



**Environmental
Science**
Processes & Impacts

**Participatory science for coastal water quality: freshwater
plume mapping and volunteer retention in a randomized
controlled experiment**

Journal:	<i>Environmental Science: Processes & Impacts</i>
Manuscript ID	EM-ART-12-2019-000571.R1
Article Type:	Paper

SCHOLARONE™
Manuscripts

1
2
3 **Participatory science for coastal water quality: freshwater plume mapping and volunteer**
4 **retention in a randomized informational intervention**
5
6
7

8 Wiley C. Jennings,¹ Sydney Cunniff,¹ Kate Lewis,¹ Hailey Deres,¹ Dan R. Reineman,^{1,†} Jennifer
9 Davis,^{1,2} Alexandria B. Boehm^{1,*}
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 ¹ Department of Civil and Environmental Engineering, Stanford University, Stanford, California
46 94305 USA

47 ² Woods Institute for the Environment, Stanford University, Stanford, California 94305, USA

48 * Corresponding author: 473 Via Ortega, Room 189, Dept. of Civil & Environmental Engineering,
49 Stanford University, Stanford, CA 94305. aboehm@stanford.edu, (650) 724-9128 (tel)

50 † Current address: Environmental Science & Resource Management Program, California State
51 University Channel Islands, Camarillo, California 93012 USA

52 Supporting Information (SI) document includes supporting text and figures. Additional
53 documents referenced in text are available as separate files, described in SI document.
54
55
56
57
58
59
60

Abstract:

Among the biggest threats to coastal water quality are freshwater discharges. It is difficult to predict the spatial extent of freshwater plumes at marine beaches because processes governing mass transport in the surf zone are complex. Participatory science approaches could facilitate collecting shoreline data, although volunteer sampling campaigns can be challenged by data quality and volunteer retention. The goals of this study were to (1) work with volunteers to estimate safe swimming distances at beaches that receive polluted discharges, and (2) test whether informational feedback to volunteers increased retention. Forty-six volunteers participated over 12 weeks in 2019 by collecting 1,452 salinity measurements at beaches near the mouths of two central California freshwater discharges and completing participation surveys. These measurements resulted in 145 distinct estimates of safe swimming distances (D_{90}), spanning a range of environmental conditions during rainy and dry periods. Median D_{90} s were 150 and 100 m at San Pedro Creek south and north, and 490 and 330 m at San Lorenzo River west and east, respectively. D_{90} was significantly associated with adjacent freshwater discharge rate at both discharges and tide level at one discharge. On average, the odds of volunteers conducting sampling decreased by 4% (95% CI: 1%, 7%) with each successive week. A randomized intervention providing repeated data feedback via email to volunteers did not affect their retention in the study.

Key words: participatory science, citizen science, water quality, beach pollution, volunteer participation, coastal discharges, surf zone, pollutant transport, community monitoring

1
2
3
4
5
6 **Environmental significance statement.** Runoff from land can carry high concentrations of
7
8 contaminants, including pathogens, nutrients, and trace organic chemicals to coastal waters.
9
10 This research investigated the factors that control the shoreline extent of runoff plumes and
11
12 found that discharge rate of the runoff source, as well as tide level at one beach, were
13
14 associated with plume extents. A unique aspect of this study is that sampling was performed by
15
16 local volunteers. Utilizing a randomized control experiment, we found that volunteers who
17
18 received project results during the study were not more likely to stay active in the study than
19
20 other volunteers. This work contributes to understanding factors that affect coastal water
21
22 quality, as well as best practices for engaging volunteers in environmental science projects.
23
24
25
26
27
28
29

30 **Introduction**

31
32 Forty-four percent of the world's population resides within 150 km of a coastline, and eight
33
34 of the world's ten largest cities are coastal.¹ Shoreline and coastal waters are valuable natural
35
36 resources that provide numerous cultural, economic, and ecological benefits.^{2,3} Despite the
37
38 economic and cultural importance of clean coastal waters, human impacts on water quality
39
40 along the world's coastlines threaten ecosystem and human health.^{4,5}
41
42
43
44

45 Among the biggest threats to coastal water quality are small-scale (flow rate on the order of
46
47 10 m³/s or less) freshwater discharges from rivers, streams, tidal outlets, and storm drains.⁶
48
49 These are common along US coastlines⁷ and often contain urban and agricultural runoff,⁸ feces
50
51 from wildlife,^{9,10} and wastewater treatment plant effluent, as well as septage¹¹ or raw sewage
52
53 when treatment systems perform poorly or are overwhelmed.^{12,13} Incoming waves affect the
54
55
56
57
58
59
60

1
2
3 mixing and transport of these freshwater discharges by opposing the freshwater jet's cross-
4 shore momentum, rapidly mixing the water column, and driving an alongshore current in the
5
6 shore momentum, rapidly mixing the water column, and driving an alongshore current in the
7
8 direction of the waves' shoreline approach.^{14,15} At wave-dominated beaches, much of the
9
10 freshwater plume from these small-scale discharges may be entrained in the surf zone, the
11
12 region between the shoreline and the wave breaker line.¹⁶ This freshwater can remain trapped
13
14 near the shoreline, where it spreads in the alongshore direction along with any pollutants it
15
16 contains.¹⁵⁻¹⁷ A surf zone tracer study conducted at two tidal outlets in southern California
17
18 found that the alongshore flux of surf zone water was 50-300 times greater than the cross-
19
20 shore flux.¹⁷ This is important for beach water quality because it means that freshwater
21
22 pollutants may spread considerable distances along the shoreline, increasing beachgoers'
23
24 pollutant exposure.

25
26
27
28
29
30 It is difficult to predict the spatial extent of freshwater plumes along the shoreline using
31
32 analytical models because the processes governing mass transport in the surf zone are
33
34 numerous and complex.^{14,17-22} One empirical method to estimate the length scale of freshwater
35
36 plumes in the surf zone is to use salinity as a naturally occurring tracer. This approach is
37
38 advantageous because salinity measurements are analytically simple to make. However, to
39
40 capture the variability in the freshwater plume extent at a beach, many measurements must be
41
42 made across space (to estimate spatial variation in freshwater content) and time (to span the
43
44 range of physical conditions present at a beach). Additionally, measurement must be
45
46 undertaken by hand, since sensor deployment in the surf zone may be impeded by waves and
47
48 tides, permitting requirements, and beachgoers, particularly if deployment across multiple days
49
50 is required.
51
52
53
54
55
56
57
58
59
60

1
2
3 Working with local community members to gather frequent samples could help meet these
4
5 sampling requirements. In recent years, participatory science (also “citizen science”) has grown
6
7 increasingly popular among academics and practitioners for gathering environmental data.^{23,24}
8
9
10 Volunteer-based participatory science offers the potential benefits of gathering many
11
12 observations in a cost-effective way,^{23,25,26} enhancing public scientific knowledge and skills,²⁷
13
14 and elevating the perceived value of the issue or environment under study.²⁸ Two primary
15
16 challenges for participatory science are data quality and volunteer retention. To mitigate
17
18 potential data quality problems, methods and protocols must be appropriate for volunteer skill
19
20 levels and the logistical challenges of numerous participants independently contributing data.²⁹
21
22
23 Quality control (QC) measures specific to the protocols used by volunteers must also be
24
25
26 developed to ensure that data are of adequate quality for their intended purpose.
27
28
29

30 Data quality concerns are common to all scientific studies. In contrast, volunteer retention
31
32 is a unique and important challenge for participatory science.³⁰⁻³² Retaining volunteers in
33
34 longitudinal studies is particularly difficult, and poor retention can threaten a study’s success.³⁰
35
36
37 Therefore, understanding how specific aspects of project design or management affect
38
39 sustained volunteer participation could improve outcomes for participatory science studies.
40
41
42 Indeed, a growing number of studies has sought to understand drivers of volunteer
43
44 participation and retention.³³⁻⁴¹ The use of cross-sectional or observational designs, however,
45
46
47 has precluded any participatory environmental science study from inferring causality between
48
49 project or volunteer characteristics and volunteer retention. The present study seeks to test the
50
51 effect of providing motivationally-targeted feedback to volunteers in a scientific study using a
52
53
54 managed experimental design.
55
56
57
58
59
60

1
2
3 The basis for this experimental treatment, which aims to enhance learning-oriented
4 benefits of volunteering by regularly presenting volunteers with data collected during the
5 study, is a framework for understanding volunteer retention from the field of functional
6 psychology. This framework posits that volunteer participation is more likely sustained when
7 specific motivations for volunteering align with perceived benefits of participating. Motivations
8 alone, divorced from corresponding perceived benefits, are insufficient to sustain volunteer
9 participation.^{42,43} Randomized experiments assessing the impact of motivationally-targeted
10 communication materials on volunteer participation have been conducted in the human
11 services sector⁴⁴ and in a massive online participatory physics study.⁴⁵ The present study is the
12 first known participatory environmental science study to test the effect of informational
13 feedback on sustaining participation with an experimental design.

14
15 This study sought to address knowledge gaps in beach water quality (WQ) and participatory
16 science (PS) research by investigating the following four questions: WQ1: How far along the
17 shoreline are two polluted freshwater discharges at popular beaches likely to have meaningful
18 public health impacts? WQ2: How does variation in environmental conditions at these sites
19 affect these alongshore extents? PS1: How much is volunteer retention increased by regularly
20 presenting to volunteers the data that they collect? PS2: To what extent is this effect
21 moderated by motivations for volunteering? This study provides data to inform safe swimming
22 distances from discharges at recreational beaches, a systematic framework for assessing
23 volunteer data quality that may be adapted to other studies, and insight into best practices for
24 sustaining volunteer engagement in participatory environmental science.

Methods

Study sites. Study sites (Figure 1) were located at the receiving beaches for two freshwater discharges in California, USA. San Pedro Creek in Pacifica (37.596560°, -122.505785°), which discharges to Pacifica State Beach, and San Lorenzo River, which discharges to Main and Seabright Beaches in Santa Cruz (36.963278°, -122.012892°), were selected because they frequently contain unsafe levels of fecal indicator bacteria^{46–49} and pathogens,^{8,50} discharge to popular recreational beaches, and differ in scale.⁸ Both discharges are located in watersheds that experience distinct wet and dry seasons, with most precipitation typically falling from November to March.⁵¹ The shoreline on each side of each discharge was considered a sampling site. These four sites are referred to as San Pedro Creek south (meaning the shoreline to the south of San Pedro Creek's mouth), San Pedro Creek north, San Lorenzo River west, and San Lorenzo River east (Figure 1).

Volunteer recruitment and communication. Volunteers were recruited online and in-person from nearby communities (see Supporting Information (SI)) and were required to attend one of several in-person trainings. Following training, all communication with volunteers was via email, until after the project's end date when all volunteers were invited to results presentations. Volunteers were required to be at least 18 years of age and were not compensated for their participation in this study. This study was approved by Stanford University's Institutional Review Board (protocol IRB-37740).

1
2
3 **Field sampling.** Volunteers were assigned to sample at one sampling site and conducted field
4
5 sampling between February and May 2019. All volunteers received a kit of sampling supplies
6
7 with a detailed sampling protocol (see SI), as well as one 90-minute in-person training held in a
8
9 classroom setting. The training included background information on beach water quality, an
10
11 overview of project goals, and detailed step-by-step instructions for performing field sampling
12
13 and project logistics. Volunteers were asked to conduct sampling once per week for twelve
14
15 weeks. During each sampling event, volunteers were asked to collect samples in ankle deep
16
17 water, on an incoming wave, at a minimum of five points along the shoreline. Volunteers
18
19 recorded the salinity of each sample using a low-cost analog refractometer (Agriculture
20
21 Solutions, Dual Scale Refractometer). The measurements made during a sampling event are
22
23 collectively referred to as a salinity profile. Other relevant meta-data were also recorded (see
24
25 SI). Volunteers were asked to select locations that spanned a full range of salinity values, from
26
27 the salinity at the mouth of the discharge (as low as 0 PSU) to background marine salinity of 34
28
29 PSU. Volunteers then recorded the GPS coordinates of each location using a smartphone app
30
31 (Save Location GPS, Rayo Infotech).
32
33
34
35
36
37
38
39

40 Data were recorded, and GPS coordinates were transcribed on paper data sheets. Following
41
42 each sampling event, volunteers deposited a data sheet along with one of the water samples
43
44 for which they recorded salinity at an adjacent drop-off point hosted by a local establishment.
45
46 Project staff collected samples each week and returned them to the lab for validation of salinity
47
48 measurement accuracy and data digitization. To prevent data digitization errors, all data sheets
49
50 were digitized twice by different project staff, and mismatches were identified and corrected.
51
52
53
54
55
56
57
58
59
60

1
2
3 **Refractometer QC.** An experiment to determine a threshold for acceptable refractometer
4 accuracy among volunteers, incorporating both variability due to field conditions and
5 differences across refractometers, was conducted as follows. One trained project staff member
6 made five measurements with each of ten refractometers in the field, mimicking the salinity
7 profiles requested of volunteers. Accuracy of these fifty measurements was determined by
8 comparing them to measurements made in the lab using a digital Sonde probe (YSI-30), and the
9 acceptable accuracy threshold was set at the level of error containing 95 percent of
10 measurements, equal to +/- 2.3 PSU. Further investigation of refractometer accuracy and
11 repeatability is described in the SI.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 **Volunteer data QC criteria.** To ensure that data collected by volunteers were sufficiently
28 accurate to produce scientifically defensible estimates of 90 percent dilution distances (D_{90S}),
29 we implemented a strict QC procedure. The QC framework was divided into three conceptual
30 categories: (1) completeness, the requirement that data collection tasks were completed
31 sufficiently for data to be interpretable; (2) reasonableness, the requirement that data were
32 plausible; and (3) usefulness, the requirement that the data collected could be used to answer
33 the scientific questions of this study. Each category was operationalized by multiple criteria
34 (Table 1). For example, measurement accuracy is one criterion under the reasonableness
35 category. Salinity accuracy was determined by using a digital salinity probe (YSI, Model 30,
36 Yellow Springs, Ohio) to measure the salinity of volunteers' submitted samples, representing
37 about 20 percent of all recorded measurements. If a volunteer's measurement of a submitted
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 sample had error exceeding 2.3 PSU, all measurements recorded by that volunteer on that day
4
5 were discarded.
6
7
8
9

10 **D₉₀ estimation.** After QC, refractometer instrument bias was corrected to increase salinity
11 measurement accuracy (see SI). Salinity was transformed to freshness, defined as $f = (sal_{bg} -$
12 $sal_{meas}) / sal_{bg}$, where f is freshness, sal_{bg} is background coastal salinity (34.5 PSU was used,
13 based on measurements made by staff at field sites during low discharge flow conditions), and
14 sal_{meas} is the measured salinity. Salinity values below 0 were assigned 0 PSU and above 34.5
15 assigned 34.5 PSU to ensure that freshness was bounded by [0, 1]. Note that measured values
16 above 37 or below -3 PSU were discarded (Table 1).
17
18
19
20
21
22
23
24
25
26

27 To compute shoreline distances from GPS-labeled sampling locations, several intermediate
28 approximations were made (see SI). Then, nonlinear regression was used to estimate how
29 freshness decayed with alongshore distance from each discharge datum. A two-parameter
30 exponential decay model was fit to each freshness profile:
31
32
33
34
35
36

$$f_{ij} = k_{1j} \exp(-k_{2j} d_{ij}) + \varepsilon_{ij} \quad \varepsilon_j \sim N(0, \sigma_j^2).$$

37 where $f \equiv$ freshness (dimensionless) of the i^{th} point in the j^{th} freshness profile; $d \equiv$ distance from
38 discharge (m); $k_1 \equiv$ best fit freshness at discharge mouth (dimensionless); $k_2 \equiv$ spatial decay rate
39 of freshness (m^{-1}); $\varepsilon \equiv$ regression residual (dimensionless). Non-linear regression of freshness
40 values was used rather than linear regression on log-transformed data because freshness
41 residuals were normal and additive on the untransformed scale.⁵²
42
43
44
45
46
47
48
49
50
51

52 Finally, D_{90} , the alongshore distance from the creek or river at which freshwater was diluted
53 by 90 percent, was estimated for each freshness profile by setting each f_j to 0.1 and solving for
54
55
56
57
58
59
60

1
2
3 *d_j*. Ninety percent was used because fecal indicator bacteria (FIB) levels in San Pedro Creek and
4
5 San Lorenzo River from 2016 to 2018 were frequently greater than the CA single sample
6
7 criterion used to protect public health, but rarely exceeded ten times the criterion, indicating
8
9 that 90 percent dilution is typically sufficient to reduce FIB to levels considered safe in
10
11 individual samples (Figure S2).
12
13
14
15
16
17

18 **Associations between D_{90} and hydrodynamic variables.** Linear association between D_{90} and
19
20 tide level, flow rate, and alongshore current were investigated. Verified six-minute tide level
21
22 data were obtained from NOAA (San Francisco Station, 9414290) with mean lower low water
23
24 datum.⁵³ Fifteen-minute flow rates of San Lorenzo River were retrieved from a USGS gage near
25
26 the river mouth.⁵⁴ Hourly flow rates of San Pedro Creek were predicted from weekly flow
27
28 measurements (handheld flow meter, FP101, Global Water) and hourly precipitation data using
29
30 random forest regression (see SI). Alongshore velocity was estimated using an equation
31
32 proposed by Inman et al.,¹⁴ which is a function of surf zone wave height and angle (output from
33
34 the Coastal Data Information Program model⁵⁵), beach slope (for San Lorenzo River, calculated
35
36 in QGIS using a USGS digital elevation model;⁵⁶ for San Pedro Creek, taken from Wong et al.¹⁶),
37
38 and bottom drag (taken from Inman et al.¹⁴) (see SI). Velocity estimates were z-standardized
39
40 and interpreted as relative values.
41
42
43
44
45
46

47 To evaluate associations between D_{90} and hydrodynamic variables, linear ordinary least
48
49 squares regression was used. One model was estimated for each of the two discharges (San
50
51 Pedro Creek and San Lorenzo River). Models were specified as:
52
53

$$\sqrt{D_{90}} = \beta_0 + \beta_1 \times tide + \beta_2 \times \log_{10}(discharge) + \beta_3 \times Z_{current} + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$

1
2
3 where *tide* is tide level (relative to mean lower low water) at the time of sampling, *discharge* is
4
5 volumetric flow rate at the time of sampling, and $Z_{current}$ is the z-standardized alongshore
6
7 current velocity at the time of sampling. D_{90} was square-root transformed to reduce
8
9 heteroscedasticity in the residuals. Variance inflation factors were used to check for
10
11 multicollinearity. Wilcoxon rank sum tests were used to test several hypotheses regarding
12
13 whether excluding the D_{90} s that failed certain QC criteria biased D_{90} statistics. Specifically, D_{90}
14
15 estimates that were excluded due to insufficient spatial coverage of samples (Figure S7) were
16
17 compared to other D_{90} s with respect to magnitude and correlations with hydrodynamic
18
19 variables. Throughout this paper, an alpha level of 0.05 was used to determine statistical
20
21 significance.
22
23
24
25
26
27
28
29

30 **Randomized informational feedback.** All participants received twice-monthly emails with
31
32 recommendations for resolving common data collection issues observed by project staff or
33
34 reported by volunteers. To test whether increasing scientific feedback to volunteers affected
35
36 sampling effort, emails sent to one half of volunteers also prominently featured a link to access
37
38 a report of data collected over the previous two weeks (“data reports”). Treatment condition
39
40 (receive or not receive links to data reports) was randomly assigned after stratifying the
41
42 volunteers by gender and sampling location. Data reports included plots and descriptions of
43
44 aggregated salinity data collected by volunteers, along with discharge flow data and FIB
45
46 measurements (see SI). Emails were sent with Mail Chimp; click data were retrieved to
47
48 determine the number of data reports viewed by each volunteer.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **Volunteer survey overview.** Three online surveys were administered to each volunteer over
4
5 the course of the project (see SI). Prior to commencing sampling, volunteers were administered
6
7 a pre-project survey, which included one battery of items based on the Volunteer Functions
8
9 Inventory⁴³ to gauge volunteer motivations for participating, and a battery of items to gauge
10
11 expected obstacles to participating (see *Data analysis: motivations and obstacles*). A mid-
12
13 project survey was administered during weeks 6-7 of sampling, and a post-project survey was
14
15 administered immediately following conclusion of field sampling. Both mid and post surveys
16
17 asked participants about sampling and participation problems they had experienced, and
18
19 repeated questions from the Volunteer Functions Inventory to gauge perceived benefits of
20
21 participating up to that point. All surveys were conducted via the Qualtrics online platform (see
22
23 SI). Links to surveys were emailed to participants, and up to four follow-up emails were sent to
24
25 volunteers to encourage survey completion.
26
27
28
29
30
31
32
33
34

35 **Survey constructs and analysis: motivations, benefits, and obstacles.** We evaluated six
36
37 motivations for and corresponding benefits of volunteering in this project: learning-, values-,
38
39 social-, career-, recreation-, and child-oriented. The first four were proposed by Clary et al.,⁴³
40
41 and we added the last two after administering an open-ended survey during a pilot version of
42
43 this project in 2017. Motivational constructs were defined in terms of volunteers' perception of
44
45 opportunities offered by volunteering with the project, as follows. Learning-oriented
46
47 motivation is the perception that volunteering with this project offers opportunities to learn or
48
49 exercise knowledge or skills. Values-oriented motivation is the perception that volunteering
50
51 with this project offers opportunities to contribute to an issue the volunteer is concerned
52
53
54
55
56
57
58
59
60

1
2
3 about. Social-oriented motivation is the perception that volunteering with this project offers
4
5 opportunities to develop social relationships. Career-oriented motivation is the perception that
6
7 volunteering with this project offers opportunities to develop one's career. Recreation-oriented
8
9 motivation is the perception that volunteering with this project offers opportunities to engage
10
11 in an enjoyable activity. Child-oriented motivation is the perception that volunteering with this
12
13 project offers opportunities to participate in an activity with one's children.
14
15
16

17
18 Corresponding benefit constructs were measured in the mid- and post-project surveys, and
19
20 differed from motivation constructs in that they addressed retrospective, rather than
21
22 prospective, project experience. Benefit constructs were defined similarly to motivation
23
24 constructs. For example, learning-oriented benefits are the perception that volunteering with
25
26 this project offered opportunities to learn or exercise knowledge or skills.
27
28
29

30
31 Construct indicators, each corresponding to one survey item, permitted integer responses
32
33 from 1 to 5, corresponding to 'strongly disagree' to 'strongly agree' with 3 as 'neither agree nor
34
35 disagree'. Three indicators operationalized each of the learning-, values-, and social-oriented
36
37 motivational constructs, while one indicator operationalized each of the career, recreational,
38
39 and child-oriented motivations (Table S1 lists individual indicators). This design was used to
40
41 constrain the length of the survey while including other crucial survey items. Learning- and
42
43 values-oriented motivational constructs were prioritized because they were expected to be
44
45 common among volunteers, based on pilot surveys conducted in 2017, and because volunteers
46
47 of either orientation were theorized to plausibly respond positively to the informational
48
49 intervention. The social-oriented construct was prioritized to serve as a control: we theorized
50
51 that volunteers with social-oriented motivations would be unaffected by the intervention.
52
53
54
55
56
57
58
59
60

1
2
3 Confirmatory factor analysis could not be performed on the motivational constructs due to
4
5 statistical concerns (see SI). To determine a volunteer's primary motivation, a score was
6
7 calculated for each motivation as either the average response value (for constructs with
8
9 multiple indicators) or the response value (for constructs with a single indicator). A volunteer
10
11 was classified as learning oriented if their learning score was greater than 3 and at least as high
12
13 as all other motivation scores.
14
15
16

17
18 Obstacles to participating were surveyed and defined as the logistical, financial, or
19
20 physiological hindrances that make participation more effortful. These hindrances included low
21
22 time availability (indicated by self-reported employment status), high travel time to project site
23
24 (volunteers' self-reported estimates), having to pay for parking at field sites (self-reported), and
25
26 physical difficulty of performing sampling tasks (self-reported). Surveyed obstacles and scoring
27
28 details are provided in Table S2 and Figure S4.
29
30
31
32
33
34

35 **Analysis of volunteer retention.** The null hypothesis corresponding to research question PS1
36
37 was that receiving data reports was not associated with retention (referred to as the
38
39 "information hypothesis"). Since receiving data reports was a randomly assigned condition, this
40
41 association was interpreted as being causal. The null hypothesis corresponding to PS2 was that
42
43 retention of volunteers who received data reports and whose motivations were learning-
44
45 oriented was equal to retention of other volunteers (referred to as the "motivation
46
47 hypothesis"). Since motivation type was not a randomly assigned condition, this association
48
49 was not interpreted as being causal.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Generalized Estimating Equations (GEE) models with logit links were fit to test each of these
4
5 two hypotheses. In both models, the binary outcome was whether a volunteer conducted
6
7 sampling on a given day. Each of the 45 volunteers that completed the pre-project survey and
8
9 attended training was included over the 85-day duration of their sampling campaign.⁴⁴ An
10
11 exchangeable correlation structure with robust standard errors was used to account for
12
13 repeated observations of each volunteer. Predictors for the information hypothesis model were
14
15 obstacles score, treatment group (coded 0/1 for control/treatment), time (in days, coded 0 for
16
17 training day to 84), and a treatment group by time interaction. Predictors for the motivation
18
19 hypothesis model were obstacles score, target group (to be distinguished from a true treatment
20
21 group, coded 0/1 for other volunteers/volunteers who received data reports and were learning
22
23 motivated), time (in days), and a target group by time interaction. The coefficient estimate of
24
25 the treatment (or target) group by time interaction term in each model tests the corresponding
26
27 hypothesis regarding retention. Exponentiating this coefficient yields an odds ratio, describing
28
29 how the odds of conducting sampling for volunteers in the treatment (target) group change
30
31 over time compared to the control group, *i.e.*, volunteer retention. Obstacles score was
32
33 included to control for factors expected to impact retention that were not accounted for in the
34
35 randomized treatment assignment, including age- and access-related factors. Participant age
36
37 was correlated with obstacles score (Pearson's $r = -0.57$; 95% CI = -0.74, -0.34; Figure S5), so age
38
39 was excluded from regression models.
40
41
42
43
44
45
46
47
48
49
50
51

52 Results

53
54
55
56
57
58
59
60

1
2
3 **Salinity measurements.** 1,452 salinity measurements were submitted by 36 participants
4
5 between Feb 2 and May 4, 2019. Sampling occurred on 55 and 73 out of 92 possible days at San
6
7 Pedro Creek and San Lorenzo River, respectively. Forty-five percent of measurements did not
8
9 meet the QC criteria for the analysis of D_{90} (Figure S6). The criterion most commonly failed was
10
11 accuracy (17 percent of all measurements), followed by failing to sample a sufficiently long
12
13 beach extent to estimate D_{90} (“sufficient spatial coverage”, 11 percent) and failing to perform
14
15 calibration (7 percent). Grouping discarded data by QC category revealed that 10 percent of
16
17 measurements were discarded due to insufficient completeness, 21 percent due to insufficient
18
19 reasonableness, and 14 percent due to insufficient usefulness with respect to the specific
20
21 analysis aims of this study. After QC, 804 observations across the four study beaches,
22
23 constituting 145 salinity profiles, were used to estimate D_{90} values. D_{90} s excluded due to
24
25 insufficient spatial coverage tended to be longer than included D_{90} s ($p < 0.05$, Figure S7) at San
26
27 Pedro south, San Lorenzo west, and San Lorenzo east. At San Pedro south, excluded D_{90} s were
28
29 associated with higher discharge rates. At San Lorenzo west, excluded D_{90} s were associated
30
31 with higher discharge rates and higher tide levels. San Lorenzo east showed no associations
32
33 between data exclusion due to spatial coverage and hydrodynamic variables. At San Pedro
34
35 north, no D_{90} s failed the sufficient spatial coverage criterion.

36
37
38
39
40
41
42
43
44
45
46
47 **Distance to 90 percent dilution.** Exponential decay models fit freshness profiles well, judged by
48
49 residual standard deviations for each profile (example fits in Figure S8). Median (minimum,
50
51 maximum) residual standard deviation across profiles was 0.07 (0.00, 0.36) in units of
52
53 freshness. Thirty-five, 39, 27, and 44 D_{90} s were estimated and passed QC criteria at San Pedro
54
55
56
57
58
59
60

1
2
3 Creek south, San Pedro Creek north, San Lorenzo River west, and San Lorenzo River east,
4
5 respectively (Figure 2). The median (25th, 75th percentiles) D_{90} s were 150 (100, 180) m at San
6
7 Pedro Creek south, 100 (70, 180) m at San Pedro Creek north, 490 (320, 720) m at San Lorenzo
8
9 River west, and 330 (210, 520) m at San Lorenzo River east.
10
11

12
13 During sampling events that passed QC, discharge ranged from 5.9×10^{-2} to 1.4×10^0 m³/s
14
15 with a mean of 3.8×10^{-1} m³/s in San Pedro Creek, and from 2.7×10^0 to 1.6×10^2 m³/s with a
16
17 mean of 1.7×10^1 m³/s in San Lorenzo River. Alongshore current velocities were northward for
18
19 all 74 D_{90} s at San Pedro Creek and eastward for all 71 profiles at San Lorenzo River. For linear
20
21 regression modeling, alongshore current velocities were coded such that the direction of the
22
23 corresponding D_{90} with respect to the discharge mouth is positive. The mean (range) velocity
24
25 magnitude was 1.8 (0.6, 3.0) m/s in Pacifica, and 1.9 (0.7, 3.2) m/s in Santa Cruz. These values
26
27 exceeded ranges previously reported for similar conditions by up to an order of magnitude.^{16,17}
28
29 Thus, velocities were interpreted as relative values.
30
31
32
33
34

35 Linear regression models of D_{90} incorporating discharge rate, alongshore current, and tide
36
37 level were estimated for each of the two discharges. The model for San Pedro Creek explained
38
39 about one-third of the variation ($R^2 = 0.38$) in D_{90} (Table S3). The model for San Lorenzo River
40
41 explained about one-tenth of the variation in D_{90} (adj. $R^2 = 0.11$). At San Pedro Creek, a 10-fold
42
43 increase in discharge (m³/s) was significantly associated with a 16 m (95% CI: 6, 30 m) increase
44
45 in D_{90} . A 1 m increase in tide level was significantly associated with a 4 m (1, 10 m) decrease in
46
47 D_{90} . Alongshore current velocity was not statistically associated with D_{90} .
48
49
50
51

52 At San Lorenzo River, a 10-fold increase in discharge was significantly associated with a 19
53
54 m (0, 17 m) increase in D_{90} . Tide level was not statistically associated with D_{90} . A 1-standard
55
56
57
58
59
60

1
2
3 deviation increase in alongshore current velocity was statistically associated with a 2 m (0, 9 m)
4
5 decrease in D_{90} .
6
7

8 Due to the perfect collinearity between alongshore current velocity direction and discharge
9
10 side (north/south at San Pedro Creek, east/west at San Lorenzo River), dummy variables for
11
12 discharge side were not included in the model. This correlation made it impossible to use
13
14 regression to distinguish the effects of alongshore current direction on D_{90} from effects of other
15
16 features specific to each side of a discharge, such as discharge mouth or coastal morphology.
17
18 Variance inflation factors for reported model variables were approximately 1, indicating that
19
20 multi-collinearity was not a problem in these models.
21
22
23
24
25
26
27

28 **Volunteer participation and data report viewing rates.** Of over 60 people who expressed
29
30 interest in volunteering by completing a preliminary survey, 46 people attended a training and
31
32 were given sampling kits. Of those 46, 45 completed the pre-project survey, 31 completed the
33
34 mid-project survey, and 29 completed the post-project survey. Thirty-six volunteers completed
35
36 sampling at least once. Participation trends are shown Figure 3 and Figure S9, and participation
37
38 rates and the number of data reports viewed by each volunteer are shown in Figure S10.
39
40

41
42 Volunteers who received links to data reports conducted an average of 5.7 ($\sigma = 4.5$) out of 12
43
44 requested sampling events, and volunteers who did not receive links conducted an average of
45
46 5.6 (4.4) sampling events. Volunteers who received data report links viewed an average of 2.2
47
48 ($\sigma = 2.1$) of the 6 data reports, and the number of reports viewed was significantly, positively
49
50 associated with participation rate (Spearman's $\rho = 0.51$, $p = 0.02$). Note that this correlation
51
52 does not indicate a causal relationship between receiving data reports and participation.
53
54
55
56
57
58
59
60

1
2
3
4
5
6 **Effect of data reports on volunteer retention.** Data from 45 volunteers who completed the pre-
7
8 project survey were included in the model testing the information hypothesis. Among the 21
9
10 volunteers who received data reports, median (range) age category was 35-44 (18-24, 65-74)
11
12 years. Fifteen were female, and 6 were male. Among the 24 volunteers who did not receive
13
14 data reports, median age was 35-44 (18-24, 75-84) years. Twenty were female, and 4 were
15
16 male. Aggregating across treatment groups, the odds of volunteers conducting sampling
17
18 decreased by 4 percent (OR = 0.96, 95% CI: [0.93, 0.99]) with each successive week. The
19
20 information hypothesis model (Table S4), described in Methods, indicates that a 1-unit increase
21
22 in composite obstacles score was associated with a 37 percent reduction (OR = 0.63, [0.46,
23
24 0.86]) in the odds of conducting sampling. A one unit increase in obstacles scores is equal to
25
26 any combination of obstacles scores in Table S2 that sums to 1. For each passing week, the odds
27
28 of conducting sampling among the control group declined by 4 percent, though this was not
29
30 statistically significant. Furthermore, the interaction between time and treatment was not
31
32 statistically significant, indicating that receiving links to data reports did not alter retention.
33
34
35
36
37
38
39
40
41

42 **Association between data reports and retention for learning-oriented volunteers.** Of 45
43
44 volunteers, 7 received data reports and were learning-oriented. Characteristics of these
45
46 volunteers are presented in Table S5. The motivation hypothesis model (Table S4) indicates
47
48 that, adjusting for obstacles to participation, learning-oriented volunteers who received data
49
50 reports had statistically significantly lower retention than other volunteers (*i.e.* volunteers who
51
52 either did not receive data reports or who received data reports but were not learning
53
54
55
56
57
58
59
60

1
2
3 motivated). Specifically, for each one week increase in time, the odds of conducting sampling
4
5 among learning-oriented volunteers who received data reports dropped 10 percent compared
6
7 to other volunteers (OR = 0.90, [0.83, 0.98]). This result is driven by the fall in participation after
8
9 day 60 among learning-oriented volunteers who received data reports, as retention prior to day
10
11 60 is not statistically different between target and control groups (Figure 3).
12
13
14
15
16
17

18 Discussion

19
20 **Safe swimming distances from freshwater discharges to two marine beaches were identified**
21
22 **through participatory science.** Median safe swimming distances (D_{90s}) were 130 m at San
23
24 Pedro Creek and 380 m at San Lorenzo River. San Mateo County, where San Pedro Creek is
25
26 located, currently encourages beachgoers to remain at least 300 ft (90 m) from a flowing creek
27
28 or storm drain when making water contact. This distance corresponds to the 30th percentile of
29
30 D_{90s} measured at San Pedro Creek in this study. To be more health protective, San Mateo
31
32 County could consider increasing the recommended swimming distance to 130 m (420 ft) or
33
34 180 m (580 ft), which would be sufficiently far to result in 90 percent dilution of San Pedro
35
36 Creek water in approximately half or three quarters, respectively, of days in winter and spring.
37
38 We did not measure FIB in our study since high levels of fecal contamination have been
39
40 extensively documented at both sites,^{8,57} resulting in both sites currently being listed on the US
41
42 Clean Water Act 303(d) list.^{48,49} Measuring salinity, rather than FIB, is advantageous because it
43
44 allows the spatial impact from freshwater discharges to be isolated from other nearby FIB
45
46 sources. Efforts similar to those described in this paper could be undertaken with volunteers at
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 other beaches to refine beach management and protect beachgoers from exposure to
4
5 pollutant-laden runoff.
6
7
8
9

10 **Freshwater discharge rate was positively associated with D_{90} at both discharge sites.** In fact,
11
12 the effect size was similar at both sites, despite site-specific differences in beach and coastline
13
14 morphology and wave action. Furthermore, the intercepts of the two models, which
15
16 correspond to the estimated average D_{90} with a 0 m tide (mean lower low), 1 m³/s discharge
17
18 rate, and mean alongshore velocity, were nearly identical, reinforcing the idea that the spatial
19
20 impact of freshwater is similar at both sites. This suggests that the estimated effects of
21
22 discharge rate on D_{90} may be generalizable to other ocean-exposed marine beaches.
23
24
25
26
27
28
29

30 **Lower tides were associated with longer D_{90} s at San Pedro Creek.** The intertidal zone at San
31
32 Pedro Creek is rocky and slopes downward. At low tides, the creek spreads along the beach,
33
34 reducing the momentum with which it meets the surf zone.¹⁶ This lower momentum may result
35
36 in greater wave trapping of freshwater against the shoreline, where salinity measurements for
37
38 this study were made.¹⁶ At San Lorenzo River, where tide level was not associated with D_{90} , the
39
40 tidally-influenced interaction between the river mouth and marine water is more complex. For
41
42 example, the San Lorenzo River mouth morphology can be highly tidally influenced, with
43
44 increasing mouth curvature and width at high tide, and exposed sand bars at low tide forming
45
46 following heavy rains. Thus, potential impacts of tide level on D_{90} may not be captured by this
47
48 simple linear model.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **Alongshore current effects on D_{90} could not be distinguished from other beach-specific**
4 **effects.** The estimated negative effect of alongshore current on D_{90} at San Lorenzo River
5
6 contradicted expectations. Wave-driven alongshore currents in the surf zone are expected to
7
8 have a positive effect on D_{90} by bending the freshwater jet in the current's direction.¹⁶
9
10
11 However, larger alongshore currents may also have resulted in strong rip currents which can
12
13 eject water from the surf zone.¹⁴ Further study, including *in situ* current measurements, is
14
15 required to understand the effects of alongshore currents on freshwater mixing and transport
16
17 in the surf zone at these sites.
18
19
20
21
22
23
24

25 **Although QC resulted in discarding nearly half of collected data, remaining data were**
26 **sufficient in quantity and quality to address the study's research questions.** A common
27
28 concern in participatory science is that measurements made by volunteers may be of poor
29
30 quality.^{31,58} Many studies cite data quality concerns, asserting that volunteer-made
31
32 measurements are inferior in accuracy or precision to measurements made by trained
33
34 scientists.^{59–63} However, studies often do not clearly define reasonable criteria for data quality
35
36 acceptability specific to the methods utilized by volunteers, and recent review articles point out
37
38 that data quality inferiority assertions are frequently supported by weak evidence or biased
39
40 assumptions.^{63,64} Although acceptable data quality will be specific to each project's methods
41
42 and objectives, there is a lack of systematic data quality evaluation present in the participatory
43
44 science literature. This study sought to rigorously assess volunteer-collected data quality by
45
46 thoroughly assessing the accuracy of the instruments used by volunteers, implementing
47
48 sampling procedures that reduce these instruments' inaccuracies, and defining accuracy criteria
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 specific to these instruments under field conditions so that instrument accuracy and volunteer-
4
5 associated errors were not confounded. Although our QC criteria resulted in discarding 45
6
7 percent of all measurements, the remaining data were of high quality and sufficient in number
8
9 to investigate the stated water quality research questions. The QC constructs of data
10
11 completeness, reasonableness, and usefulness proposed in this study are sufficiently broad that
12
13 they could be applied to other studies and operationalized with project-specific QC criteria. QC
14
15 failures in each of these three categories can be ameliorated by improving sampling protocols
16
17 and providing feedback to volunteers about how to make data more complete, reasonable, and
18
19 useful.
20
21
22
23
24
25
26
27

28 **Volunteer retention did not differ between volunteers who received data reports and those**
29 **who did not.** Previous findings from survey-based analyses indicate that it is highly important
30
31 to volunteers that results from their work are shared with them.³⁴ Experimental findings from
32
33 this study do not support the proposition that regularly sharing results with volunteers via a link
34
35 in an email affects retention. This may be due to data reports not providing additional learning-
36
37 oriented benefits beyond those experienced by all participants through other project
38
39 experiences, such as training and sampling. This hypothesis is supported by responses in mid-
40
41 and post-project surveys indicating that learning-oriented benefits of participating were high
42
43 (>4 out of 5) for both treatment and control groups. Furthermore, volunteers who received
44
45 data reports did not report experiencing learning-oriented benefits different from other
46
47 volunteers in either survey (Wilcoxon rank sum, $p > 0.05$). Note that a randomized experimental
48
49 design enables us to distinguish the positive correlation between viewing data reports and
50
51
52
53
54
55
56
57
58
59
60

1
2
3 participation rates among volunteers in the treatment group from the null causal effect of
4
5 receiving data reports.
6
7
8
9

10 **Volunteers who were learning-oriented and received data reports had lower retention than**
11 **other volunteers.** A second, explicitly functional psychological hypothesis was also tested to
12
13 determine whether volunteer motivations moderated the effect of the informational
14
15 intervention. Specifically, since the informational intervention was designed to increase
16
17 learning-oriented benefits of volunteering via increased access to study data, it was
18
19 hypothesized that learning-oriented volunteers would be most positively affected. Model
20
21 results, however, indicated the opposite: learning-oriented volunteers who received the
22
23 informational intervention had statistically lower retention than other volunteers. Since
24
25 learning-oriented volunteers who received data reports did not report a statistically significant
26
27 difference in learning-oriented benefits compared to other volunteers in either mid- or post-
28
29 project survey, this finding is not easily explained by the functional psychology framework used
30
31 in this paper.
32
33
34
35
36
37
38
39

40 Several post-hoc explorations were undertaken to better understand the reduced retention
41
42 of learning-oriented volunteers. First, we hypothesized that, by providing evidence to
43
44 volunteers that the data they collected were being quickly processed and used by project staff,
45
46 data reports may have enhanced volunteers' perception that they were contributing to a
47
48 worthwhile project. This perception would correspond to an increase in perceived values-
49
50 oriented benefits relative to learning-oriented benefits. To test whether the intervention
51
52 affected retention of values-oriented volunteers, the motivation regression model was fit for
53
54
55
56
57
58
59
60

1
2
3 values-oriented volunteers. Model results did not indicate that the intervention statistically
4
5 affected retention of values-oriented volunteers compared to other volunteers.
6
7

8 Second, we surmised that learning-oriented volunteers may have had lower retention than
9
10 other volunteers, regardless of receiving data reports. However, our modeling suggested that,
11
12 ignoring treatment condition, learning-oriented volunteers did not differ from other volunteers
13
14 in terms of retention. Thus, no evidence was found that lower retention for learning-oriented
15
16 volunteers who received data reports was driven by lower retention among all learning-
17
18 oriented volunteers.
19
20
21

22 Finally, it is possible, although counter to expectations, that learning-oriented volunteers
23
24 perceived the data reports particularly negatively, driving down their retention. In the final
25
26 survey, volunteers who received data reports were asked whether data reports had a positive
27
28 impact on their volunteering experience. Unfortunately, an insufficient number of people
29
30 responded to the final survey to partition respondents by motivational orientation. Thus, the
31
32 hypothesis that learning-oriented volunteers perceived data reports more negatively than
33
34 others could not be tested.
35
36
37
38

39 These post-hoc analyses did not explain the finding that learning-oriented volunteers who
40
41 received data reports, intended to be a learning-oriented benefit, had lower retention than
42
43 other volunteers. Together, these post-hoc analyses and the survey responses regarding
44
45 learning-oriented benefits suggest that receiving data reports had little impact on the perceived
46
47 benefits of volunteering. Given this finding and the small number of volunteers who were
48
49 learning oriented and received data reports, the association between learning-oriented
50
51
52
53
54
55
56
57
58
59
60

1
2
3 volunteers who received data reports and retention may be attributable to unmeasured factors
4
5
6 or spurious.

7
8 It also worth noting that although we classified volunteers as learning oriented or not, many
9
10 volunteers had multiple motivational orientations. We defined a motivational orientation as a
11
12 motivation with a score greater than three and at least as high as all other motivation scores,
13
14 which is aligns with previous work.^{43,44} Under this definition, thirty-one volunteers had a single
15
16 motivational orientation, while fourteen volunteers had more than one motivational
17
18 orientation (Table S6), which supports Clary and Snyder's finding that volunteers are often
19
20 multiply motivated.⁴² In this study, the values orientation was most common among volunteers
21
22 with single motivational orientations (seventeen of thirty-one) and among volunteers with
23
24 multiple orientations, followed by the enjoyment and then the learning orientation. Future
25
26 participatory environmental science studies should consider strategies for increasing retention
27
28 that explicitly target values-oriented motivations, potentially in conjunction with enjoyment or
29
30 learning motivations.
31
32
33
34
35
36
37
38
39

40 **Participation rates did not differ between the 2019 study described in this paper and a 2017**
41 **pilot study.** In contrast to the 2019 study, the pilot study was designed to have minimal contact
42
43 and feedback between volunteers and project staff. No in-person training was conducted, and
44
45 no project update emails were sent during the project unless volunteers explicitly asked for
46
47 input. Sampling protocols were very similar, occurred at the same sites during the same season,
48
49 and the volunteers (n = 53), although all different individuals, were recruited by the same
50
51 means from the same population of local NGOs and beachgoers. Interestingly, average
52
53
54
55
56
57
58
59
60

1
2
3 participation rates did not statistically differ between the pilot study and the study presented in
4
5 this manuscript (Figure S11), despite the increased time required of project staff in the 2019
6
7 study to provide training and data reports.
8
9

10 11 12 13 **Conclusions**

14
15 This study presents a novel framework for estimating safe swimming distances at beaches and
16
17 is the first participatory environmental science study to experimentally test strategies for
18
19 increasing volunteer retention. The empirical estimates of freshwater plume extents presented
20
21 in this work will be useful for future surf zone circulation studies. They also provide scientific
22
23 grounding for guidance provided to swimmers by beach managers about safe swimming
24
25 distances from freshwater discharges. This work shows that by implementing thorough data
26
27 quality control measures, volunteers using low-cost instruments can implement complex
28
29 sampling schemes to collect water quality data in the surf zone, which is a complex, data-poor
30
31 environment. Furthermore, although receiving project feedback is commonly cited as
32
33 motivating to volunteers, results from this study's experimental intervention suggest that
34
35 managers of participatory science projects should not rely on disseminating project data via
36
37 email to promote retention.
38
39
40
41
42
43
44
45
46

47 **Acknowledgments**

48
49 We acknowledge the anonymous volunteers for contributing to this project; Monterey Bay
50
51 Aquarium Sanctuary Exploration Center, Santa Cruz Museum of Natural History, Sonlight
52
53
54
55
56
57
58
59
60

1
2
3 Surfshop, and Soul Grind Coffee for serving as sample drop-off locations; and Jana Hennig,
4
5 Jackie Mogensen, and Emily Williams for producing a volunteer recruitment video.
6
7
8
9

10 **Disclosures**

11
12 This study was supported by NSF CBET grant number 1644300. WCJ was supported by the NSF
13
14 GRFP. The authors declare no competing financial interest.
15
16
17
18
19

20 **Supporting Information**

21
22 SI document includes supporting text and figures. Additional documents referenced in text are
23
24 available as separate files, described in SI document. Data gathered and used in this study are
25
26 publically available at <https://purl.stanford.edu/pb938sy7980>.⁶⁵
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Volunteer data QC criteria. Criteria are presented in the order in which they were applied to the data.

Category	Criterion	Description	Data discarded if failed
Completeness	Calibration performed	Calibration measurement recorded	Entire salinity profile
Reasonableness	Validated sample accurate	Volunteer measurement within interval [validation measurement +/- 2.3 PSU]	Entire salinity profile
Completeness	GPS coords fully transcribed	GPS coordinates must have 6 digits following decimal	Single data point
Reasonableness	Salinity range reasonable	Salinity reasonable given acceptable refractometer error of +/- 2.3 PSU: within [-3, 37] PSU	Single data point
Reasonableness	GPS coords on target beach	Coords lie within reasonable beach polygon	Single data point
Reasonableness	GPS coords on shoreline	Shoreline approximated by quadratic fit to all data points; coords must fall within 3 standard deviations of quadratic fit	Single data point
Usefulness	Sufficient number of points per profile	Minimum of 4 points per profile pass quality control criteria; necessary to fit exponential function	Entire salinity profile
Usefulness	Sufficient spatial coverage	Sufficient distance along shoreline and sufficient salinity range measured to accurately estimate D_{90} . See SI for details.	Entire salinity profile

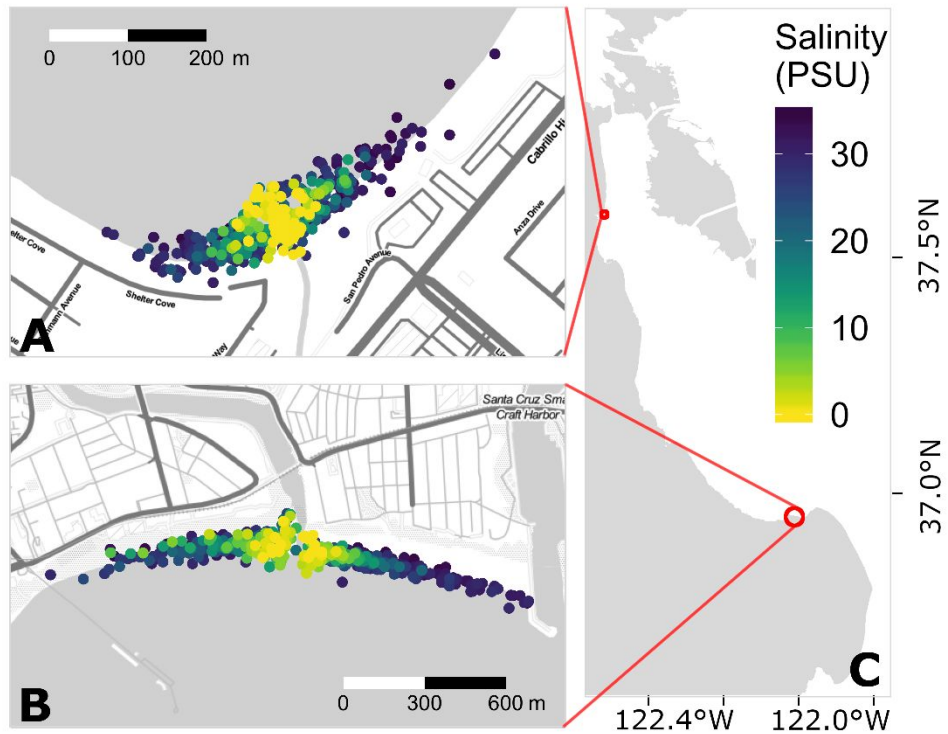


Figure 1. All salinity measurements that passed QC at (A) San Pedro Creek and (B) San Lorenzo River. Note that apparent differences in distance from the water reflect the dynamic location of the land-water interface, which is influenced by tides. (C) depicts the California coast spanning from San Francisco in the north to Monterey Bay in the south.

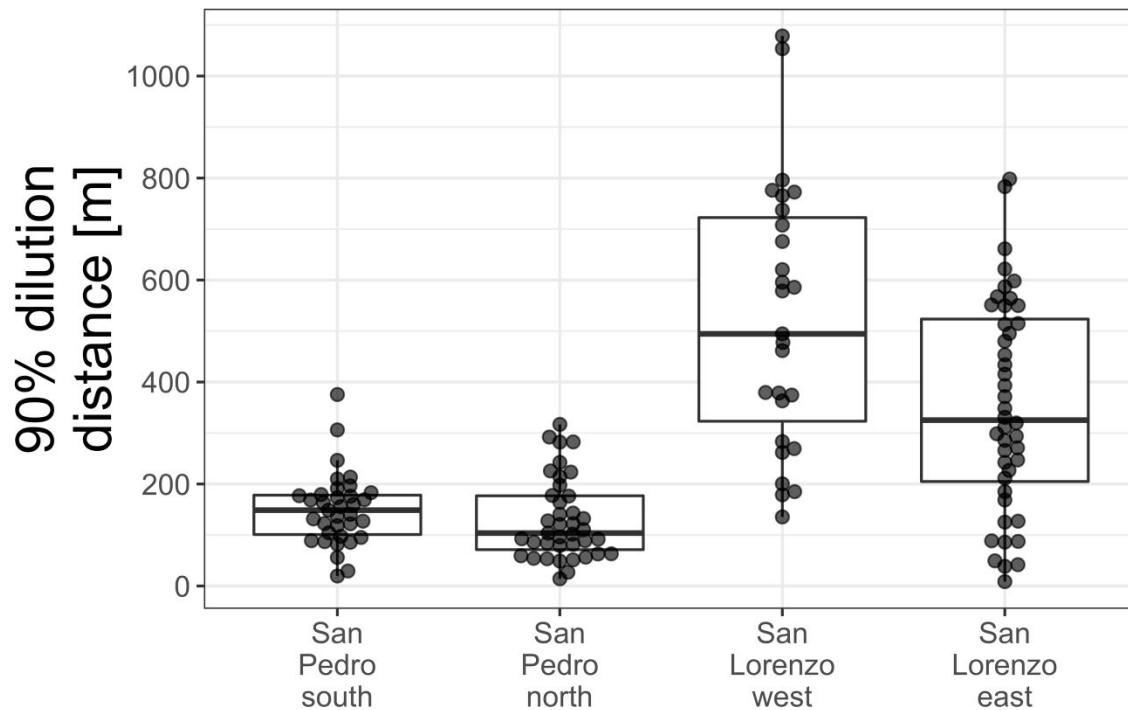


Figure 2. Boxplots representing distributions of D_{90} distances at each beach. Boxplot centerlines are medians, bottoms and tops of boxes are 25th and 75th percentiles, and whiskers extend to the furthest data point within a distance of $1.5 \times$ interquartile range from the 25th and 75th percentiles, respectively. Individual points are displaced horizontally to aid visualization.

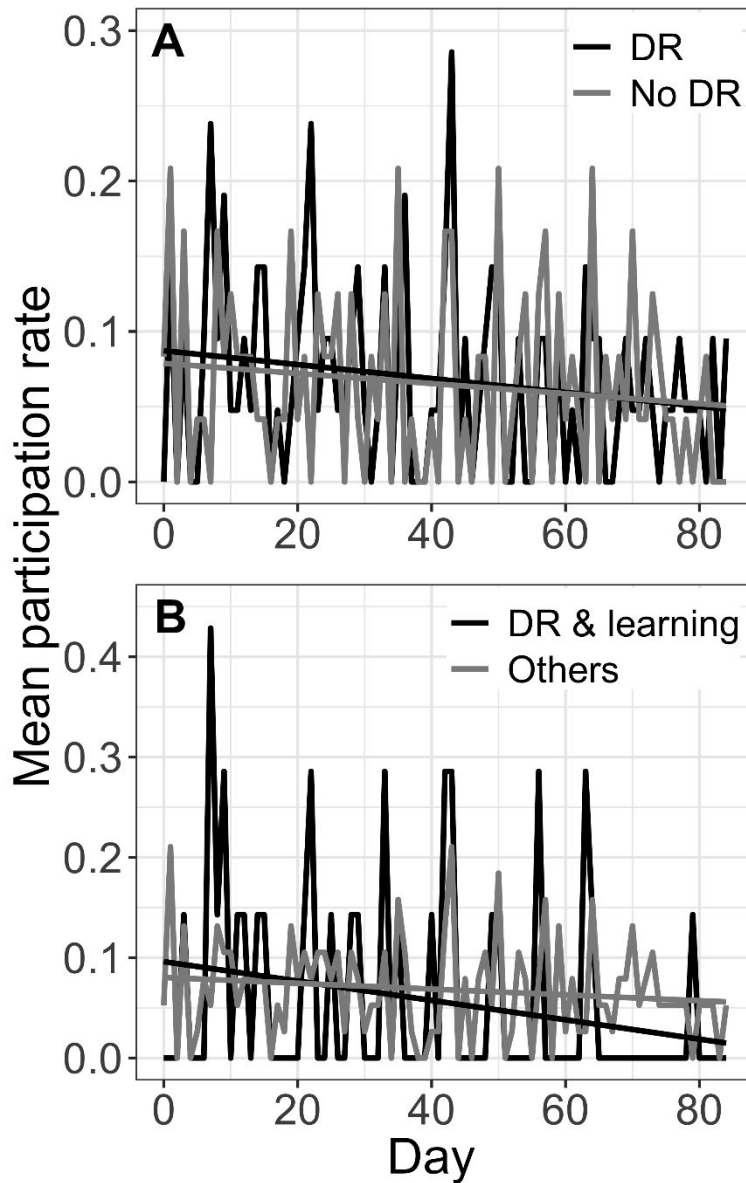


Figure 3. Mean participation rate by study day. Mean participation rate is the number of volunteers out of 45 total who sampled on a given day. Volunteers conducted sampling over an 85-day window. Volunteers are disaggregated by (A) treatment group for the informational hypothesis and (B) target group for the motivational hypothesis. “DR” means received links to data reports; “learning” means volunteers who were learning oriented. Trend lines are linear fits.

References

- (1) Anonymous. The largest cities in the world by land area, population, and density <http://www.citymayors.com/statistics/largest-cities-population-125.html> (accessed Oct 2, 2017).
- (2) Defeo, O.; McLachlan, A.; Schoeman, D. S.; Schlacher, T. A.; Dugan, J.; Jones, A.; Lastra, M.; Scapini, F. Threats to Sandy Beach Ecosystems: A Review. *Estuar. Coast. Shelf Sci.* **2009**, *81* (1), 1–12. <https://doi.org/10.1016/j.ecss.2008.09.022>.
- (3) Barbier, E. B.; Hacker, S. D.; Kennedy, C.; Koch, E. W.; Stier, A. C.; Silliman, B. R. The Value of Estuarine and Coastal Ecosystem Services. *Ecol. Monogr.* **2011**, *81* (2), 169–193. <https://doi.org/10.1890/10-1510.1>.
- (4) Halpern, B. S.; Walbridge, S.; Selkoe, K. A.; Kappel, C. V.; Micheli, F.; D'Agrosa, C.; Bruno, J. F.; Casey, K. S.; Ebert, C.; Fox, H. E.; et al. A Global Map of Human Impact on Marine Ecosystems. *Science* **2008**, *319* (5865), 948–952.
- (5) Halpern, B. S.; Longo, C.; Hardy, D.; McLeod, K. L.; Samhuri, J. F.; Katona, S. K.; Kleisner, K.; Lester, S. E.; O'Leary, J.; Ranelletti, M.; et al. An Index to Assess the Health and Benefits of the Global Ocean. *Nature* **2012**, *488* (7413), 615.
- (6) Schiff, K. C.; Allen, M. J.; Zeng, E. Y.; Bay, S. M. Southern California; 2000.
- (7) Heal the Bay. *2018-2019 Beach Report Card*; 2019.
- (8) Walters, S. P.; Thebo, A. L.; Boehm, A. B. Impact of Urbanization and Agriculture on the Occurrence of Bacterial Pathogens and Stx Genes in Coastal Waterbodies of Central California. *Water Res.* **2011**, *45* (4), 1752–1762. <https://doi.org/10.1016/j.watres.2010.11.032>.
- (9) Harwood, V. J.; Whitlock, J.; Withington, V. Classification of Antibiotic Resistance Patterns of Indicator Bacteria by Discriminant Analysis: Use in Predicting the Source of Fecal Contamination in Subtropical Waters. *Appl. Environ. Microbiol.* **2000**, *66* (9), 3698–3704.
- (10) Layton, B. A.; Walters, S. P.; Boehm, A. B. Distribution and Diversity of the Enterococcal Surface Protein (Esp) Gene in Animal Hosts and the Pacific Coast Environment. *J. Appl. Microbiol.* **2009**, *106* (5), 1521–1531. <https://doi.org/10.1111/j.1365-2672.2008.04113.x>.
- (11) Withers, P. J.; Jordan, P.; May, L.; Jarvie, H. P.; Deal, N. E. Do Septic Tank Systems Pose a Hidden Threat to Water Quality? *Front. Ecol. Environ.* **2014**, *12* (2), 123–130. <https://doi.org/10.1890/130131>.
- (12) Sercu, B.; Van De Werfhorst, L. C.; Murray, J. L. S.; Holden, P. A. Sewage Exfiltration As a Source of Storm Drain Contamination during Dry Weather in Urban Watersheds. *Environ. Sci. Technol.* **2011**, *45* (17), 7151–7157. <https://doi.org/10.1021/es200981k>.
- (13) Sauer, E. P.; VandeWalle, J. L.; Bootsma, M. J.; McLellan, S. L. Detection of the Human Specific Bacteroides Genetic Marker Provides Evidence of Widespread Sewage Contamination of Stormwater in the Urban Environment. *Water Res.* **2011**, *45* (14), 4081–4091. <https://doi.org/10.1016/j.watres.2011.04.049>.
- (14) Inman, D. L.; Tait, R. J.; Nordstrom, C. E. Mixing in the Surf Zone. *J. Geophys. Res.* **1971**, *76* (15), 3493–3514. <https://doi.org/10.1029/JC076i015p03493>.
- (15) Rodriguez, A. R.; Giddings, S. N.; Kumar, N. Impacts of Nearshore Wave-Current Interaction on Transport and Mixing of Small Scale Buoyant Plumes. *Geophys. Res. Lett.* *0* (ja). <https://doi.org/10.1029/2018GL078328>.

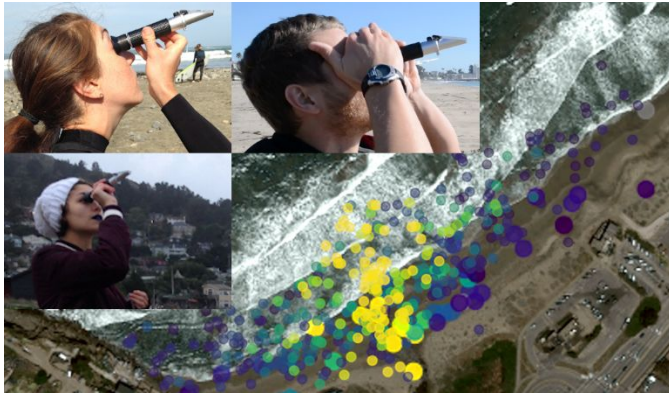
- 1
2
3 (16) Wong, S. H. C.; Monismith, S. G.; Boehm, A. B. Simple Estimate of Entrainment Rate of
4 Pollutants from a Coastal Discharge into the Surf Zone. *Environ. Sci. Technol.* **2013**, *47*
5 (20), 11554–11561. <https://doi.org/10.1021/es402492f>.
6
7 (17) Grant, S. B.; Kim, J. H.; Jones, B. H.; Jenkins, S. A.; Wasyl, J.; Cudaback, C. Surf Zone
8 Entrainment, along-Shore Transport, and Human Health Implications of Pollution from
9 Tidal Outlets. *J. Geophys. Res. Oceans* **2005**, *110* (C10).
10 <https://doi.org/10.1029/2004JC002401>.
11
12 (18) Clarke, L. B.; Ackerman, D.; Largier, J. Dye Dispersion in the Surf Zone: Measurements and
13 Simple Models. *Cont. Shelf Res.* **2007**, *27* (5), 650–669.
14 <https://doi.org/10.1016/j.csr.2006.10.010>.
15
16 (19) Horner-Devine, A. R.; Hetland, R. D.; MacDonald, D. G. Mixing and Transport in Coastal
17 River Plumes. *Annu. Rev. Fluid Mech.* **2015**, *47* (1), 569–594.
18 <https://doi.org/10.1146/annurev-fluid-010313-141408>.
19
20 (20) Feddersen, F.; Olabarrieta, M.; Guza, R. T.; Winters, D.; Raubenheimer, B.; Elgar, S.
21 Observations and Modeling of a Tidal Inlet Dye Tracer Plume. *J. Geophys. Res. Oceans*
22 **2016**, *121* (10), 7819–7844. <https://doi.org/10.1002/2016JC011922>.
23
24 (21) Moulton, M.; Elgar, S.; Raubenheimer, B.; Warner, J. C.; Kumar, N. Rip Currents and
25 Alongshore Flows in Single Channels Dredged in the Surf Zone. *J. Geophys. Res. Oceans*
26 **2017**, *122* (5), 3799–3816. <https://doi.org/10.1002/2016JC012222>.
27
28 (22) Clark, D. B.; Feddersen, F.; Guza, R. T. Cross-Shore Surfzone Tracer Dispersion in an
29 Alongshore Current. *J. Geophys. Res. Oceans* **2010**, *115* (C10).
30 <https://doi.org/10.1029/2009JC005683>.
31
32 (23) McKinley, D. C.; Miller-Rushing, A. J.; Ballard, H. L.; Bonney, R.; Brown, H.; Evans, D. M.;
33 French, R. A.; Parrish, J. K.; Phillips, T. B.; Ryan, S. F.; et al. Investing in Citizen Science Can
34 Improve Natural Resource Management and Environmental Protection. *Issues Ecol.* **2015**,
35 *19*.
36
37 (24) Follett, R.; Strezov, V. An Analysis of Citizen Science Based Research: Usage and
38 Publication Patterns. *PLOS ONE* **2015**, *10* (11), e0143687.
39 <https://doi.org/10.1371/journal.pone.0143687>.
40
41 (25) Delaney, D. G.; Sperling, C. D.; Adams, C. S.; Leung, B. Marine Invasive Species: Validation
42 of Citizen Science and Implications for National Monitoring Networks. *Biol. Invasions*
43 **2008**, *10* (1), 117–128. <https://doi.org/10.1007/s10530-007-9114-0>.
44
45 (26) Bonney, R.; Cooper, C. B.; Dickinson, J.; Kelling, S.; Phillips, T.; Rosenberg, K. V.; Shirk, J.
46 Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific
47 Literacy. *BioScience* **2009**, *59* (11), 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>.
48
49 (27) Bonney, R.; Ballard, H.; Jordan, R.; McCallie, E.; Phillips, T.; Shirk, J.; Wilderman, C. C.
50 Public Participation in Scientific Research: Defining the Field and Assessing Its Potential
51 for Informal Science Education. A CAISE Inquiry Group Report. *Online Submiss.* **2009**.
52
53 (28) McKinley, D. C.; Miller-Rushing, A. J.; Ballard, H. L.; Bonney, R.; Brown, H.; Cook-Patton, S.
54 C.; Evans, D. M.; French, R. A.; Parrish, J. K.; Phillips, T. B.; et al. Citizen Science Can
55 Improve Conservation Science, Natural Resource Management, and Environmental
56 Protection. *Biol. Conserv.* **2017**, *208*, 15–28.
57 <https://doi.org/10.1016/j.biocon.2016.05.015>.
58
59
60

- 1
2
3 (29) Bliss, J.; Aplet, G.; Hartzell, C.; Harwood, P.; Jahnige, P.; Kittredge, D.; Lewandowski, S.;
4 Soscia, M. L. Community-Based Ecosystem Monitoring. *J. Sustain. For.* **2001**, *12* (3–4),
5 143–167. https://doi.org/10.1300/J091v12n03_07.
6
7 (30) Serret, H.; Deguines, N.; Jang, Y.; Lois, G.; Julliard, R. Data Quality and Participant
8 Engagement in Citizen Science: Comparing Two Approaches for Monitoring Pollinators in
9 France and South Korea. *Citiz. Sci. Theory Pract.* **2019**, *4* (1), 22.
10 <https://doi.org/10.5334/cstp.200>.
11
12 (31) Whitelaw, G.; Vaughan, H.; Craig, B.; Atkinson, D. Establishing the Canadian Community
13 Monitoring Network. *Environ. Monit. Assess.* **2003**, *88* (1), 409–418.
14 <https://doi.org/10.1023/A:1025545813057>.
15
16 (32) Brouwer, S.; Hessels, L. K. Increasing Research Impact with Citizen Science: The Influence
17 of Recruitment Strategies on Sample Diversity. *Public Underst. Sci.* **2019**,
18 0963662519840934. <https://doi.org/10.1177/0963662519840934>.
19
20 (33) Wright, D. R.; Underhill, L. G.; Keene, M.; Knight, A. T. Understanding the Motivations and
21 Satisfactions of Volunteers to Improve the Effectiveness of Citizen Science Programs. *Soc.*
22 *Nat. Resour.* **2015**, *28* (9), 1013–1029. <https://doi.org/10.1080/08941920.2015.1054976>.
23
24 (34) Alender, B. Understanding Volunteer Motivations to Participate in Citizen Science
25 Projects: A Deeper Look at Water Quality Monitoring. *J. Sci. Commun.* **2016**, *15* (3), A04.
26 <https://doi.org/10.22323/2.15030204>.
27
28 (35) August, T. A.; West, S. E.; Robson, H.; Lyon, J.; Huddart, J.; Velasquez, L. F.; Thornhill, I.;
29 Freshwater and Citizen Science Research Hackathon Group. Citizen Meets Social Science:
30 Predicting Volunteer Involvement in a Global Freshwater Monitoring Experiment.
31 *Freshw. Sci.* **2019**.
32
33 (36) Domroese, M. C.; Johnson, E. A. Why Watch Bees? Motivations of Citizen Science
34 Volunteers in the Great Pollinator Project. *Biol. Conserv.* **2017**, *208*, 40–47.
35 <https://doi.org/10.1016/j.biocon.2016.08.020>.
36
37 (37) Geoghegan, H.; Dyke, A.; Pateman, R.; West, S.; Everett, G. *Understanding Motivations*
38 *for Citizen Science; UK Environmental Observation Framework, 2016.*
39
40 (38) Measham, T. G.; Barnett, G. B. Environmental Volunteering: Motivations, Modes and
41 Outcomes. *Aust. Geogr.* **2008**, *39* (4), 537–552.
42 <https://doi.org/10.1080/00049180802419237>.
43
44 (39) Nov, O.; Arazy, O.; Anderson, D. Scientists@Home: What Drives the Quantity and Quality
45 of Online Citizen Science Participation? *PLOS ONE* **2014**, *9* (4), e90375.
46 <https://doi.org/10.1371/journal.pone.0090375>.
47
48 (40) Raddick, M. J.; Bracey, G.; Gay, P. L.; Lintott, C. J.; Cardamone, C.; Murray, P.; Schawinski,
49 K.; Szalay, A. S.; Vandenberg, J. Galaxy Zoo: Motivations of Citizen Scientists.
50 *ArXiv13036886 Astro-Ph Physicsphysics* **2013**.
51
52 (41) Tiago, P.; Gouveia, M. J.; Capinha, C.; Santos-Reis, M.; Pereira, H. M. The Influence of
53 Motivational Factors on the Frequency of Participation in Citizen Science Activities. *Nat.*
54 *Conserv.* **2017**, *18*, 61–78. <https://doi.org/10.3897/natureconservation.18.13429>.
55
56 (42) Clary, E. G.; Snyder, M. The Motivations to Volunteer: Theoretical and Practical
57 Considerations. *Curr. Dir. Psychol. Sci.* **1999**, *8* (5), 156–159.
58
59 (43) Clary, E. G.; Snyder, M.; Ridge, R. D.; Copeland, J.; Stukas, A. A.; Haugen, J.; Miene, P.
60 Understanding and Assessing the Motivations of Volunteers: A Functional Approach. *J.*

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Pers. Soc. Psychol.* **1998**, 74 (6), 1516–1530. <https://doi.org/10.1037/0022-3514.74.6.1516>.
- (44) Al-Ubaydli, O.; Lee, M. Can Tailored Communications Motivate Environmental Volunteers? A Natural Field Experiment. *Am. Econ. Rev.* **2011**, 101 (3), 323–328. <https://doi.org/10.1257/aer.101.3.323>.
- (45) Lee, T. K.; Crowston, K.; Harandi, M.; Osterlund, C.; Miller, G. Appealing to Different Motivations in a Message to Recruit Citizen Scientists: Results of a Field Experiment. *JCOM J. Sci. Commun.* **2018**, 17 (1), A1-.
- (46) Santa Cruz County Environmental Health. Water Quality Status <http://scceh.com/waterquality.aspx>.
- (47) California Water Boards. California Environmental Data Exchange Network <http://ceden.org/index.shtml>.
- (48) California Regional Water Quality Control Board Central Coast Region. *San Lorenzo River Watershed Pathogen TMDL*; 2009.
- (49) California Regional Water Quality Control Board; San Francisco Bay Region. *San Pedro Creek and Pacifica State Beach Bacteria TMDL*; 2012.
- (50) Sassoubre, L. M.; Walters, S. P.; Russell, T. L.; Boehm, A. B. Sources and Fate of Salmonella and Fecal Indicator Bacteria in an Urban Creek. *J. Environ. Monit.* **2011**, No. 13, 2206–2212.
- (51) National Oceanic and Atmospheric Administration. 1981-2010 Normals <https://www.ncdc.noaa.gov/cdo-web/datatools/normals>.
- (52) Xiao, X.; White, E. P.; Hooten, M. B.; Durham, S. L. On the Use of Log-Transformation vs. Nonlinear Regression for Analyzing Biological Power Laws. *Ecology* **2011**, 92 (10), 1887–1894. <https://doi.org/10.1890/11-0538.1>.
- (53) US NOAA. Tides and currents <https://tidesandcurrents.noaa.gov/waterlevels.html?id=9414290> (accessed Jul 1, 2019).
- (54) U.S. Geological Survey. National Water Information System data available on the World Wide Web (USGS Water Data for the Nation) https://waterdata.usgs.gov/ca/nwis/uv?site_no=11161000 (accessed May 10, 2019).
- (55) Scripps Institution of Oceanography. Coastal Data Information Program. <https://cdip.ucsd.edu/> (accessed Jun 1, 2019).
- (56) Stevens, A. W.; Logan, J. B.; Snyder, A. G.; Hoover, D. J.; Barnard, P. L.; Warrick, J. A. Beach topography and nearshore bathymetry of northern Monterey Bay, California: U.S. Geological Survey data release <https://doi.org/10.5066/F76H4GCW> (accessed May 1, 2019).
- (57) San Pedro Creek Watershed Coalition. *San Pedro Creek Watershed Assessment and Enhancement Plan*; 2002.
- (58) Conrad, C. C.; Hilchey, K. G. A Review of Citizen Science and Community-Based Environmental Monitoring: Issues and Opportunities. *Environ. Monit. Assess.* **2011**, 176 (1), 273–291. <https://doi.org/10.1007/s10661-010-1582-5>.
- (59) Loperfido, J. V.; Beyer, P.; Just, C. L.; Schnoor, J. L. Uses and Biases of Volunteer Water Quality Data. *Environ. Sci. Technol.* **2010**, 44 (19), 7193–7199. <https://doi.org/10.1021/es100164c>.

- 1
2
3 (60) Nicholson, E.; Ryan, J.; Hodgkins, D. Community Data - Where Does the Value Lie?
4 Assessing Confidence Limits of Community Collected Water Quality Data. *Water Sci.*
5 *Technol.* **2002**, *45* (11), 193–200.
6
7 (61) Safford, H.; Peters, C. A. Citizen Science for Dissolved Oxygen Monitoring: Case Studies
8 from Georgia and Rhode Island. *Environ. Eng. Sci.* **2017**, *35* (4), 362–372.
9 <https://doi.org/10.1089/ees.2017.0218>.
10
11 (62) Lewandowski, E.; Specht, H. Influence of volunteer and project characteristics on data
12 quality of biological surveys. *Conserv. Biol.* *29* (3), 713–723.
13 <https://doi.org/10.1111/cobi.12481>.
14
15 (63) Kosmala, M.; Wiggins, A.; Swanson, A.; Simmons, B. Assessing Data Quality in Citizen
16 Science. *Front. Ecol. Environ.* *14* (10), 551–560. <https://doi.org/10.1002/fee.1436>.
17
18 (64) Specht Hannah; Lewandowski Eva. Biased Assumptions and Oversimplifications in
19 Evaluations of Citizen Science Data Quality. *Bull. Ecol. Soc. Am.* *99* (2), 251–256.
20 <https://doi.org/10.1002/bes2.1388>.
21
22 (65) Jennings, W. C.; Cunniff, S.; Lewis, K.; Deres, H.; Reineman, D. R.; Davis, J.; Boehm, A. B.
23 Raw Data for “Participatory Science for Coastal Water Quality: Freshwater Plume
24 Mapping and Volunteer Retention in a Randomized Informational Intervention.” *Stanf.*
25 *Digit. Repos.* **2020**.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Table of Contents Art**
5
6



20 **Novelty of work.** This study presents a novel framework for estimating safe swimming
21
22 distances at beaches and is the first participatory environmental science study to
23
24 experimentally test strategies for increasing volunteer retention.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60