, incorporating 157 projects and collectively representing 151,854 unique within-project participants. The strongest signals we found across the English language, peer-reviewed literature that reported demographics of participants is three-part: high participation from educated, white adults; a rising tide of female participation; and project focus influences participant demographics. We found that health-focused projects attracted a younger, relatively more diverse participant base than biodiversity projects, which were older and whiter.

These patterns are echoed in the STEM workforce, at least in the USA. Women make up only ~34% of USA STEM workers (Okrent and Brent 2021). Blacks/African Americans, Hispanic/Latinos, and American Indians/Alaska Natives constitute ~30% of employed USA workers but only ~23% of STEM workers (Okrent and Brent 2021). The USA volunteering population also displays overrepresentation of majority demographic groups. Schlachter (2021) reported that volunteers were disproportionately middle-aged, white non-Hispanic females with a college degree. If there are demographic biases in both the STEM workforce and the

volunteering population, citizen science may be at the intersection of those two groups, and display some of the same (social and/or statistical) biases.

In their citizen science meta-analysis, Theobald et al. (2013) found that only ~12% of the 388 biodiversity projects they reported on published data in the scientific literature. Our study indicates that single-digit percentages of English language citizen science papers include participant demographics. In a recent survey of 140 German citizen science projects, Moczek et al. (2021) found that only 13–26% of projects reported information on participant age, gender, or education level. Thus, the vast majority of projects, and certainly participants, remain invisible. As Pateman et al. (2021) opine: "practitioners [need to] document and publish participant demographics." In fact, even in this study of 157 projects, sample sizes for certain variables (e.g., race/ethnicity, n = 30 out of all projects) were quite small. It is therefore important to consider whether the results found in this research are generally applicable.

Trends in gender of citizen science participants

Findings with respect to gender have been mixed across the literature. Pateman et al. (2021) found males in Great Britain were more likely to participate in citizen science than females. In a broad review of published citizen science projects, NASEM (2018) also found males predominating. Our data suggest wide variation in gender (Figures 3a, 4, 5a), and that the proportion of women participants overall is on the rise (Figure 4). However, we also found evidence to suggest that women were persistently underrepresented in particular project foci—namely, in the physical sciences (Table 2). Male overrepresentation appears to be particularly true in astronomy, where participants in the five projects reported on herein (including both

hands-on and online) were only 19–26% female. These disparities can be even more stark. For instance, Curtis (2018) found only 13.2% of participants were female in the physics-based online game Quantum Moves (N = 674).

Why might this be true? Cooper and Smith (2010) suggest gender patterns may be related to the degree to which citizen science project activities are competitive or authoritative (male dominated) versus supportive or participatory (female dominated). Paleco et al. (2021) reported more men participating in technical-related events (e.g., data quality), and more women participating in social science-related events. These studies suggest that gender-based participation may have roots in project focus, and/or in project activities or organization—findings that are echoed in USA academic science graduate degrees, and particularly in mathematics, computer sciences, engineering, physical sciences, and geosciences, which all continue to trend male, despite the continued accession of women (NCSES 2023). However, we also believe that while the rates predicted in our modeling are probably an overestimate (e.g., Figure 4), the trend towards greater female participation over time is real. This suggests that citizen science may be able to push boundaries, and further equality, at least in some demographic groups.

Educated white adults

Our analysis indicated an overrepresentation of graduate/professional degrees among citizen science participants, at least relative to the USA average, with extremely high percentages in two online projects (Citizen Sort: 61.8%, Tang and Prestopnik 2019; The COVID-19 Citizen Science Study: 45.4%, Lin et al. 2021). Allf et al. (2022) reported that participants electing to disclose

their information as SciStarter members (n = 423) were also highly educated (at 53% with graduate/professional degrees). That citizen science participants trend toward college and higher education graduates is a result repeatedly demonstrated across the citizen science literature (e.g, Evans et al. 2005; Curtis 2018; Haklay 2018).

Studies of regional-to-national populations have indicated that citizen science is practiced more often by people who are white and/or those who also enjoy a moderately high socioeconomic status (i.e., middle class; Pateman et al. 2021; Mahmoudi et al. 2022). This finding is repeated in surveys across citizen science populations, which also indicate an overrepresentation of older, white non-Hispanics (Merenlender et al. 2016; NASEM 2018; Allf et al. 2022).

Our overall findings reinforce literature results: a propensity, even preponderance, of educated, white adults (Figures 3, 5). Our modeling results (n = 121 projects) suggest that the focus of the project may explain some of this statistical and/or social bias, in that health projects had a higher proportion of non-white (Figure 5*b*) and younger (Figure 5*d*) participants as compared with biodiversity projects. One reason may be that health projects often include the intersection of public health, environmental justice, and geographic community (e.g., the Flint water crisis: Pauli 2019; STI/HIV research in LGBTQ+ communities: Mann-Jackson et al. 2021), and in that sense are centering in the margins (Cooper et al. 2021). Of course, the inverse is also likely true, that biodiversity projects differentially attract older, more affluent, white participants—those with the luxury of time to participate.

There are, however, examples of participant-diverse biodiversity-focused citizen science (e.g., Open Air Laboratories network, Paleco et al. 2021), which may be explained by intentional

partnership with community organizations with a mission to support underserved and marginalized groups (Open Air Laboratories 2013). Collectively, these studies suggest that it may be possible to diversify participation in citizen science, and specifically increase non-white participation through purposeful recruitment, including reaching out directly to communities, aligning the project with community interests and goals, and creating projects that do not require significant amounts of time and/or a specific schedule.

Social stratification and solution strategies

Underrepresentation in citizen science, as for science more broadly, is likely driven by a combination of multi-faceted, intersectional factors. Barriers to participation by those who are not white, less wealthy, and/or less highly educated include: time constraints such that citizen science is not a priority (Merenlender et al. 2016; Domroese and Johnson 2017; Walker, Smigaj, and Tani 2021), the economic constraint of having to work in lieu of days off (Evans et al. 2005; Merenlender et al. 2016; Walker, Smigaj, and Tani 2021), and/or feeling unwelcome in scientific spaces (Evans et al. 2005; Merenlender et al. 2016). Dawson (2018) explored why low-income, minority populations may be broadly excluded from science, using focus groups and interviews. A combination of factors emerged from interviewees, including awareness, feelings of powerlessness, and predetermined exclusion as a function of race/ethnicity. Interviewees imagined participants in science being not like them (in particular, they imagined participants as people who were white, with free time, and with a comfortable income). Dawson (2018) describes these as forms of social stratification within which doing begets acquisition of cultural capital which promotes more doing, whereas not doing promotes exclusion. Many studies have identified the importance of prior participation in citizen science as a significant factor predicting

participation (NASEM 2018; Pateman et al. 2021; Allf et al. 2022). The question becomes, how can public participation in informal science activities, including but not limited to citizen science, become more representational of the geographic public from which participants are drawn?

Strategies that have been suggested for diversifying citizen science, and specifically to increase minoritized and marginalized groups, include: constructing authentic community science centered within the traditions of social and environmental justice (Pandya 2012; Cooper et al. 2021); framing citizen science around the location, language, norms, and priorities of communities (Chesser, Porter, and Tuckett 2020; Pierce 2022); reducing barriers to participation by providing both online and hands-on ways to participate (Puhan et al. 2018; Paleco et al. 2021); incorporating multiple kinds of knowledge (Wallerstein and Duran 2006; Pandya 2012,); providing role models that are themselves underrepresented (Hite et al. 2019); and using recruitment strategies specific to diverse audiences (Brouwer and Hessels 2019; West, Pateman, and Dyke 2021). We point out that significant demographic shifts in the "citizen and community science population" are most likely to occur when place-based, community science centered in the margins increases everywhere, at the same time that large-scale citizen science projects realize a more diverse participant base. We posit that the latter may require the scientific community engaged in designing public data collection projects to become more accepting of non–college credentialled participants (Burgess et al. 2017), and more willing to take the time and make the effort to honestly engage individuals and groups on their terms (Harris et al. 2021).

Conclusion

This study indicates that citizen science has yet to realize its promise of equal and equitable representation, at least as can be captured in peer-reviewed English language publications. To thoroughly investigate the complex, intersectional topic that is diversity in citizen science, more demographic data needs to be published. Regardless, increasing participant diversity has benefits both for marginalized groups (e.g., Wallerstein and Duran 2006; Bonney et al. 2016) and for science (e.g., Blake, Rhanor, and Pajic 2020; Paleco et al. 2021). Thus, we suggest that all projects commit to the goal of realizing representation equal to the demographics of their geographic distribution, and further, that the big tent of citizen science (as discussed in Cooper et al. 2021) collectively works to welcome locally based, community projects created by and for those most marginalized from science.

Acknowledgements

We would like to thank the many scientists and citizen science participants represented in this paper, along with Jennifer Ruesink, Michael Pocock, Abigail Swann, and anonymous reviewers for providing critical feedback.

Funding information

JKP funded by the Lowell A. and Frankie L. Wakefield Endowment for Ocean Fishery Sciences.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

JKW contributed to the conception, data collection, analysis, visualization, and writing. JKL contributed to the conception, data collection, and writing. MZS contributed to the data collection, data analysis, and writing. JCW contributed to the data collection and writing. JKP contributed to the conception, data collection, data analysis, and writing.

References

Allf, BC, Cooper, CB, Larson, LR, Dunn, RR, Futch, SE, Sharova, M and Cavalier, D. 2022. Citizen science as an ecosystem of engagement: Implications for learning and broadening participation. *BioScience*, 72(7): pp.651–663. DOI: https://doi.org/10.1093/biosci/biac035

America Counts Staff. 2021. *About 13.1 percent have a master's, professional degree or doctorate*. United States Census Bureau.

Bates, D, Mächler, M, Bolker, B and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 71(1): 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Blake, C, Rhanor, A and Pajic, C. 2020. The demographics of citizen science participation and its implications for data quality and environmental justice. *Citizen Science: Theory and Practice*, 5(1). DOI: https://doi.org/10.5334/cstp.320

Bolthausen, E and Wüthrich, MV. 2013. Bernoulli's law of large numbers. *ASTIN Bulletin: The Journal of the IAA*, 43(2): 73–79. DOI: https://doi.org/10.1017/asb.2013.11

Bonney, R, Phillips, TB, Ballard, HL and Enck, JW. 2016. Can citizen science enhance public understanding of science? *Public understanding of science*, 25(1): 2–16. DOI: https://doi.org/10.1177/0963662515607406

Brouwer, S and Hessels, LK. 2019. Increasing research impact with citizen science: The influence of recruitment strategies on sample diversity. *Public Understanding of Science*, 28(5): 606–621. DOI: https://doi.org/10.1177/0963662519840934

Burgess, HK, DeBey, LB, Froehlich, HE, Schmidt, N, Theobald, EJ, Ettinger, AK, HilleRisLambers, J, Tewksbury, J and Parrish, JK. 2017. The science of citizen science:

Exploring barriers to use as a primary research tool. *Biological Conservation*, 208: 113–120. DOI: https://doi.org/10.1016/j.biocon.2016.05.014

Chesser, S, Porter, MM and Tuckett, AG. 2020. Cultivating citizen science for all: Ethical considerations for research projects involving diverse and marginalized populations. *International Journal of Social Research Methodology*, 23(5): 497–508. DOI: https://doi.org/10.1080/13645579.2019.1704355

Cooper, CB, Hawn, CL, Larson, LR, Parrish, JK, Bowser, G, Cavalier, D, Dunn, RR, Haklay, M, Gupta, KK, Jelks, NTO and Johnson, VA. 2021. Inclusion in citizen science: The conundrum of rebranding. *Science*, 372(6549): 1386–1388. DOI: https://doi.org/10.1126/science.abi6487

Cooper, CB and Smith, JA. 2010. Gender patterns in bird-related recreation in the USA and UK. *Ecology and Society*, 15(4). DOI: https://doi.org/10.5751/ES-03603-150404

Curtis, V. 2018. *Online citizen science and the widening of academia*. Cham, Switzerland: Palgrave Macmillan. DOI: https://doi.org/10.1007/978-3-319-77664-4

Dawson, E. 2018. Reimagining publics and (non) participation: Exploring exclusion from science communication through the experiences of low-income, minority ethnic groups. *Public Understanding of Science*, 27(7): 772–786. DOI: https://doi.org/10.1177/0963662517750072

Domroese, MC and Johnson, EA. 2017. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation*, 208: 40–47. DOI: https://doi.org/10.1016/j.biocon.2016.08.020

Evans, C, Abrams, E, Reitsma, R, Roux, K, Salmonsen, L and Marra, PP. 2005. The Neighborhood Nestwatch program: Participant outcomes of a citizen-science ecological research project. *Conservation Biology*, 19(3): 589–594. DOI: https://doi.org/10.1111/j.1523-1739.2005.00s01.x

Follett, R and Strezov, V. 2015. An analysis of citizen science based research: Usage and publication patterns. *PloS one*, 10(11): e0143687. DOI: https://doi.org/10.1371/journal.pone.0143687

Gugliucci, N, Bracey, G, Buxner, S, Bakerman, M, Gay, PL, Glushko, A, Southard, H and Breedon Smith, J. 2021. Motivations and Patterns of Engagement of CosmoQuest Participants. *Communicating Astronomy with the Public Journal*, 29: 151.

Haklay, M. 2013. Citizen science and volunteered geographic information: Overview and typology of participation. In: Sui, D, Elwood, S and Goodchild, M (eds.) *Crowdsourcing Geographic Knowledge*. Dordrecht, Netherlands: Springer. pp. 105–122. DOI: https://doi.org/10.1007/978-94-007-4587-2_7

Haklay, M. 2018. Participatory citizen science. In: Hecker, S, Haklay, M, Bowser, A, Makuch, Z, Vogel, J and Bonn, A (eds.) *Citizen science: Innovation in open science, society and policy*. London, England: UCL Press. pp. 52–62. DOI: https://doi.org/10.2307/j.ctv550cf2.11

Harris, L, Garza, C, Hatch, M, Parrish, JK, Posselt, J, Alvarez Rosario, JP, Davidson, E, Eckert, G, Wilson Grimes, K, Garcia, JE, Haacker, R, Horner-Devine, MC, Johnson, A, Lemus, J, Prakash, A, Thompson, L, Vitousek, P, Martin Bras, MP, Reyes, K. 2021. Equitable exchange: A framework for diversity and inclusion in the geosciences. *AGU Advances*, 2(2): e2020AV000359. DOI: https://doi.org/10.1029/2020AV000359

Hite, R, Jones, MG, Andre, T, Childers, G and Corin, E. 2019. Female and minority experiences in an astronomy-based science hobby. *Cultural Studies of Science Education*, 14(4): 937–962. DOI: https://doi.org/10.1007/s11422-018-9884-y

Jones, N, Marks, R, Ramirez, R and Ríos-Vargas, M. 2022. 2020 census illuminates racial and ethnic composition of the country, 21 August 2021. Available at https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html (Last accessed 4 October 2023).

Lin, A, Vittinghoff, E, Olgin, J, Peyser, N, Aung, S, Joyce, S, Yang, V, Hwang, J, Avram, R, Nah, G and Tison, GH. 2021. Predictors of incident SARS-CoV-2 infections in an international prospective cohort study. *BMJ open*, 11: e052025. DOI: https://doi.org/10.1136/bmjopen-2021-052025

Long, JA. 2019. *jtools: Analysis and presentation of social scientific data*. (Version 2.2.2) [R Package]. Available at: https://jtools.jacob-long.com/.

Mahmoudi, D, Hawn, CL, Henry, EH, Perkins, DJ, Cooper, CB and Wilson, SM. 2022. Mapping for whom? Communities of color and the citizen science gap. *UMBC Faculty Collection*.

Mann-Jackson, L, Alonzo, J, Garcia, M, Trent, S, Bell, J, Horridge, DN and Rhodes, SD. 2021. Using community-based participatory research to address STI/HIV disparities and social determinants of health among young GBMSM and transgender women of colour in North Carolina, USA. *Health & Social Care in the Community*, 29(5): e192–e203. DOI: https://doi.org/10.1111/hsc.13268

Merenlender, AM, Crall, AW, Drill, S, Prysby, M and Ballard, H. 2016. Evaluating environmental education, citizen science, and stewardship through naturalist programs. *Conservation Biology*, 30(6): 1255–1265. DOI: https://doi.org/10.1111/cobi.12737

National Academies of Sciences, Engineering, and Medicine (NASEM). 2018. *Learning through citizen science: Enhancing opportunities by design*. Washington, DC: The National Academies Press.

National Center for Science and Engineering Statistics (NCSES). 2023. *Diversity and STEM: Women, Minorities, and Persons with Disabilities 2023*. Special Report NSF 23–315. Alexandria, VA: National Science Foundation.

Okrent, A and Brent, A. 2021. *The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers*. Report NSB-2021-2. Alexandria, VA: National Science Foundation.

Open Air Laboratories. 2013. *Community Environment Report*. London, UK: Open Air Laboratories.

Paleco, C, García Peter, S, Salas Seoane, N, Kaufmann, J and Argyri, P. 2021. Inclusiveness and diversity in citizen science. In: Vohland, K, Land-Zastra, A, Ceccaroni, L, Lemmens, R, Perelló, J, Ponti, M, Samson, R and Wagenknecht, K (eds.) *The science of citizen science*. Cham, Switzerland: Springer. pp. 261–281. DOI: https://doi.org/10.1007/978-3-030-58278-4_14

Pandya, RE. 2012. A framework for engaging diverse communities in citizen science in the US. *Frontiers in Ecology and the Environment*, 10(6): 314–317. DOI: https://doi.org/10.1890/120007

Pateman, RM, Dyke, A and West, SE. 2021. The diversity of participants in environmental citizen science. *Citizen Science: Theory and Practice*, 6(1): 9. DOI: https://doi.org/10.5334/cstp.369

Pateman, RM and West, SE. 2023. Citizen Science: Pathways to Impact and why Participant Diversity Matters. *Citizen Science: Theory and Practice*, 8(1): 50. DOI: https://doi.org/10.5334/cstp.569

Pauli, BJ. 2019. *Flint fights back: Environmental justice and democracy in the Flint water crisis*. Cambridge, Massachusetts: MIT Press. DOI: https://doi.org/10.7551/mitpress/11363.001.0001

Pierce, RL. 2022. Interrupting pathways to health inequities in citizen science health research. *Health Promotion International*, 37(Supplement_2): ii21–ii34. DOI: https://doi.org/10.1093/heapro/daac032

Puhan, MA, Steinemann, N, Kamm, CP, Muller, S, Kuhle, J, Kurmann, R, Calabrese, P, Kesselring, J and von Wyl, V. 2018. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. *Swiss medical weekly*, 148(w14623): w14623. DOI: https://doi.org/10.4414/smw.2018.14623

Schlachter, LH. 2021. *Key Findings from the 2019 Current Population Survey: Civic Engagement and Vounteering Supplement*. Washington, DC: AmeriCorps, Office of Research and Evaluation.

Shirk, JL, Ballard, HL, Wilderman, CC, Phillips, T, Wiggins, A, Jordan, R, McCallie, E, Minarchek, M, Lewenstein, BV, Krasny, ME and Bonney, R. 2012. Public participation in scientific research: a framework for deliberate design. *Ecology and society*, 17(2). DOI: https://doi.org/10.5751/ES-04705-170229

Tang, J and Prestopnik, NR. 2019. Exploring the impact of game framing and task framing on user participation in citizen science projects. *Aslib Journal of Information Management*, 71(2): 260–280. DOI: https://doi.org/10.1108/AJIM-09-2018-0214

Theobald, EJ, Ettinger, AK, Burgess, HK, DeBey, LB, Schmidt, NR, Froehlich, HE, Wagner, C, HilleRisLambers, J, Tewksbury, J, Harsch, MA and Parrish, JK. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, 181: 236–244. DOI: https://doi.org/10.1016/j.biocon.2014.10.021

Thigpen, CL and Funk, C. 2020. *Younger, more educated U.S. adults are more likely to take part in citizen science research*. Washington, DC: Pew Research Center.

Walker, DW, Smigaj, M and Tani, M. 2021. The benefits and negative impacts of citizen science applications to water as experienced by participants and communities. *Wiley Interdisciplinary Reviews: Water*, 8(1): e1488. DOI: https://doi.org/10.1002/wat2.1488

Wallerstein, NB and Duran, B. 2006. Using community-based participatory research to address health disparities. *Health promotion practice*, 7(3): 312–323. DOI: https://doi.org/10.1177/1524839906289376

West, SE, Pateman, RM and Dyke, A. 2021. Variations in the motivations of environmental citizen scientists. *Citizen Science: Theory and Practice*, 6(1): 14. DOI: https://doi.org/10.5334/cstp.370

Wiggins, A and Crowston, K. 2011). From conservation to crowdsourcing: A typology of citizen science. In: 2011 44th Hawaii International Conference on System Sciences, Kauai, HI on 4 January 2011, pp. 1–10. DOI: https://doi.org/10.1109/HICSS.2011.207

Zuur, AF, Ieno, EN, Walker, NJ, Saveliev, AA and Smith, GM. 2009. *Mixed effects models and extensions in ecology with R (Vol. 574)*. New York: Springer. DOI: https://doi.org/10.1007/978-0-387-87458-6

Chapter 3- Supplementary Materials

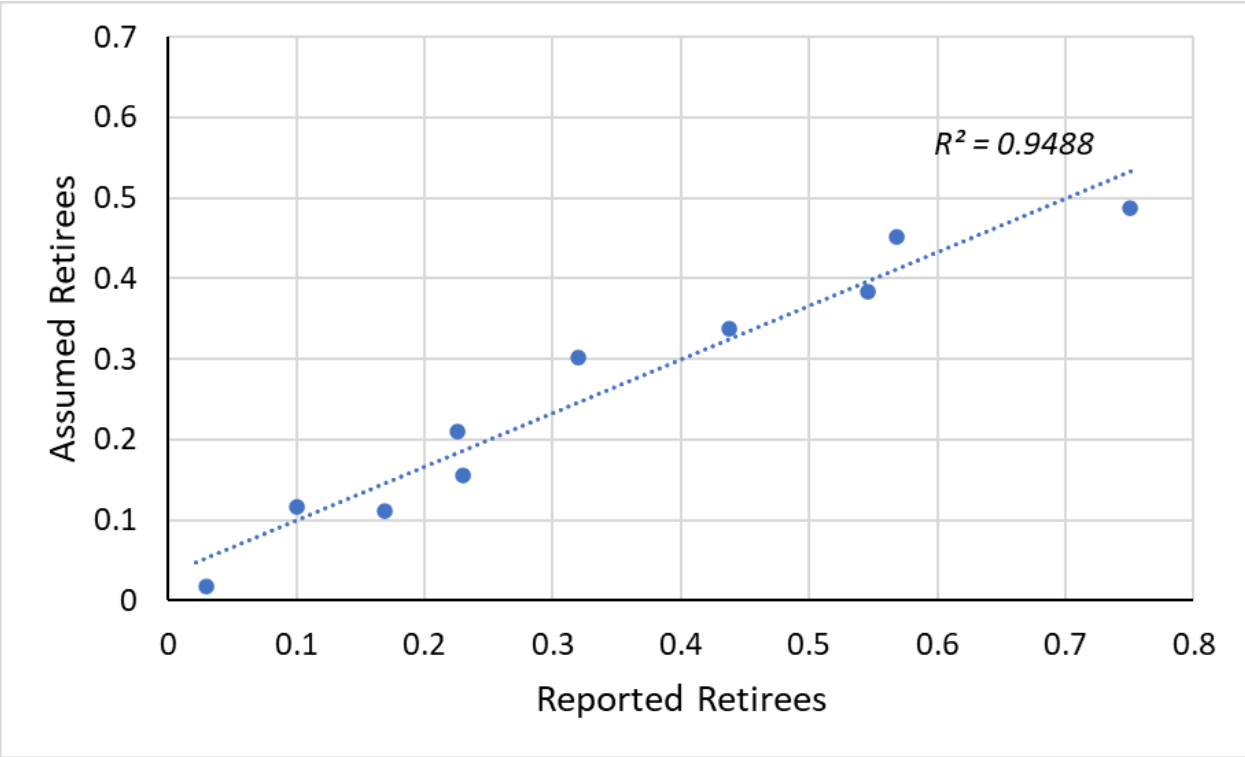
Demographic variables for which we collected data

Supplemental Table 1: Broad demographic categories for which we collected data ("Broad category"), the specific variables within those categories for which we collected data ("Specific variable"), and number of projects that reported that kind of data ("N").

Broad Category	Specific Variable	N
Gender	All data- all data related to gender	144
	Non-Cisgender Male/Female Data- all data for gender categories other than cisgender Male/ Female, including transgender, gender diverse, other, mixed, non-conforming and/or non-binary	12
	% F- % of participants that were female	143
Age	All data- all data related to age	150
	Mean or median- The mean and/or median age	52
	Average- Unspecified "average" age	22
	Minimum- The minimum age	35
	Maximum- The maximum age	34
	Range- The maximum- minimum age	34
Race/ Ethnicity	All data- all data related to race/ethnicity	30
	% White- % of participants that were white	17
Education	All data- all data related to education	88
	% College + - % of participants with some college education	58

	% Grad degree- % of participants with a graduate or professional degree	40
Occupation	Job- all data about participant jobs	40
	Income- all data about income	19
	Retiree- % of participants that were retired	65
Recreation	Frequent outdoor recreation- % of participants that frequently undergo outdoor recreation	1
Citizen science	Previous Citizen science- % of participants with previous Citizen science participation	12
Political views	All data- all data about political views	1
Residence	Living area- all data about what area participants lived in	14
	Dwelling type- all data about what type of dwelling the participants lived in, e.g. apartment	3
	Rural/Urban- all data about whether the participants lived in an urban or rural area	6
Family structure	Head of House- % of participants that were the head of their household	4
	Cohabitation- all data about number of people the participants lived with	6
	Marital Status- all data about marital status of the participants	3
	Children- all data about number of children the participants had	9
	Pets- all data about number of pets the participants had	3

Reported versus "assumed" retirees



Supplemental Figure 1: Relationship between reported % retirees ("Reported Retirees") and estimated % retirees calculated based on reported age data ("Assumed Retirees"), with a trend-line added and R^2 provided.

Descriptive statistics of the five chosen demographic variables

Supplemental Table 2: For each demographic proportion (female, people with graduate degree (grad), white, retiree) or mean average (Age), the unweighted (UW) and weighted (W) mean, unweighted standard deviation (SD), minimum (Min), maximum (Max), range (maximum – minimum), number of projects (Project N), number of participants (Participant N), and comparison census data (U.S. Pop.).

Demo. Var.	Mean (UW)	Mean (W)	SD	Min	Max	Range	Proj. N	Part. N	U.S. Pop.
Female	0.45	0.51	0.20	0.03	0.94	0.91	143	129,681	0.51[a]
Grad	0.35	0.43	0.17	0.00	0.67	0.67	40	38,522	0.13[b]
White	0.88	0.87	0.17	0.42	1.00	0.58	17	37,699	0.64[c]
Retiree	0.21	0.16	0.19	0.00	0.76	0.76	65	62,231	0.17[d]
Age	48.1	51.5	9.5	16.6	64.0	47.4	52	27,985	38.8[e]

a- U.S. adults in 2020. Source: U.S. Census Bureau 2020 (https://data.census.gov/)
b- For 2019, percentage of U.S. aged 25+ that have a graduate or professional degree. Source: U.S. Census Bureau (https://www.census.gov/library/stories/2019/02/number-of-people-with-masters-and-phd-degrees-double-since-2000.html)
c- U.S. adults in 2020. Source: U.S. Census Bureau (https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html)
d- U.S. adults in 2021. Source: Administration on Aging 2021 (https://acl.gov/sites/default/files/Profile%20of%20OA/2021%20Profile%20of%20OA/2021ProfileOlderAmericans_508.pdf)
e- U.S. population in 2021. Source: U.S. Census Bureau (https://www.census.gov/newsroom/press-releases/2022/population-estimates-characteristics.html)

Detailed overview of project foci

Supplemental Table 3: For each category of project focus, the % of projects that we determined fell into that category (%), the number of projects in the category (N), a description of the category (description), and an example.

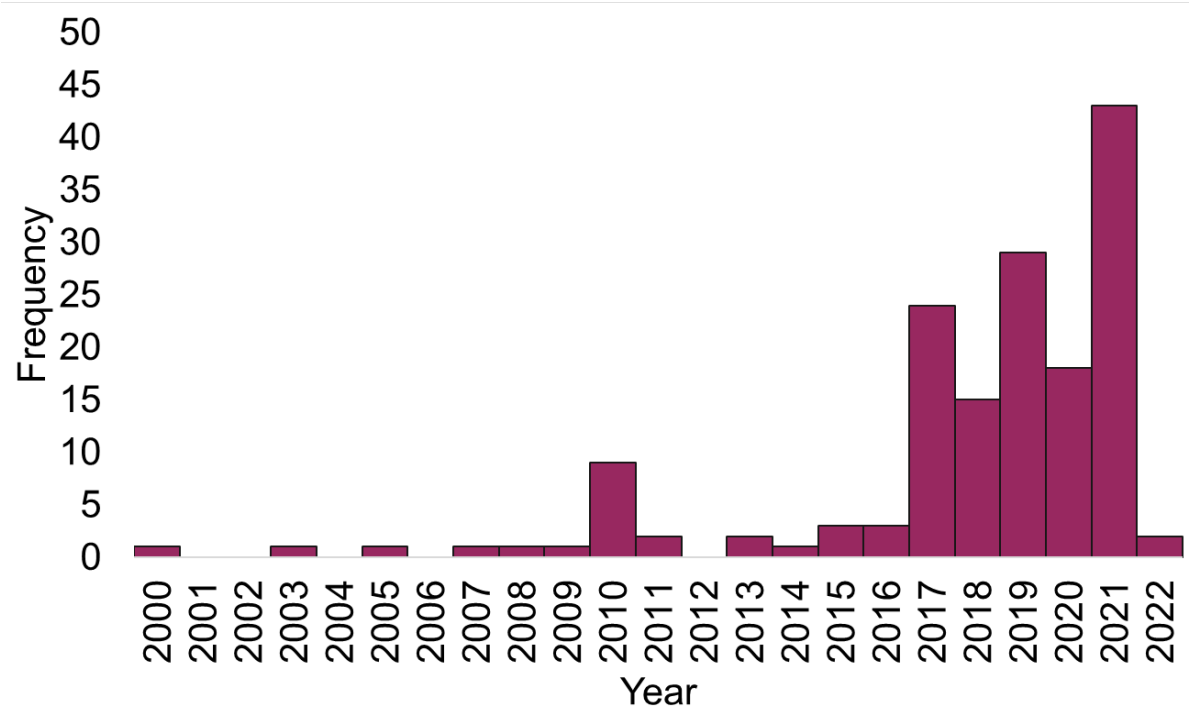
Category	%	N	Description	Example
Physical Science	8.3	13	the nonliving world but including molecular structure even if attached to biochemistry, as well as extra-terrestrial (e.g., astronomy)	amateur astronomy - Jones, M.G., Corin, E.N., Andre, T., Childers, G.M. and Stevens, V., 2017. Factors contributing to lifelong science learning: Amateur astronomers and birders. Journal of Research in Science Teaching, 54(3), pp.412-433.
Health	26.8	42	collection/monitoring of participant activity (e.g., steps, sleep) and/or health measures (e.g., heart rate); projects centering on data collection in service of determining whether human health is being affected (e.g., air pollution projects where the pollution is about people first)	Malaria Control Project - Asingizwe, D., Poortvliet, P.M., Koenraadt, C.J., van Vliet, A.J., Ingabire, C.M., Mutesa, L. and Leeuwis, C., 2020. Why (not) participate in citizen science? Motivational factors and barriers to participate in a citizen science program for malaria control in Rwanda. PloS one, 15(8), p.e0237396.
Biodiversity	56.7	89	the living natural world, including conservation projects	Candid Critters - Allf, B.C., Cooper, C.B., Larson, L.R., Dunn, R.R., Futch, S.E., Sharova, M. and Cavalier, D., 2022. Citizen science as an ecosystem of engagement: implications for learning and broadening participation. BioScience, 72(7), pp.651-663.
Other	6.4	10	a catch-all for projects focused on data and issues clearly outside of the three categories described above	Tomnod focused on digital imaging crowdsourcing- - Baruch, A., May, A. and Yu, D., 2016. The motivations, enablers and barriers for voluntary participation in an online

				crowdsourcing platform. Computers in Human Behavior, 64, pp.923-931.
Unknown	1.9	3	the project focus could not be determined from the publication, and website information was not available	Lake water quality - Bos, J.S., Nanayakkara, L., Hurlbert, M. and Finlay, K., 2019. Citizen science for Saskatchewan lakes: a pilot project. Lake and Reservoir Management, 35(1), pp.77-89.

Comparison between this study and SciStarter

Supplemental Table 4: Project attributes used in our study ("this study", N = 157), compared to SciStarter (N = 1,599). Data collected on 25 Sept 2022.			
This Study	**This Study Project %**	**SciStarter**	**SciStarter Project %**
biodiversity	56.7	Ecology & Environment	54.9
human health	26.8	Health & Medicine	27.6
physical science	8.3	Astronomy & Space, Chemistry, Physics, Geology & Earth Science	5.6
other	6.4	remaining projects, after subtracting all the projects counted in other categories from the total projects on SciStarter	11.9
online	13.2	Online only	11.9
Hands-on	86.8	not Online	88.1

Project counts per year



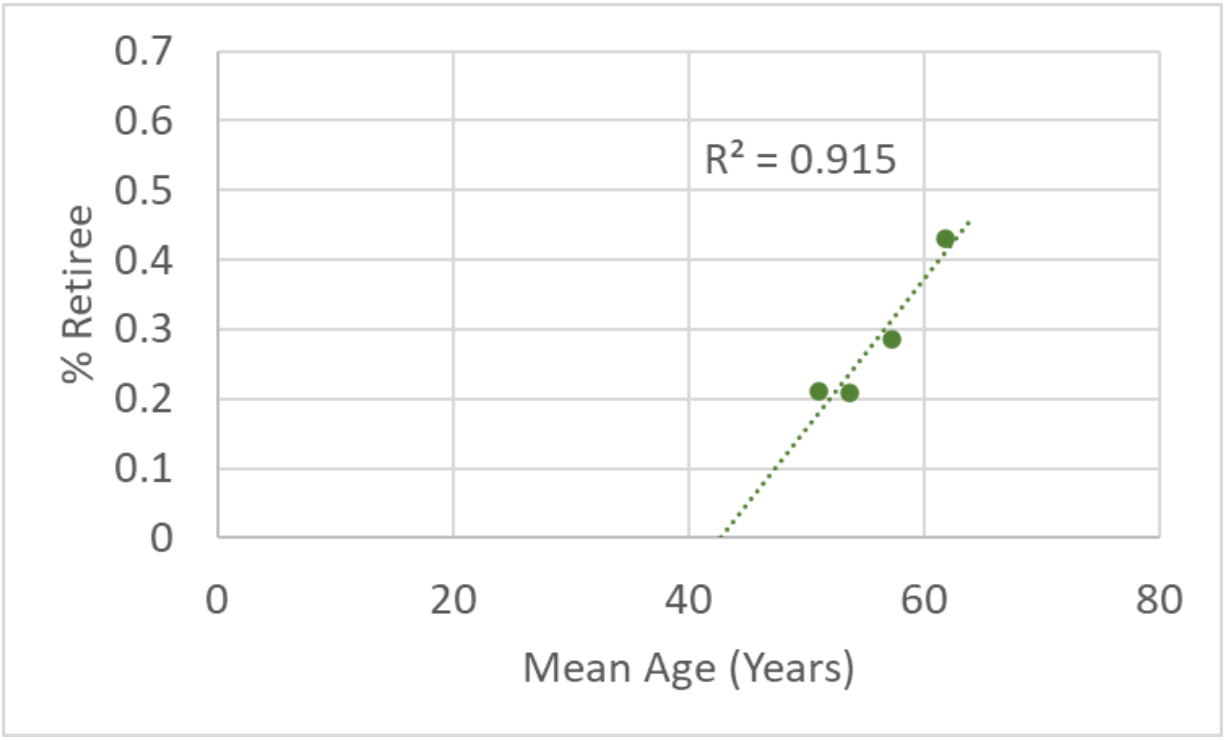
Supplemental Figure 2: Project count ("frequency") as a function of publication year.

Effect of adding 2011-2016 data

Supplemental Table 5: Results of quasi-binomial GLMMs of the proportion of individuals in CS projects who fell into different demographic categories (being female, having a graduate degree, being a retiree, and being white). Here, data from 2011-2016 was not included. For each coefficient, variable estimates, standard error (SE), the lower (2.5% CI) and upper (97.5% CI) confidence intervals, and P-values for each variable in the final models are shown. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Significance is assessed against hands-on for project access and biodiversity focus for focus.

Coefficient	Estimate	SE	2.5% CI	97.5% CI	P-value
Female (n = 106)					
Intercept	-511.33	109.24	-727.59	-298.79	8.73×10^{-6} ***
Year	0.25	0.05	0.15	0.36	8.85×10^{-6} ***
Project Access					4.21×10^{-5} ***
Online	0.63	0.14	0.35	0.91	1.13×10^{-5} ***
White (n = 13)					
Intercept	2.85	0.38	2.19	3.68	1.87×10^{-5} ***
Focus					0.01 **
Health	-1.14	0.38	-1.97	-0.46	0.01 *
Grad (n = 31)					
Intercept	-0.75	0.12	-0.99	-0.52	7.65×10^{-7} ***
Project Access					0.00 ***
Online	0.56	0.13	0.31	0.82	0.00 ***
Retiree (n = 49)					
Intercept	-0.73	0.11	-0.95	-0.51	4.23×10^{-8} ***
Focus					4.90×10^{-12} ***
Health	-1.50	0.14	-1.78	-1.22	9.86×10^{-14} ***
Physical Sciences	-0.19	3.05	-14.08	6.67	0.95

Relationship between age and retirees



Supplemental Figure 3: Relationship between mean age (years) and % retiree (n=4).

Effect of changing maximum weight for the linear models

Supplemental Table 6: Results of quasi-binomial GLMMs of the proportion of individuals in CS projects who fell into different demographic categories (being female, having a graduate degree, being a retiree, and being white). Here, the maximum weight a project can have is 75. For each coefficient, variable estimates, standard error (SE), the lower (2.5% CI) and upper (97.5% CI) confidence intervals, and P-values for each variable in the final models are shown. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Significance is assessed against hands-on for project access and biodiversity focus for focus.

Coefficient	Estimate	SE	2.5% CI	97.5% CI	P-value
Female (n = 115)					
Intercept	-511.73	88.84	-687.84	-339.22	7.49×10^{-8} ***
Year	0.25	0.04	0.17	0.34	7.66×10^{-8} ***
Project Access					1.98×10^{-5} ***
Online	0.62	0.14	0.36	0.89	1.13×10^{-5} ***
White (n = 14)					
Intercept	2.94	0.40	2.24	3.83	1.36×10^{-5} ***
Focus					0.00 **
Health	-1.23	0.40	-2.12	-0.52	0.01 *
Grad (n = 33)					
Intercept	-0.75	0.12	-0.98	-0.52	4.15×10^{-7} ***
Project Access					0.00 ***
Online	0.55	0.13	0.31	0.80	0.00 ***
Retiree (n = 53)					
Intercept	-0.75	0.11	-0.97	-0.54	8.61×10^{-9} ***
Focus					2.64×10^{-12} ***
Health	-1.48	0.14	-1.76	-1.20	3.96×10^{-14} ***
Physical Sciences	-0.96	0.48	-2.00	-0.10	0.05 *

Supplemental Table 7: Results of quasi-binomial GLMMs of the proportion of individuals in CS projects who fell into different demographic categories (being female, having a graduate degree, being a retiree, and being white). Here, the maximum weight a project can have is 125. For each coefficient, variable estimates, standard error (SE), the lower (2.5% CI) and upper (97.5% CI) confidence intervals, and P-values for each variable in the final models are shown. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Significance is assessed against hands-on for project access and biodiversity focus for focus.

Coefficient	Estimate	SE	2.5% CI	97.5% CI	P-value
Female (n = 115)					
Intercept	-522.55	88.67	-698.36	-350.42	4.06×10^{-8} ***
Year	0.26	0.04	0.17	0.35	4.15×10^{-8} ***
Project Access					1.84×10^{-5} ***
Online	0.62	0.13	0.36	0.88	1.04×10^{-5} ***
White (n = 14)					
Intercept	2.97	0.37	2.32	3.77	5.78×10^{-6} ***
Focus					0.00 **
Health	-1.25	0.37	-2.06	-0.59	0.01 **
Grad (n = 33)					
Intercept	-0.76	0.11	-0.98	-0.53	2.26×10^{-7} ***
Project Access					9.9×10^{-5} ***
Online	0.56	0.12	0.32	0.80	7.15×10^{-5} ***
Retiree (n = 53)					
Intercept	-0.74	0.11	-0.95	-0.54	7.32×10^{-9} ***
Focus					1.05×10^{-12} ***
Health	-1.49	0.14	-1.76	-1.21	1.53×10^{-14} ***
Physical Sciences	-0.98	0.47	-2.00	-0.14	0.04 *