Spatial Economics of Citizen Science

David M. Kling\textsuperscript{a,}\textsuperscript{*}, Sonja H. Kolstoe\textsuperscript{b}, Tu Ngoc Nguyen\textsuperscript{a}

Abstract

We develop a model of the citizen science data generating process based on random utility theory. Citizen science programs seek to obtain data about natural phenomena from observations collected by non-experts. For some programs, sampling is an outcome of utility-maximizing spatial behavior. As a consequence, sampling by citizen scientists is nonrandom but predictable as a function of individual preferences and site characteristics. Citizen scientists may also measure with error. Our spatial-dynamic model characterizes the dynamics of information generated by citizen scientists about a target ecosystem service across a stylized landscape. We conduct a simulation analysis of learning from citizen scientist data compared to sampling counterfactuals that do not build on individual preferences. We find that while learning from utility-maximizing citizen scientists produces relatively less accurate estimates of ecosystem service states, performance is heterogeneous and in popular locations can exceed counterfactuals that prioritize minimizing average uncertainty. Utility-seeking citizen scientists generate better estimates when provided with updated information about site-level characteristics they value. Our model may be adapted to study design elements of citizen science programs, in particular incentives or behavioral nudges to improve sampling coverage. Measurement error has a sharply negative effect on learning, however spatially-targeted incentives can serve as a tool to compensate locally for error by increasing citizen scientist participation in one or more targeted locations.

Keywords: Uncertainty, learning, measurement error, spatial-dynamic, ecosystem services, citizen science, random utility

JEL Codes: C25, C63, D83, Q26, Q57

\textsuperscript{*} Corresponding author. Email address: david.kling@oregonstate.edu
\textsuperscript{a} Department of Applied Economics, Oregon State University
\textsuperscript{b} Department of Economics and Finance, Salisbury University

This paper was prepared for the USDA Workshop Applications and Potential of Ecosystem Services Valuation within USDA – Advancing the Science, Washington, DC, April 23-24, 2019.
1. Introduction

A rapidly-developing literature on ecosystem service valuation is uncovering the contribution of natural capital and the services it provides to economies. Prioritization of ecosystem service values is also increasingly observed in regulatory decision making. For example, U.S. National Forest land management plans are now explicitly required to focus on promoting ecosystem service (ES) provision (USFS 2017). Most ecosystem service valuation (ESV) methods, and policy instruments intended to promote ES service provision, depend on substantial amounts of information. Requirements may include some combination of the “ecosystem service production function,” or functional relationship mapping of biological and physical inputs to service outputs, and the current magnitudes of the stocks and flows involved in service production. To these needs, many methods additionally rely on data and models that can describe linked economic sectors, such as a regional tourism economy that benefits from habitat provision. Nonconsumptive outdoor recreational activities, such as bird watching, in general add to the measurement challenge since passes or permits are usually not required to participate, unlike consumptive activities such as fishing or hunting.

When considering information needs, the economics of ecosystem services may be interpreted as adding to well-known challenges presented by traditional single-stock natural resource systems. Natural capital is unlike built capital in that it is usually not engineered—meaning its dynamics may only be partially understood—and it is often measured irregularly, subject to error or patchy coverage (LaRiviere et al. 2017). The recent controversy about a potential rapid decline in insect abundance point to the surprising level of uncertainty about abundances of ecologically important stocks (Jarvis 2018). As a consequence, relative to valuation methods applied to other forms of capital, empirical methods for ESV are adapted to make best use of coarse or incomplete data, even when markets exist for final ES goods and services (e.g., Smith et al. 2017).

A direct route to addressing information demands of ESV is, of course, more and better data, and more accurate sub-models of service provision and linked economic sectors. A growing source of data on ES is citizen science. Citizen science programs enlist volunteers, mostly without formal scientific training, to monitor one or more ecosystem services (McKinley et al. 2017). These projects fill a critical void, often providing data where previously none or little was gathered due to it being cost prohibitive for researchers or government agencies. However, a continuing debate about the role of citizen science is characterized by skepticism about its usefulness as a monitoring tool. Johnston et al. (2018) and van Strien et al. (2013) use occupancy models to look at the sampling bias created by opportunistic reports, such as those produced by citizen scientists. One finding of note from Johnston et al. (2018) is that a citizen scientist’s expertise increases species reporting rates, and that by recording their identity and controlling for detectable covariates, they can control for sources of biases associated with such reports. Robinson et al. (2018) point to the potential for machine-learning to address the spatial bias commonly observed in citizen science projects. This literature largely ignores the fact citizen scientists are individuals with preferences that determine their participation behavior.

Our paper focuses on the provision of information from citizen science programs. Rather than study what citizen scientist behavior signals about individual welfare contributed by ES, as done by Kolstoe and Cameron (2017) and Kolstoe, Cameron and Wilsey (2018), our focus is on
what that same behavior may reveal about uncertain ecosystem services. Information about the physical status of an ecosystem service is likely to have public good value additional to the service’s contribution to current citizen scientist welfare (e.g., bird watcher utility). In addition, understanding the behavioral properties of a citizen science program can help predict its information endpoints, and perhaps also identify opportunities for improving its design.

This paper introduces what we believe is the first economic model of the citizen science data generating process. The model describes the spatial-dynamics of information about a target ecosystem service that is generated by citizen scientists as “samplers.” The foundation of our analysis is a random utility model that maps ecological and economic information to citizen scientist location choice. Trip behavior in turn generates observation data that is used to update the citizen science program’s beliefs about the target ES over time. Our numerical implementation of the model draws functional forms and parameter values from newly-available empirical work on citizen scientist behavior (Kolstoe, Cameron, and Wilsey 2018). We use the model to investigate the influence of feedback between learning about the ES and citizen scientist trip choice. In addition to exploring how utility-maximizing behavior drives the data generating process, we also consider policies for influencing ES monitoring using incentives. Our view is that this system presents the natural resource economic literature with a novel spatial-dynamic process where the endpoint of interest is information rather than extraction.

Our results characterize the influence of individual citizen scientist preferences on aggregate learning outcomes. Compared to sampling counterfactuals that prioritize minimizing uncertainty (“guided sampling”), estimates of ES states obtained from utility-maximizing citizen scientists are overall less accurate. Although error is generally higher, it is spatially heterogeneous: lower in popular sites and higher in unpopular sites. The pattern of error generated by utility-maximizers is influenced by information sharing. When the citizen science program shares updated information on ES estimates over time, typically through a web and/or app based platform for the project, local as well as overall error can decrease. Introducing a small degree of misclassification, or measurement error in citizen scientist observations, sharply reduces the accuracy of ES estimates, although this effect extends to all counterfactuals considered. When we consider the performance of spatially-targeted incentives, one surprising result is the relative robustness of learning in the targeted site as measurement error increases. This finding points to the potential information value of citizen scientist program design elements that account for individual preferences. One message from our analysis is that an understanding of preferences, in addition to citizen scientist properties as observers, is important for accurately predicting the effects of incentives on learning.

The remainder of the paper is organized as follows. The next section details our methods, including the structure of our spatial-dynamic model and approach to analyzing it numerically. Results from our analysis are presented and discussed in Section 3. Following a concluding discussion (Sections 4 and 5), we include an appendix that details our numerical model structure.
2. Methods

2.1. Model

We introduce our spatial-dynamic numerical simulation model in this section. The foundation is a random utility model (RUM) of citizen scientist spatial behavior. A linked ES observation model describes how data is produced from individual choice. Alongside the outcome of utility-maximizing citizen scientist behavior, we generate counterfactual patterns corresponding to random sampling and “guided” sampling, or sampling designed to prioritize reducing uncertainty. The numerical implementation of the model involves specifying functional forms and a simulation protocol. We review each of these components in turn below.

2.1.1. Citizen scientist behavior

Our model of an individual citizen scientist generates spatial behavior during the course of a year, and has two levels of choice. First, on a monthly time step $t$, individual $n$ decides how many “trips,” or citizen science participation events, to take within a fixed landscape consisting of $J$ sites. A natural RUM framework for this choice would be a multi-level nest, however for reasons explained below related to current empirical limitations, we instead adopt an unstructured approach here and define the monthly individual-level trip count $k_{nt}$ as an iid random number drawn from an individual-specific density.\(^1\) Note that a stochastic realization of $k_{nt}$ may be zero, meaning citizen scientist $n$ does not participate in month $t$.

The indirect utility an individual $n$ receives from visiting site $j$ in month $t$, $U_{njt}$, is defined as follows:

$$U_{njt} = V_{njt} + \varepsilon_{njt} = X_{nj} \beta_n + W_{njt} \gamma_{nt} + \varepsilon_{njt}$$

Here $X_{nj}$ and $W_{njt}$ are, respectively, length $d$ and length $h$ row vectors of attributes for individual $n$ corresponding to site $j$. The vector $X_{nj}$ contains attributes that may vary over space but do not change with time. Individual income less the cost of a trip to site $j$ is one example. The vector $W_{njt}$ contains attributes that may change from one month to another. Parameter vectors $\beta_n$ and $\gamma_{nt}$ interact with site attributes and may differ across citizen scientists. Each individual receives an idiosyncratic shock $\varepsilon_{njt}$ to indirect utility. We assume $\varepsilon_{njt}$ is an iid draw from an extreme value distribution that is common to all citizen scientists.

Combining these assumptions, a well-known and convenient property of this form of random utility model is that, conditional on participation, utility maximization implies a multinomial density that characterizes stochastic spatial behavior. While choice is stochastic, the density that produces choices is a deterministic function of $\{V_{n1t}, \ldots, V_{nJt}\}$ (Train 2009). A consequence is that the probability of a citizen scientist visiting a site is heterogeneous across

---

\(^1\) We elaborate on this choice further in Sections 2.1.3 and Appendix Section A.1 below.
individuals and may change over time in response to choice determinants (through $W_{njt}$) or tastes (through $\gamma_{nt}$).

### 2.1.2. Observation, learning, and sampling counterfactuals

The ecosystem service of interest in our model is habitat provision. We assume that citizen scientists derive utility from observing a rare species. The spatial distribution of this rare species is taken by the citizen science program to be a useful indicator of habitat provision. The rare species occupies a subset of the $J$ sites. The true abundance of the species in a given site is uncertain. Instead, citizen scientists base their site choices in part on the expected probability of sighting the species. Participation outcomes generate data on the probability of sighting the rare species, which the citizen science program uses to learn about the true probability that the species could be observed. Sighting probability serves as a proxy for habitat provision, by way of its correlation with true (but unobserved) underlying abundance.

As samplers, citizen scientists may misclassify the rare species. Adopting terminology from the public health literature, given a true probability of the rare species being observable at a site, $\rho_j$, we consider a sensitivity $s_n$ (Rahme and Joseph 1998). This is the probability of individual $n$ correctly classifying the species upon observing it on a trip. There is also a specificity probability $c_n$ of the individual reporting that the species was not observed when in fact it was truly not observed. This term can also capture the probability of non-reporting. The pair $\{s_n, c_n\}$ summarizes what we call the individual’s misclassification propensity.

Assembling these components, the probability that an individual will report observing the rare species is:

$$\tilde{\rho}_{nj} \equiv \tilde{\rho}(\rho_j; s_n, c_n) = s_n \rho_j + (1 - \rho_j)(1 - c_n)$$  

The second term of Eq. (2) measures the probability of a false positive observation. Observations $y_{nkj}$ are generated as $k_{njt}$ within-month draws from a binomial density with $\tilde{\rho}_{nj}$ as the probability of success.

The citizen science program maintains a belief state consisting of a collection of site-specific univariate beta distributions over possible values of $\rho_j$ on the unit interval. The belief state may be summarized as length $J$ parameter vectors $\alpha_t$ and $\omega_t$. At the beginning of month $t$, the expected value of $\rho_j$ (labeled $\hat{\rho}_{jt}$) and variance $\text{var}_{jt}$ are (Gelman et al. 2013, p. 581):

$$\hat{\rho}_{jt} = \frac{\alpha_{jt}}{\alpha_{jt} + \omega_{jt}}$$  

$$\text{var}_{jt} = \frac{\alpha_{jt} \omega_{jt}}{(\alpha_{jt} + \omega_{jt})^2 (\alpha_{jt} + \omega_{jt} + 1)}$$

The citizen science program learns from observations in a Bayesian sense. Assuming perfect sensitivity and specificity ($c_n = s_n = 1$), site-level Beta-beliefs are conjugate to the binomial observation distributions (Gelman et al. 2013). Given $Y_{jt}$ total successful observations out of $K_{jt}$
site visits across all citizen scientists and prior site-level belief $[\alpha_{jt}, \omega_{jt}]$, the end of month belief is:

$$\alpha_{jt+1} = \alpha_{jt} + Y_{jt}$$

$$\beta_{jt+1} = \beta_{jt} + K_{jt} - Y_{jt}$$

In some configurations of the model, we assume that citizen science program beliefs are common knowledge, and $[\alpha_{jt}, \omega_{jt}]$ enter into time-varying site-level utility component $W_{njt}$ (Eq. (1)).

If either sensitivity or specificity of observations are known but imperfect ($c_n$ or $s_n$ less than 1), Bayesian learning is possible to model but more complicated computationally because Beta-Binomial conjugacy is lost. In order to facilitate comparison of results that include misclassification with those that do not, we build on an approach to approximate Bayesian updating that has been used in the literature on dynamic optimization under uncertainty (MacLachlan et al. 2016; Kling et al. 2017). This method combines numerical Bayesian integration with density projection. Assume, temporarily, that all citizen scientists share the same misclassification propensity ($c_n = c$ and $s_n = s$ for all $n$). The likelihood of observing $Y_{jt}$ reported successful observations out of $K_{jt}$ visits would then be $B(Y_{jt}, K_{jt}, \hat{\rho}(\rho_{jt}))$, where $B(\cdot)$ is the binomial pdf and $\hat{\rho}(\rho_{jt})$ is a compact form of Eq. (2). Using this likelihood function it is possible to numerically compute a posterior belief density for each site. This posterior will in general not be a Beta distribution. To preserve the Beta structure of the model, we “project” a Beta distribution onto the simulated posterior by minimizing the Kullback-Leibler (KL) distance between the simulated posterior and a Beta distribution.3 The result is an approximate Bayes update that remains within the Beta family while accounting for misclassification.

In order to interpret the learning dynamics arising from citizen scientists acting as utility-maximizing samplers, we construct two counterfactuals. Associated with each is a counterfactual belief state. Both are based on how realizations of individual monthly trip count $k_{nt}$ are distributed across the landscape. This ensures that differences in learning outcomes up to month $t$ are due to the sampling pattern and not differences in cumulative sample count.

The first counterfactual involves individuals acting as “random samplers”. As the label implies, each period $k_{nt}$ is distributed randomly across the landscape, with each site having equal probability of being selected. As with the RUM-based model, associated with this counterfactual is a belief state comprised of a collection of beta densities summarized by parameter vectors $\alpha^r_{jt}$ and $\omega^r_{jt}$ (these may be transformed to produce corresponding estimates $\hat{\beta}^r_{jt}$ and $\text{var}_{jt}$ using Eqs. (3)-(4)). Beliefs are updated via Bayesian learning from observations.

---

2 This assumes that the distributions of $c_n$ and $s_n$ are known, which for some applications may not be realistic (Hutchinson, He, and Emerson 2017). Modeling joint learning about $\{\rho_j\}$ and $\{c_n, s_n\}$ is an interesting but complicated Bayesian computation. We consider opportunities for future work on this problem in our concluding discussion.

3 This operation is numerically convenient but does involve root finding. Let $\hat{h}(\rho_{jt})$ be the simulated posterior belief about the true sighting probability $\rho_{jt}$. The Beta distribution $\tilde{\alpha}(\tilde{\omega})$ that minimizes the KL distance with $\hat{h}(\rho_{jt})$ solves the system of equations $E_{\tilde{h}}[\ln \rho_{jt}] = \psi(\tilde{\alpha}) + \psi(\tilde{\alpha} + \tilde{\omega})$ and $E_{\tilde{h}}[\ln (1 - \rho_{jt})] = \psi(\tilde{\omega}) + \psi(\tilde{\alpha} + \tilde{\omega})$, where $\psi(\cdot)$ is the digamma function (MacLachlan et al. 2016).
The second counterfactual is “guided sampling”. This involves forming site selection matrix $P_t^g$ with a total of $J$ entries. Each entry of $P_t^g$ is inversely proportional to the corresponding variance implied by the belief state component $[\alpha_t^g, \omega_t^g]$ (specifically, the site-level guided sampling belief state variances $\text{var}_{i_t}^g$ derived from Eq. (4)). Guided sampling amounts to Bayesian learning via a process of (noisy) targeting of belief variance.

Before advancing, we first note two simplifying assumptions. The true probability of sighting the rare species, $\rho_j$, is assumed to be fixed across months. This choice is made to avoid a much more complicated dynamic Bayesian tracking problem (e.g., MacLachlan et al. 2016). Second, while it would be reasonable to assume that individual citizen scientists maintain private beliefs about site-level target species sighting probability, we assume these beliefs are either fixed across periods or updated in common with other individuals (common knowledge) through information feedback from the citizen science program.

### 2.1.3. Site arrangement and parameter values

Our analysis is based on $5 \times 5$ grid of sites, which vary in terms of characteristics that influence site choice. Wherever feasible, numerical choices for individual citizen scientist utility functions, site attributes, and the observation model are based on research on the eBird program by Kolstoe, Cameron, and Wilsey (2018) (hereinafter abbreviated KCW). We draw from KCW due to the unique status of eBird as a long-running and comparatively data-rich citizen science program. However, while KCW provides a convenient empirical anchor, we intend for our model to be stylized and easy to interpret. Our goal is to capture essential economic content of a broad range of citizen science programs.

The appendix provides full details on the functional forms and parameter values we adopt to implement the model numerically. Here we highlight selected features and summarize the combined results of the model parameters in our baseline case through the resulting influence on citizen scientist spatial behavior. Our strategy is to establish a stylized but plausible “spatially-explicit” landscape arrangement of sites. The pattern is spatially-explicit because attributes are arranged to represent a spatial relationship among sites. For example, we assume there is a “home site” in the top-most middle site (site 11), from which citizen scientists originate. The monetary cost of a trip increases radially out from this site (see also Section A.1 of the Appendix). We assume an opposing orientation of the true probability of sighting the rare target species (Figure 1 Panel B). The probability is highest in the bottom left-most cell (site 5), and radiates out in an exponentially decreasing pattern.
Figure 1. Spatial structure of inputs to the citizen science model. (A) Normalized average of individual-level initial probability of choosing to take a trip to a site (conditional on choosing to take a trip) for the RUM sampling alternative, baseline case. (B) True site-level probability of sighting a rare species.

Note: Axis ticks are intended to assist with identification of site number. Sites are enumerated from top-to-bottom, left-to-right column-wise. For example, the upper left-hand corner cell in each panel is site 1, and the middle cell in the top row is site 11. See also the numbering provided in Figure 3 below.

To illustrate the cumulative effect of exogenous site attributes and baseline model parameters on citizen scientist spatial behavior, we compute the initial normalized average probability of visiting a particular site in the baseline case, $\bar{P}_0$. This computation is most representative of spatial behavior across periods for the baseline case, since in later cases we allow for influences like information feedback that change probabilities over time. Normalization is necessary for the site-level probability averages to sum to one because, as explained in the next section, we follow empirical evidence from KCW and incorporate behavioral heterogeneity across citizen scientists, which produces different conditional probabilities for each individual $(P_{n0})$. The result is shown in Figure 1 Panel A. Choice probability (conditional on taking a trip) on average is highest for sites in the middle range of distance from the home site.

Two additional points are important to observe about $\bar{P}_0$ here. First, as discussed in greater detail below in our analysis of the baseline case for the model, this configuration is based on a spatially-uniform prior expectation of observing the targeted rare species in a particular site. Second, extensions to the baseline case result in site choice probability varying more substantially over time and also across citizen scientists due to preference heterogeneity.

---

4 These conditional probabilities are described as initial since in all cases we incorporate seasonal preference variation following empirical evidence provided by KCW. This has the effect of modestly changing individual level conditional probabilities across months, which in turn shifts the normalized average slightly in later months relative to $\bar{P}_0$. 

8
Where needed, below we point out where versions of the model induce dynamics in site choice probability. Unless otherwise noted, the pattern shown in Figure 1 Panel A is roughly the average of initial conditions for individual site choice probability in the utility maximizer model and therefore a useful reference point for interpreting the results that follow.

2.2. Simulations

The last component of our numerical analysis is the simulation protocol. We study the dynamic behavior of 100 utility-maximizing citizen scientists alongside the random sampler and guided sampler counterfactuals. As described at length in the appendix, citizen scientists are generated as “draws” from empirical distributions derived primarily from KCW. While there are a number of avenues for inducing realistic population heterogeneity, we focus on two: income and taste heterogeneity. Income heterogeneity is maintained throughout our analysis, which means citizen scientists are not uniform even in our simplest baseline case. The result is that citizen scientists, across all sampling counterfactuals, are heterogeneous in terms of differing sequences of monthly trip participation counts \( \{k_{nt}\} \). Following the analysis of the baseline, we activate preference heterogeneity that induces additional heterogeneity via differences in utility parameter vector \( \beta_n \) (Eq. (1)). Lastly, there is also temporal heterogeneity of preferences via \( \gamma_{nt} \), which we include to capture seasonal variation in trip frequency observed by KCW. See Appendix Section A.1 for additional information.

The initial conditions for this model are the initial beliefs for the citizen science program and the trip counts and initial conditional choice probabilities for the citizen scientists. Unless otherwise noted, we assume the citizen science program initially has a uniform prior for the sighting probability in each site \( \alpha_{j0} = \sigma_{j0} = 1 \). We consider this initial condition in order to disentangle the effect of learning from the influence of an informed prior. Citizen scientists incorporate this initial belief into the time-varying vector of site attributes \( W_{njt} \) (Eq. (1)). Our baseline case does no incorporate feedback between citizen science program learning and individual beliefs, however we allow for feedback later in the analysis.

Simulations are performed in Matlab. Each set of results is based on 5,000 dynamic realizations. We take care to construct the simulations so the same exogenous shocks are used to determine utility maximizer, random sampler, and guided sampler behavior. That is, each realization of a particular version of the model may be matched to true counterfactuals, with the deviations in learning outcomes among counterfactuals attributable to differences in sampling behavior.

3. Results

We develop results in this section by starting from a relatively simple baseline case of learning from citizen scientist utility-maximizing behavior. The counterfactual random sampler and guided sampler results serve as comparisons. Subsequent sections introduce results produced from redoing the simulation after adding more behavioral realism to the baseline case. We then consider the effect of incentives designed by the citizen science program on learning, including
when citizen scientists may misclassify observations. A summary of key numerical results is included in Table 1 of Section 4.

3.1. Baseline case

In the baseline case, we assume there is no within-month individual preference heterogeneity, and feedback between citizen science program learning and individual beliefs does not occur. These assumptions act to reduce differences in individual conditional site choice probabilities, however heterogeneity remains due to differences in time-invariant attributes $X_{nj}$ and seasonal variation in preferences via $W_{jt}Y_{jt}$ common to all $n$ (Eq. (1)). For the learning model, we assume no misclassification occurs ($s_n = c_n = 1$). The initial condition for the sighting probability is a uniform prior in each site ($\alpha_0 = \omega_0 = 1$). We note, however, that the assumption of no feedback to individuals in the RUM sampling alternative does not carry over to guided sampling. As described in section 2, feedback between beliefs $[\alpha_t^g, \omega_t^g]$ and conditional probability $P_t^g$ under that alternative through prioritization of high belief variance sites.

Figure 2. Mean absolute error in expected probability of sighting at the end of month 12, baseline case. Panel (A) Utility-maximizing citizen scientists. (B) Random sampler counterfactual. (C) Guided sampler counterfactual.

Results from the simulation illustrate how utility-maximizing citizen scientists results in learning outcomes that differ spatially from the random and guided sampler counterfactuals (Figure 2). For each site $j$, we calculate the mean absolute deviation between the true sighting probability $\rho_j$ and expected sighting probability under the utility-maximizing citizen scientists ($\hat{\rho}_j$) (Panel A), random samplers ($\hat{\rho}_j^r$) (Panel B), and guided samplers ($\hat{\rho}_j^g$) (Panel C).

\footnote{See also section A.1.2 of the appendix.}
alternatives at the conclusion of a year of learning from observations \((t = 12)\). The random and guided sampler counterfactuals are similar in terms of average absolute error (mean of approximately 0.11 versus approximately 0.1 across sites, respectively). The average error of both counterfactuals is also fairly spatially uniform. Comparing the random and guided sampler error results with the spatial distribution of true sighting probability, the relatively higher error in sites at the top and right-hand borders of the landscape is attributable to the absence of the rare species \((\rho_j = 0\) in these sites). As shown below, away from these sites guided sampling produces more accurate estimates relative to random sampling.

The spatial pattern of average absolute error that results from utility-maximizing citizen science choice differs sharply from the counterfactuals (compare Panel A vs. Panels B and C). Average absolute error is the highest of the three alternatives (mean of approximately 0.12 across sites), but spatial heterogeneity is more pronounced. Learning from utility maximizing behavior produces both the most accurate estimate (average absolute deviation of about 0.07 in site 13, the central site in the landscape) and the least accurate estimate (\(~0.18\) in site 8). The preference-driven heterogeneity in site-level trip count for utility-maximizers leads to more or less sampling on average across space relative to the comparatively spatially homogenous random and guided sampler counterfactuals. Unsurprisingly, the sites with the highest and lowest average absolute error are associated with the highest (0.072) and lowest (0.019) normalized average conditional site choice probability, respectively.
Figure 3. One realization of initial and end-of-year beliefs about rare species sighting probability in the baseline case, varying sampling model. True sighting probability overlaid for comparison.
As an illustration, we consider a single realization of matched sampling counterfactuals. Figure 3 shows the Beta belief densities at the end of 12 months of sampling and associated learning.\(^6\) This figure brings the spatial heterogeneity that can arise into focus. For example, in site 4, where the true probability of sighting the rare species is relatively high convergence to the true probability is evident among each of the three behavioral counterfactuals. In contrast, guided sampling noticeably outperforms the utility-maximizer counterfactual in sites 8 and 9, which are unpopular with citizen scientist (Figure 1 Panel A).

![Figure 3. Beta belief densities at the end of 12 months of sampling and associated learning.](image)

**Figure 3.** Beta belief densities at the end of 12 months of sampling and associated learning.

Shifting back to average performance across all realizations, to provide another illustration of the results we compare outcomes from utility maximization and random sampling relative to guided sampling, which recall on average achieves the best results in terms of accuracy. Figure 4 shows relative differences in site-level average absolute error.\(^7\) Comparing utility maximization with guided sampling highlights the spatial heterogeneity in performance (Panel A). Relative to guided sampling, adjacent sites can have either have substantially lower (-20.9% in site 3) or higher (+109% in site 8) average absolute deviation. The pattern of relative difference closely parallels the normalized average site choice probability matrix \(\bar{P}_0\) (Figure 1 Panel A), although the correspondence is not 1-to-1 due to factors including the feedback dynamics of guided sampling and the relationship between the uniform belief about sighting

---

\(^6\) This is realization 1,234 out of 5,000.

\(^7\) While it would be desirable to instead compute the average of relative differences in absolute error using matched counterfactual realizations, we are unable to do so for all realizations because the guided sampler counterfactual occasionally returns the true sighting probability for some sites by the end of month 12. This leads a division by zero problem.
probability, which the citizen scientists maintain in this simple case, and the spatially heterogeneous true sighting probability (Figure 1 Panel B).

3.2. The role of information feedback

In the baseline case, we assume there is no dynamic feedback between citizen scientist choices and preferences in the utility maximization alternative. Specifically, because preferences remain fixed, individual conditional site choice probabilities \( P_n \) are also fixed. This treatment of spatial behavior is a useful starting point for developing properties of the model, but it is at odds with the assumption of rational behavior on the part of citizen scientists. Individuals can be expected to adapt to new information, and empirical evidence from KCW and the broader economic literature on recreator behavior suggests that citizen scientists may experience positive or negative externalities from the trip behavior of other individuals.

![Figure 5. Mean absolute error in expected probability of sighting at the end of month 12, with information feedback and preference heterogeneity. Panel (A) Utility-maximizing citizen scientists. (B) Random sampler counterfactual. (C) Guided sampler counterfactual.](image)

We explore the influence of information feedback on the data generating process by studying a case where the citizen scientist program shares its updated estimates of rare species sighting probability with individuals at the end of each month. In this case, citizen scientist indirect utility of visiting sights is modified by assuming expected rare species sighting is no longer time invariant (included in \( X_{nj} \)). Instead, the citizen science program’s estimate \( \hat{\beta}_{jt} \) is incorporated into each individual’s time-varying \( W_{njt} \) vector in the site-level indirect utility function (Eq. (1)). As a consequence, while the normalized average conditional site probabilities
\(P_0\) reviewed above describe initial propensity to choose a given site for a trip, preferences and therefore individual spatial behavior now shifts over time in response to the dynamics of learning.

In addition to information feedback, we also introduce individual preference heterogeneity for fixed aspects of sites (now \(\beta_n\) differs across individuals). On its own (without information feedback), this latter change allows the model to more closely mimic measured citizen scientist preference heterogeneity reported by KCW, but does not significantly affect learning outcomes relative to the baseline case. We summarize results when preference heterogeneity is introduced on its own in Table 1 (Section 4). The counterfactual random sampler and guided sampler models work as before since neither modification we make to the base case affects the density of monthly trip selections for those versions of the model.

The joint influence of feedback and more realistic behavioral heterogeneity produces potentially counterintuitive results when compared to the baseline case. First, updates to site-level estimates of the probability of sighting the rare species acts to pull citizen scientists to sites that were unpopular given fixed expectations in the baseline case (Figure 5). As a result, the largest average absolute error across sites with feedback is roughly 14% lower relative to the baseline case.8

![Figure 6](image)

**Figure 6.** Performance of utility maximizer relative to guided sampling: result with preference heterogeneity and information feedback vs. baseline case. Panels show the percent difference in average absolute error at the site level.

Relative performance of utility maximizers and random samplers when compared to guided sampling reveals more about the influence of feedback in this case (Figure 6). Since the

---

8 Highest percent difference in average absolute error in both cases occurs in site 8: ~0.16 when there is both information feedback and preference heterogeneity, vs. ~0.18 in the baseline case. See also Table 1.
random sampler counterfactual is unaffected by changes to the utility maximizer behavior model, we instead again show the RUM vs. guided sampler results from the baseline case as a comparison (Panel B). Information feedback to citizen scientists leads to overall better performance. Taking the average across the landscape, relative mean absolute error compared to guided sampling is cut by almost half when citizen scientists are provided updated information on expecting sighting probability (about 10% higher compared to 18% higher in the baseline case). Information feedback also tends to smooth the accuracy of RUM-based sampling. With feedback, utility maximizers improve on the worst site-level performance relative to guided sampling. In site 8, relative mean absolute error is about 80% higher relative to guided sampling, compared to 109% higher in the baseline case. On the other hand, feedback erodes the best site-level relative performance of utility maximizer sampling (relative mean absolute error is -28% in site 23 compared to -35% in the baseline case).

Figure 7. Normalized average of initial individual-level conditional probability of choosing to take a trip to a site, with feedback and the incentive policy.

3.3. Incentivizing information provision

A powerful property of random utility theory is the ability to predict how individual choice drives spatial behavior given a policy intervention, for example the closure of a site to citizen scientists. In what follows, we assume the citizen scientist program chooses to incentivize participants to visit site 10, the cell at the bottom and second from the left in the

---

9 Specifically, Panel (B) of Figure 6 and Panel (A) of Figure 4 show the same results. Because our focus is on comparing results with feedback and preference heterogeneity with the baseline model, Figure 6 shows both panels on the same scale. This may result in slight coloration differences between Figure 6 Panel (B) and Figure 4 Panel (A).
landscape (Figure 7). This site has relative high probability of sighting the rare species (Figure 1 Panel B), but this is not initially known to the citizen science program given the assumption of a flat prior. One motivation for the incentive based on observables may be the unpopularity of the site (Figure 1 Panel A). We carry through the assumptions of preference heterogeneity and information feedback considered in the previous section. The random and guided sampler alternatives are unaffected by the incentive.

The result of the incentive for normalized average conditional probability of site choice is shown in Figure 7. An implication of random utility theory for spatial behavior is that the incentive need only have a desired influence on individual utility. The most direct mechanism would be monetary compensation to reduce travel costs. This is how we operationalize the incentive in this example; relative to the baseline, travel cost to site 10 is reduced from the highest tier to the second-lowest tier (a 69% reduction). However, with knowledge of individual preferences, citizen scientist programs could potentially have an understanding of non-monetary incentives that would be predicted to achieve the same effect. For example, organizing contests or rankings, bounties with non-cash rewards, or facilitating peer effects may achieve the same result.

![Figure 8](image)

**Figure 8.** Mean absolute error in expected probability of sighting at the end of month 12, with information feedback and the incentive policy. Panel (A) Utility-maximizing citizen scientists. (B) Random sampler counterfactual. (C) Guided sampler counterfactual.

Results for this case point to a nuanced effect of the incentive on the performance of learning from the resulting citizen science behavior (Figures 8 and 9). One predictable outcome is that the incentive draws individuals to the targeted cell, driving down the average absolute error relative to guided sampling (-36%). The RUM framework allows us to capture the opportunity cost of the incentive in terms of learning elsewhere in the landscape. The incentive
draws citizen scientists away from the rest of the landscape, increasing the relative absolute error in sites with non-zero probability of sighting on average (~16%) compared to when the incentive is removed (~10%). An implication is that citizen science programs that attempt to influence the spatial behavior of participants should also consider effects on data generated in sites that are not targeted.

![Figure 9](image1.png)

**Figure 9.** Performance of utility maximizer relative to guided sampling: result with and without incentive. Panels show the percent difference in average absolute error at the site level.

### 3.4. Misclassification

Up to this point, we have considered cases where citizen scientists do not misclassify the rare species. The specificity ($s_n$) and sensitivity ($c_n$) terms of the observation equation (Eq. (2)) were each set to 1 for all individuals. In this section, we relax this assumption. We begin with a simple misclassification scenario. We then reevaluate the incentive policy considered above when citizen scientists differ in terms of their misclassification propensity.

#### 3.4.1. Learning with misclassification

We return to the case where preferences are heterogeneous and utility-maximizing citizen scientists are provided an information feedback (Section 3.2). Now, however, individuals may report inaccurate information to the citizen science program. Beliefs about true sighting probability are updated approximately as described above in Section 2.1.2. While the citizen science program does the best it can to learn given the risk of misclassification, errors affect
belief updates. Error in turn propagates back to citizen scientists via information feedback each month.

**Figure 10.** Mean absolute error in expected probability of sighting at the end of month 12, with misclassification. Panel (A) Utility-maximizing citizen scientists. (B) Random sampler counterfactual. (C) Guided sampler counterfactual.

**Figure 11.** Performance of utility maximizer relative to guided sampling: learning with and without misclassification. Panels show the percent difference in average absolute error at the site level. Guided sampling counterfactual in each panel is generated without misclassification.
A simple form of homogenous misclassification involves each citizen scientist having a 10% probability of not reporting an accurate sighting of the rare species ($s_n = s = 0.9$) and a 10% chance of incorrectly reporting a sighting when it does not occur ($c_n = c = 0.9$). We implement this assumption and re-run the dynamic counterfactuals (Figure 10). The influence of even a seemingly minor amount of misclassification error is significant. Compared to the case with perfect classification, across-site average of mean absolute error is nearly 40% higher (approximately 0.17 with misclassification versus ~0.11 without).

Next, we again consider the effect of misclassification on learning from utility-maximizers relative to guided sampling. Misclassification decreases the guided sampler counterfactual overall (across-site average mean absolute error increases to approximately 0.15 up from about 0.1) in a pattern consistent with the broadly uniform pattern observed previously (e.g., Figure 5 Panel C). To illustrate the spatial pattern of error under misclassification, we compare the RUM results to the guided sampler counterfactual without misclassification (Figure 11). This allows us to compare outcomes with a common baseline. The results drive home the spatially-heterogeneous cost of misclassification in terms of the accuracy of sighting probability estimates. Compared to the baseline with guided sampling and perfect classification, misclassification on average across sites leads to sharply higher error in estimates obtained by learning from utility-maximizers (roughly 63% higher mean absolute error on average, compared to ~10%). Estimates for initially-unpopular sites fare the worst (compare Figure 11 Panel A and Figure 1 Panel A). One interesting outcome is that misclassifying utility maximizers do best in relative terms in site 5 (lower left-hand corner), which has the highest true sighting probability (Figure 1 Panel B). Here information feedback leads to a slightly higher average trip count (~9.4 compared to ~8.7) and combines with a relatively high underlying chance of successful sighting to mitigate the effect of misclassification error on final estimates.

3.4.2. Misclassification and incentives

One measure of the value of understanding citizen scientist misclassification propensity is improvement in the ability to predict learning endpoints, which in this system is learning about a rare species. From an economic perspective, misclassification propensity is one of several dimensions of heterogeneity. The spatial-dynamics of learning from utility-maximizing citizen scientists depends on multiple individual characteristics, including misclassification propensity and also other sources of heterogeneity like preferences for site characteristics. To illustrate, we consider the problem of predicting the effect of the same incentive policy considered previously (Section 3.3) under two misclassification scenarios. The “homogenous scenario”

10 If the counterfactual is instead guided sampling with misclassification, utility-maximizers somewhat counterintuitively appear to do better in relative terms: the across-site average of relative mean absolute error is only ~8% higher. In isolation, this increase in relative performance obscures the overall deterioration of learning (on average) from the introduction of a small amount of misclassification.
returns to the uniform misclassification assumption \( s_n = s = 0.9 \) and \( c_n = c = 0.9 \) introduced above, modified this time by the effect of the incentive.

A second “heterogeneous scenario” splits the population of citizen scientists. One group has the same propensity as the \( s_B = c_B = 0.9 \). The other is somewhat less accurate \( s_A = c_A = 0.7 \). Group membership is randomly assigned as a function of income, with higher income increasing the probability of assignment to the more accurate group.\(^{11}\) In this example, 31% of citizen scientists are assigned to the more accurate classifier group.

\( s_B = c_B = 0.9 \) and \( s_A = c_A = 0.7 \)

\( \text{Figure 12.} \) Performance of utility maximizer relative to guided sampling: the effect of an incentive on learning from utility-maximizers with and without misclassification heterogeneity. Panels show the percent difference in average absolute error at the site level. Guided sampling counterfactual in each panel is generated without misclassification.

We compare the homogenous and heterogeneous misclassification scenarios against the common baseline of guided sampling with information feedback and preference heterogeneity but without misclassification (Figure 12). The latter we use again for illustration purposes as an overall more accurate and spatially-uniform counterfactual. As expected, heterogeneity that reduces the average classification accuracy among citizen scientists leads to less accurate beliefs on average (Panel A). The magnitude of the difference is somewhat

\(^{11}\) Individual income \( I_n \) may be interpreted as an element of \( X_{nj} \). See also Section A.1 for more information on the role of individual income in the base model. Here, assignment in group B (instead of group A) follows a binomial distribution: \( B(I_n / \max(I_k)) \). In other words, probability of being an accurate classifier is increasing in income. We adopt this assumption to tie assignment to observable characteristics in the model. Our maintained hypothesis for the purpose of this analysis is that more income correlates with characteristics including leisure time and education, which in turn reduce the likelihood of misclassification.
surprising. The across-site average of relative mean absolute error is more than double in the heterogeneous scenario (\(\sim 131\%\) vs. \(\sim 68\%\)) despite the seemingly modest difference between the two groups.

The outcome for site 10, which is the target of the incentive, is of particular interest. When citizen scientists are overall more accurate classifiers, mean absolute error is driven below the guided sampler outcome by the incentive (-24% approximately). The same reduction in error is not delivered by the policy when misclassification propensity is heterogeneous and on average lower. Instead, incentivized utility-maximizers do slightly worse than guided sampling without misclassification (about 0.73% higher). The gap in performance of the incentive between the two scenarios points to the information endpoint gain provided by more accurate understanding of citizen scientist characteristics. Importantly, characterizing the differences in the prediction of learning dynamics associated with each misclassification scenario depends on the underlying model of preference-driven citizen scientist spatial behavior. However, an unexpected result is that the performance of learning in the targeted site is relatively robust: mean absolute error is about 0.07 with lower and homogenous misclassification error, and 0.09 with higher heterogeneous error among citizen scientists. We return to this point in the discussion that follows.

4. Discussion

In this section, we briefly review numerical results from the cases considered in Section 3. Table 1 below provides a summary. Viewed together, results from the cases developed above indicate that categories of individual preference heterogeneity have varying levels of influence on learning from utility-maximizing citizen scientists. Introducing greater preference heterogeneity to the baseline case has a small effect on learning, while the addition of information feedback has a more substantial positive effect on accuracy. Introducing an apparently small chance of misclassification to the information feedback case increases mean absolute error across sites by over 50% (on average). Misclassification also affects counterfactuals including guided sampling, resulting in utility-maximizer sampling performing better in relative terms despite overall deterioration of learning. The sharp negative effect of misclassification reinforces the emphasis placed by the ecology and data science communities on the challenging inference problem posed by citizen science data (e.g., Robinson et al. 2018).

The incentive policy applied to site 10 has the effect of increasing the accuracy of the final belief for that site, at the cost of lowering accuracy in at least some other sites as a result of diverted trips. One somewhat surprising result is the robustness of the effect of the policy on learning in the targeted site in terms of final mean absolute error. This finding suggests that employing incentives is a potentially important tool for dealing with misclassification. However, as noted above, predicting the effect of incentives and other citizen science program design components on learning should take into account the influence of individual preferences in addition to observation model properties like misclassification. In particular, the degree to which an incentive slows learning in non-target sites will depend on preference-driven substitution patterns in the citizen scientist population in addition to observation error.
Table 1. Summary of numerical results for RUM sampling

<table>
<thead>
<tr>
<th>Case (Figure #s)</th>
<th>Mean absolute error relative to true sighting probability</th>
<th>Mean absolute error compared to guided sampling (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min. (Min. site)</td>
<td>Max. (Max. site)</td>
</tr>
<tr>
<td>Baseline (Figs. 2, 4)</td>
<td>0.0678 (13)</td>
<td>0.1822 (8)</td>
</tr>
<tr>
<td>Preference heterogeneity (N/A)</td>
<td>0.0681 (13)</td>
<td>0.1842 (8)</td>
</tr>
<tr>
<td>Information feedback and preference heterogeneity (Figs. 5, 6)</td>
<td>0.065 (13)</td>
<td>0.1572 (8)</td>
</tr>
<tr>
<td>Incentive policy (Figs. 8, 9)</td>
<td>0.055 (10)</td>
<td>0.1704 (8)</td>
</tr>
<tr>
<td>Misclassification (Figs. 10, 11)</td>
<td>0.0983 (4)</td>
<td>0.2252 (6)</td>
</tr>
<tr>
<td>Misclassification and incentive policy (Fig. 12)</td>
<td>0.0661 (10)</td>
<td>0.2365 (6)</td>
</tr>
<tr>
<td>Heterogeneous misclassification propensity and incentive policy (Fig. 12)</td>
<td>0.0876 (10)</td>
<td>0.3232 (16)</td>
</tr>
</tbody>
</table>

Notes: (a) Guided sampler counterfactual used for each case matches the corresponding model assumptions. For example, guided sampling with misclassification is used as the counterfactual to compute relative performance or RUM sampling with misclassification.

5. Conclusion

Ecosystem service valuation aims to extend conventional non-market valuation to natural capital and derived services that tend to be characterized by difficult-to-measure functional relationships and abundances. As a consequence, economic and natural science information demands are often substantially higher relative to traditional valuation exercises. More plentiful and cheap remote sensing data will be essential to refine ESV methods and open up new classes to study. However, the fundamental uncertainties that characterizes natural capital and derived ES are likely to remain. Information about services—and not just their value—will continue to remain an economically significant endpoint. The economic literature describing
how and why this information is generated is underdeveloped, making this a promising area of future research as new citizen science projects continue to come online every year. Similar concerns apply to geospatial data from volunteered geographic information (VGI) sources such as social media (Flickr, Instagram, etc.) and cell phones, making models using VGI data on human behaviors subject to similar problems.

Citizen science is an example of a spatial-dynamic data generating process that is both relatively new to economics and likely to be significant for the study of ecosystem service values going forward. Our model indirectly follows from an earlier recognition in bioeconomics that human behavior can reveal properties of uncertain natural capital, or ecological “bads” like hypoxia in the Gulf of Mexico (Newbold and Massey 2010; Smith et al. 2017). For the first time, to the best of our knowledge, we describe economic properties of citizen science. Our main contribution to the interdisciplinary literature on citizen scientists is proposing random utility theory as a link between determinants of individual citizen scientist preferences and the spatial-dynamic pattern of observation data. We study how individual adaptation to new information feeds back to the data generating process. Our analysis also illustrates how a behavioral model can be used to predict the influence of program design features, in our example spatially-targeted incentives, might influence learning.

We adopt a variety of simplifying assumptions which could be revisited in future work. First, we model learning about only one ES. Many citizen science programs are interested in multiple information endpoints. Our model could be extended to consider how individual citizen science preferences can influence learning about multiple ES. Another potentially strong assumption is that structure of misclassification, operationalized here as individual-level sensitivity and specificity, are known. In reality, a core challenge of citizen science data is correcting for human error. An important but challenging extension to this paper would address how simultaneous learning about both the true distribution of ES and individual misclassification propensity is influenced by citizen scientist behavior and heterogeneity.
References


Appendix

A.1. Details on model parameter values

We base much of the structure of our numerical simulation model, and our approach to choosing parameter values, on Kolstoe, Cameron and coauthor’s (2017, 2018) (KCW) research on the eBird program in the states of Oregon and Washington.

A.1.1. Trip participation model

We use a two-stage procedure for simulating the probability density of individual $n$ taking a trip in a particular month. Unfortunately, due to the structure of the data KCW consider, a monthly trip participation model is not readily available. To produce a useful structure for our purposes, we assume that the annual level of participation is proportional to income. KCW have only indirect evidence for citizen scientist income. Instead, we produce the binned distribution (histogram) of 5-year estimate annual income in 2011 from the American Community Survey, pooling the states of Oregon (OR) and Washington (WA).\footnote{Query generated from the U.S. Census Bureau American Fact Finder website: factfinder.census.gov.} This particular survey version is cited as a reference point by KCW. We create an income distribution by drawing from this histogram. For each of $n$ draws, individual $n$ is assigned income $I_n$ equal to the midpoint of the bin drawn.

Next, we standardize incomes using the ACS estimates of the OR-WA pooled mean and simulated standard deviation of income. Using these standardized incomes as probability weights via an inverse standard normal CDF transform, we draw from a Poisson distribution

![Figure A1: Inputs to the monthly individual trip count model. (A) Average individual-level trip frequency within a year. (B) Monthly probability of a trip occurring conditional on choice to take a trip during the year.](image)
with mean $\lambda = 1.65$, where 1.65 is the average individual monthly trip count reported by KCW. The result is shown in Panel A of Fig. A1.

To complete our description of trip participation, we use the observed relative frequency of trip month from KCW as multinomial probability weights (Fig. A1, Panel B). The annual number of trips for an individual is used as the number of draws from this empirical distribution. Note that across simulated realizations of citizen science choice dynamics, we fix individual-level trip count, similar to demographic variables like income (see below), but we re-sample the distribution of that count across months (monthly trip choice). In principle, annual trip count could also be re-sampled to increase heterogeneity and perhaps better capture inter-year variability. However, under our approach this would require constructing a more elaborate trip count frequency model, or resampling income – effectively sampling different individual citizen scientists. We regard our current approach to this component of this model in keeping with our focus on behavioral rather than demographic heterogeneity.

### Table A1. Trip site choice determinants conditional on taking a trip

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Range</th>
<th>Source$^a$</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_j$</td>
<td>Trip cost to site $j$</td>
<td>$[6.34,75.86]$</td>
<td>KCW</td>
<td>Unit is USD (2011)</td>
</tr>
<tr>
<td>$\xi_n$</td>
<td>Individual-specific taste heterogeneity</td>
<td>$[-2.31,2.2]$</td>
<td>Distribution of $\xi_n$ drawn from N(0,1)</td>
<td></td>
</tr>
<tr>
<td>$Yd_n$</td>
<td>Deviation in individual $n$’s income from mean</td>
<td>$[-5.73,16.27]$</td>
<td>Simulated from empirical distribution of American Community Survey (ACS) annual income data for Oregon and Washington.</td>
<td></td>
</tr>
<tr>
<td>$R_j$</td>
<td>Other species richness</td>
<td>$[55.46,96.02]$</td>
<td>KCW</td>
<td>Average value reported by KCW is 75.74.</td>
</tr>
<tr>
<td>$S_j$</td>
<td>Scenic site indicator</td>
<td>${0,1}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$U_j$</td>
<td>Urban site indicator</td>
<td>${0,1}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_{nt}$</td>
<td>Expected rare species sighting probability</td>
<td>Simulation-determined</td>
<td>Outcomes for $\hat{\rho}_{jt}$ determined by model simulations and initial conditions.</td>
<td></td>
</tr>
<tr>
<td>$Cg_{jt}$</td>
<td>Congestion in site $j$</td>
<td>Simulation-determined</td>
<td>Outcomes for $Cg_{jt}$ determined by model simulations and initial conditions.</td>
<td></td>
</tr>
</tbody>
</table>

Notes: (a) Unless otherwise noted, source for variable is Kolstoe, Cameron, and Wilsey (2018) (KCW).

Lastly, we note that one consequence of this model, some sampled citizen scientists opt not to take a trip in the single year we consider (Panel A). There are 9 of these individuals generated by the monthly trip count model; zero assignments are due to low income draws. If
annual trip choice were instead re-sampled as described above, this number would likely vary but would often be non-zero.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Valuea</th>
<th>Commentsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_C$</td>
<td>Marginal utility (MU) of income ($x - 1$)</td>
<td>$-0.0362$</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_R$</td>
<td>Fixed MU of other species richness</td>
<td>$0.0105$</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{RT}(t)$</td>
<td>Month-specific change in MU of other species richness</td>
<td>Multiple</td>
<td>Base month in from KCW “Ecological Economics” model specification is January, so $\beta_{RT}(1) = 0$. Range of $\beta_{RT}(t)$ for $t &gt; 1$ is $[-0.038, 0.112]$. Both $\beta_{RT}(2)$ (February) and $\beta_{RT}(7)$ (July) are negative.</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>Standard deviation of fixed MU of other species richness</td>
<td>$0.0219$</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{Yd}$</td>
<td>Shift in MU of other species richness associated with marginal deviation from mean income</td>
<td>$0.00512$</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_S$</td>
<td>Utility from visiting a scenic site</td>
<td>$0.0617$</td>
<td>Average of coefficient estimates for 1(National Wildlife Refuge), 1(National Parks, etc.), and 1(National Forests, etc.) in from KCW “Ecological Economics” model specification.</td>
</tr>
<tr>
<td>$\beta_U$</td>
<td>Utility from visiting an urban site</td>
<td>$-0.651$</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_{p}$</td>
<td>MU of rare species sighting probability</td>
<td>$1.674$</td>
<td>Set equal to coefficient estimate for 1(Expected Endangered Bird Species) in from KCW “Ecological Economics” model specification.</td>
</tr>
<tr>
<td>$\beta_{Cg,1}$</td>
<td>MU of congestion intercept</td>
<td>$1.47$</td>
<td>Calibrated</td>
</tr>
<tr>
<td>$\beta_{Cg,2}$</td>
<td>MU of congestion slope</td>
<td>$-0.62$</td>
<td>Calibrated</td>
</tr>
</tbody>
</table>

Notes: (a) Unless otherwise noted, source for parameter values is the “Ecological Economics” 60-minute maximum travel time to site specification, for pooled Oregon and Washington sample, reported by KCW (Table A2 of the supplementary information).

An important point to note is that congestion ($C_{g,t}$) is excluded from the cases considered in the main paper.
A.1.2. Trip choice model

The indirect utility model (Eq. (1) of the main paper) that drives site choice probability conditional on taking a trip, excluding the idiosyncratic shocks, is:

$$V_{n,j,t} = \beta_C c_j + [\beta_R + \beta_{RT}(t) + \sigma_R \xi_n + \beta_{Ya} Y_{d_n}] R_j + \beta_S S_j + \beta_U U_j + \beta_A A_{n,j,t} + [\beta_{c,g,1} + \beta_{c,g,2} C g_{j,t}] C g_{j,t} \ (A1)$$
Parameters ($\sigma_R$ and the various $\beta_l$) and variables, along with their values and derivations, which appear in Eq. A1 are described in Tables A1 and A2 above. We strive for balance in the main paper between the stylized nature of the analysis and the nuanced empirical foundation for the numerical analysis provided by KCW. In particular, we abuse notation slightly by drawing distinction between time varying attributes and parameters in Eq. (1) of the main paper that is absent from Eq. (A1). One key difference is that in the main paper, the interaction between month indicators and $R_j$ are embedded in $W_{njt}Y_{jt}$. Here, for parsimony they are instead represented as the equivalent time-varying parameter $\beta_{RT}(t)$.

To interpret the baseline normalized average initial probability of site choice (Figure 1 Panel A of the main paper), it is helpful to visualize the spatial configuration of the variables that vary by site (Figure A2 above). Our approach to structuring the baseline case provides it with a spatially-explicit interpretation. As described in the main text, our assumption is that citizen scientists originate from an urban core, the “home site”, located in site 11 (top row, middle column) (compare Figure A2, Panels A and D). Trip costs reflect this assumption, increasing radially out from that site (Panel A). It would be possible to modify this baseline case to perhaps be more consistent with the configuration of the urban site indicator (Panel D), however this simplifying assumption facilitates interpretation of choice probability. Offsetting trip cost is “other species richness”, a desirable attribute which again increases out from the home site. The overall effect of this spatial configuration of attributes is to increase expected utility radially out from the home site, adjusted by a spatially explicit pattern of trip costs and fixed site quality.