VALUABLES
AN RFF/NASA CONSORTIUM

Valuing Satellite Data for Harmful Algal Bloom Early Warning Systems

Stephen C. Newbold, Sarah Lindley, Shannon Albeke, Joshua Viers, George Parsons, and Robert Johnston

WP 22-23
October 2022

About the Authors

Stephen C. Newbold is an associate professor in the Department of Economics at the University of Wyoming.

Sarah Lindley is a PhD candidate in the Department of Economics at the University of Wyoming.

Shannon Albeke is a research scientist with the Wyoming Geographic Information Science Center at the University of Wyoming.

Joshua Viers is a professor and director of the Center for Information Technology Research in the Interest of Society (CITRIS) at the University of California, Merced.

George Parsons is a professor in the School of Marine Science & Policy at the University of Delaware.

Robert Johnston is a research professor and director of the George Perkins Marsh Institute at Clark University.

About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

Sharing Our Work

Valuing satellite data for harmful algal bloom early warning systems[1]

Stephen C. Newbold[1]
Sarah Lindley[2]
Shannon Albeke[3]
Joshua Viers[4]
George Parsons[5]
Robert Johnston[6]

[1] Associate professor, Department of Economics, University of Wyoming.

[2] PhD candidate, Department of Economics, University of Wyoming.

[3] Research scientist, Wyoming Geographic Information Science Center, University of Wyoming.

[4] Professor, Director of Center for Information Technology Research in the Interest of Society (CITRIS), University of California, Merced.

[5] Professor, School of Marine Science & Policy, University of Delaware.

[6] Research Professor, Director of The George Perkins Marsh Institute, Clark University.

1. This research was supported in part through the National Aeronautics and Space Administration cooperative agreement number NNX17AD26A with Resources for the Future to estimate the value of information obtained form satellite-based remote sensing.

Contents

Abstract

In this study we develop an information valuation framework for harmful algal blooms (HABs), and we apply the framework to a case study of outdoor recreation in California. We obtained estimates of the concentration of cyanobacteria from remote sensing satellite data at 100 lakes in California in 2019 from the San Francisco Estuary Institute (SFEI). We developed a new approach to estimate a recreation demand site-choice model that includes a full set of fixed effects for both destinations and origins. We examined the statistical performance of the estimator using simulated data in a Monte Carlo analysis, and we applied the approach using cell phone mobility data, which indicate the total number of visitors from around 1400 ZIP codes to the study lakes. We estimated the value of a perfect early warning system by comparing the total willingness-to-pay for access to the lakes under a counterfactual scenario wherein the presence or absence of HABs at all lakes could be known with certainty before recreators select which site to visit to the value of access under the status quo scenario assuming recreators form expectations about HAB occurrences based on the historic frequencies of HABs at each site. Our benchmark results suggest that the total value of the complete mitigation of HABs at the 100 California lakes selected for our study between April and September 2019 would have been $7.41 million, and the total value of a perfect early warning system would have been $2.46 million.

1 Introduction

Harmful algal blooms (HABs) are excessive growths of otherwise naturally occurring algae that can form green mats on the surfaces of lakes, rivers, streams, and estuaries and often contain species of cyanobacteria that are toxic to animals and people [1]. HABs can have adverse consequences for commercial fisheries [2], recreation [3, 4], property values [5], pets and livestock [6], and human health [7, 8]. HABs have been recorded in all 50 states in the U.S. [6], are growing in frequency and severity due to increased loads of nitrogen and phosphorus in urban and agricultural runoff [9], and are expected to be further exacerbated by climate change [10, 11].

The main objective of this study is to develop and demonstrate a computational framework for estimating the value of satellite data for identifying the presence of HABs in surface waters. Specifically, we examine the role of satellite data in early warning systems for water-based outdoor recreation activities such as boating, fishing, and swimming in the United States. HABs are generally intermittent but vary in their frequency, severity, extent, and duration depending on the timing of nutrient exports from upstream land parcels, rain events, and surface water flow patterns. Reliable early warning systems supported by near real-time satellite imaging data would allow recreators to divert their visits away from water bodies currently experiencing a HAB to un-impacted sites and thereby increase the overall enjoyment of water-based recreation activities, reduce the risks of adverse health effects, and possibly mitigate the regional economic impacts associated with lost visitation days.

The framework developed here can be used to estimate and compare the net benefits of adapting to HABs through the expanded use of early warning systems, and the net benefits of mitigating HABs by adopting agricultural best management practices or other means of reducing nutrient loads to receiving water bodies [12]. We anticipate that analogous questions about the relative benefits and costs of adaptation and mitigation will become increasingly important for a wide range of environmental issues as climate change progresses in the coming decades [13, 14, 15, 16], so this research should have general policy relevance beyond our focus on HABs. We describe several other potential follow-up activities in the Discussion section.

The basic logic of our information valuation approach is summarized in the diagram shown in Figure 1, which describes the information characteristics, decision-maker actions, and environmental outcomes for a generic pair of reference and counterfactual scenarios for which the value of satellite information can be estimated. The reference scenario describes a state of the world with no early warning system in place, and the counterfactual scenario represents an alternative state of the world with all else equal except an early warning system is operational.

In our application described below, we estimate the value of implementing a perfect early warning system relative to a baseline scenario in which recreators know only the historic frequency of algal blooms at each recreation site in their choice set. We use satellite data collected over a span of recent years to characterize the historic frequencies of blooms, and we envision in the counterfactual scenario a state of world in which a fully functioning early warning system uses improved satellite data and highly accurate near-future prediction models to provide forecasts of HABs at each recreation site. The comparison is analogous to a time before reliable weather forecasts when one would choose to carry an umbrella or not based on the historic frequency of rain this time of year (the reference scenario) versus today when highly reliable day-ahead rain forecasts are readily available on widely used personal electronic devices (the counterfactual scenario). Our

INFORMATION	
(Reference scenario)	*(Counterfactual scenario)*
No early warning system in place, recreators' expectations of future HAB occurrences based on average long-run frequencies of previously observed HABs.	Early warning system (e.g., MyWaterQuality Portal) provides near real-time predictions of HABs in major recreational use surface water bodies in California.
DECISIONMAKER ACTIONS	
On each outing, recreators choose a water body to visit for boating, fishing, or swimming given knowledge of average long-run frequencies of HABs.	On each outing, recreators choose a water body to visit for boating, fishing, or swimming activities given predictions of HABs by the early warning system.
OUTCOMES FOR PEOPLE AND THE ENVIRONMENT	
Relatively low HAB avoidance rate by recreators, relatively high rate of disruption of recreational activities and relatively high rate of exposure to cyanotoxins.	Relatively high HAB avoidance rate by recreators, relatively low rate of disruption of recreational activities and relatively low rate of exposure to cyanotoxins.

Figure 1. Information, actions, and outcomes for a reference scenario and a counterfactual scenario involving the influence of harmful algal blooms (HABs) on outdoor recreation activities in California. The reference scenario is meant to represent current conditions, and the counterfactual scenario represents a possible near-future when highly accurate real-time forecasts of HABs are widely available.

impression is that a close approximation of real-time perfect information about HABs is plausible in the near future, but that our reference scenario is a better representation of the current state of the world in California and elsewhere. This comparison produces an ex ante estimate of the value of additional information about HABs, i.e., what is the maximum amount recreators would be willing to pay to know whether a HAB is present at each recreation site in their choice set *before* deciding which site to visit on each choice occasion.

It also is possible to examine other reference-counterfactual comparisons including a range of possible improvements to an existing early warning system based on the availability of more accurate, precise, and timely satellite data. The framework also can be used to estimate the value of joint implementation of early warnings and improvements in site conditions, represented by reductions in the frequency and severity of HABs. These and similar applications of the model will allow us to investigate whether adaptation and mitigation are substitutes or complements in this setting.

This study makes several contributions to the environmental economics literature and to the toolkit of computational methods available for use by economists and environmental scientists to estimate the monetized value of remotely sensed data from satellites or other sources. A large literature exists on the use of outdoor recreation demand models for valuing site access, envi-

ronmental quality, and lost visits [17, 18, 19]. There are a number of foundational studies in environmental economics on the relationship between the value of information and (quasi-)option value [20, 21, 22], and there is a newer literature on VOI applications to health and environmental issues [23, 24, 25, 26]. But to our knowledge no previous studies have used the approach proposed here to estimate the value of near real-time information about environmental conditions at outdoor recreation sites. Our approach can also provide a revealed preference point of comparison to stated preference studies of the value of information on outdoor recreation site conditions [e.g., 27]. At least one previous study has examined the value of HAB predictions for commercial fisheries [28], and work is underway at Resources for the Future on the value of satellite information for predicting HABs [29, 30]. Our study complements previous and ongoing work in this area by providing a random utility maximization framework for estimating the value of satellite data in a large geographic region (not confined to a single water body) on an ongoing basis (not confined to a specific HAB event) accounting for the prevailing (or simulated counterfactual) spatial and temporal distribution of HABs.

Our model generates predictions of recreators' avoidance behaviors based on their own implicit assessments of the relative desirability of alternative recreation sites conditional on the available information at the time of choice [31, 32]. This can provide a means to conduct exposure-response calculations to estimate changes in infection risks [33], economic impacts [34], and health damages [35] with versus without an early warning system in place, or with improvements in the accuracy and precision of an existing early warning system. Finally, we note that much of the previous non-market valuation research on HABs has been conducted in Lake Erie [36, 37], so our California case study expands the geographic scope of the literature and provides an initial indication of the degree of heterogeneity in the social costs of HABs in different regions of the U.S.

2 Methods

The information valuation framework developed in this study is built on a repeated discrete choice random utility maximization (RUM) modeling paradigm [38], which provides a ready link to a class of revealed preference econometric models commonly used in outdoor recreation demand studies [18]. We adapt the standard discrete choice recreation demand model to account for uncertainty in site conditions, and we use the model to estimate the value to recreators of resolving the uncertainty prior to their recreation site choices on each choice occasion.

2.1 Empirical strategy

Our empirical strategy involves three steps, which we outline here in short form and describe in detail in the sub-sections that follow.

In step 1, we develop a model of outdoor recreation site choices based on the conventional RUM modeling framework. The RUM approach was originally developed by McFadden and others [38, 39] and is now used in a wide variety of fields including marketing, transportation, residential location choice, and more. In the standard RUM model, individuals choose one of $j = 1, 2, ..., J$ exhaustive and mutually exclusive options. Individuals are assumed to always choose the option that confers the highest utility, U_{ij}, among all J options. Individuals observe all U_{ij}'s for all options, but the researcher cannot observe all factors that compose U_{ij}, so for econometric analysis the utilities

are decomposed into an observable component, V_{ij}, and an unobservable component, ε_{ij}, such that $U_{ij} = V_{ij} + \varepsilon_{ij}$. Assuming the ε_{ij}'s are identically and independently distributed following a Gumbel (type I extreme value) distribution, the probability that individual i will choose option j, denoted as $y_{ij} = 1$, yields the standard multinomial logit choice probability,

$$\Pr\left[y_{ij} = 1\right] = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}, \tag{1}$$

and the associated log-sum formula for the expected maximum utility per choice occasion,

$$\mathbb{E}\left[\max U_i\right] = \ln\left[\sum_{j=1}^{J} e^{V_{ij}}\right]. \tag{2}$$

Several additional assumptions will allow us to define our principal measure of willingness to pay, which in turn forms the basis of our value of information measure. Specifically, we assume that all recreators share a common expected gross utility (excluding the disutility from foregone travel costs) from visiting a site where a HAB is not present, δ_j, and we assume that utility is linear in income over the relevant range of changes to be evaluated[2] This allows us to write $V_{ij} = \delta_{ij} - \lambda c_{ij}$, where c_{ij} is the cost (in dollars) that individual i must incur to visit site j, and λ is the marginal utility of income. To represent uncertainty about the presence of a HAB at one or more recreation sites, we denote the probability that a HAB will occur at site j on any choice occasion as H_j, and we denote the loss in utility from the presence of a HAB as β, so the site utility if a HAB is present is $\delta_{ij} - \beta$ and the expected utility from visiting site j is $\theta_j = \delta_j - \beta H_j$. Finally, we denote the observable utility of the outside option for individual i as α_i, which represents the utility of not visiting any of the specified recreation sites on a choice occasion and doing something else instead.

With these definitions in place, individual i's willingness to pay for access to the set of recreation sites with no early warning system is [e.g., 40, p 220-235]:

$$WTP_i^0 = \frac{1}{\lambda}\left\{\ln\left[e^{\alpha_i} + \sum_{j=1}^{J} e^{\delta_j - \beta H_j - \lambda c_{ij}}\right] - \alpha_i\right\}. \tag{3}$$

This is a dollar-equivalent value suitable for use in economic benefit-cost analysis and will be our principle measure of the surplus economic value provided by the set of outdoor recreation sites in our study area.

In step 2, we estimate the parameters of the recreation demand model—the θ_j's, α_i's, and λ—using mobility data collected from personal electronic devices that indicate the number of trips taken from each origin (ZIP code) to each destination (lake) in our study area. We develop and apply a new estimation algorithm that allows identification of a full set of fixed effects for all origins and destinations, which to our knowledge is the first application of its kind. We calibrate the utility loss associated with the presence of a HAB at a visited site, β, using summary statistics obtained from several previously-published non-market valuation studies that estimated recreators' willingness-to-pay to eliminate an aglal bloom at a visited lake or a closely related measure.

In step 3, we use the parameterized model to compute the value of information from a perfect

2. This linearity assumption is conventional but is not strictly necessary. It can be relaxed in future applications, but we maintain it in the present study to keep our early experiments with the framework as simple as possible.

early warning system for HABs at 100 lakes in California. With all parameters specified in step 2, we can compute the value of access to the lakes given the status quo information regime, under which we assume that all recreators form their expectations of HAB events based on the historic frequency of HABs at each lake, and given an idealized information regime, under which we assume that all recreators know with certainty at which lakes HABs occur on each choice occasion. The difference between the values of access under the two information regimes is the value of perfect information.

If recreators can learn with certainty whether or not a HAB is present at every site before they choose which (if any) recreation site to visit on each choice occasion, then the willingness-to-pay by individual i for access to the sites is

$$WTP_i^{ew} = \frac{1}{\lambda}\left\{ \sum\nolimits_{s=1}^{S} P_s \ln\left[e^{\alpha_i} + \sum\nolimits_{j\notin s} e^{\delta_j - \lambda c_{ij}} + \sum\nolimits_{j\in s} e^{\delta_j - \beta H_j - \lambda c_{ij}}\right] - \alpha_i \right\}, \tag{4}$$

where $s = 1, 2, ..., , S$ indexes all possible sets of sites where a HAB is present, and P_s is the probability that set s obtains—i.e., the probability that HABs are present at all sites in set s and absent at all other sites. The H_j's and P_s's in equations (3) and (4) provide a means to represent the spatial and temporal frequency distribution of HABs in the study area and can be calculated from historic observations of HABs at each of the recreation sites.

Note that WTP_i^0 in equation (3) represents the *maximized expected value* of access with no early warning system, and WTP_i^{ew} in equation (4) represents the *expected maximized value* of access with perfect information. The difference between these quantities is the expected value of perfect information from the early warning system [e.g., 41, p 258-259], which we denote as VOI_i. This corresponds to individual i's maximum willingness to pay for an early warning system that would provide perfectly accurate and precise HAB forecasts.

Before proceeding to the details of our methods, we can illustrate the basic logic of the estimation strategy and subsequent calculations of VOI—steps 2 and 3 in our empirical strategy—in a simple example with two recreation sites that are equi-distant from a recreator's home and so are equally costly to visit. The gross utility conferred by visiting either site (excluding the dollar cost of travel to and from each site) is δ if a HAB is not present and $\delta - \beta$ if a HAB is present, and the probability that a HAB will be present at each site is H. Without loss of generality, we also can normalize α to 1. This is the most parsimonious version of the model possible that retains the key features necessary for valuing information about unknown site conditions. In this minimal model, the equations above reduce to far simpler functions of only δ, β, λ, and H. Specifically, equation (3) simplifies to

$$WTP^0 = \frac{1}{\lambda}\left[1 + 2e^{\delta - \beta H}\right], \tag{5}$$

and equation (4) simplifies to

$$WTP^{ew} = \frac{1}{\lambda}\left[\left(1-2H+H^2\right)\ln\left(1+2e^{\delta}\right) + 2H\left(1-H\right)\ln\left(1+e^{\delta-\beta}+e^{\delta}\right) + H^2\ln\left(1+2e^{\delta-\beta}\right)\right]. \tag{6}$$

To parameterize the model, we make the following assumptions: if HABs never occur at either site then the average visitation rate is $r_0 = 0.1$ per choice occasion; if HABs always occur at both sites then the average visitation rate is $r_1 = 0.001$ per choice occasion (or 0.007 or 0.03, to examine

the sensitivity of VOI to variations in β relative to δ); and the individual's willingness to pay for access to the sites if HABs never occur at either site is $WTP^0 = 1$. (This assumption has the effect of normalizing VOI relative to WTP^0.) Jointly, these assumptions are sufficient to calibrate the three preference parameters of this minimal model: δ, β, and λ.

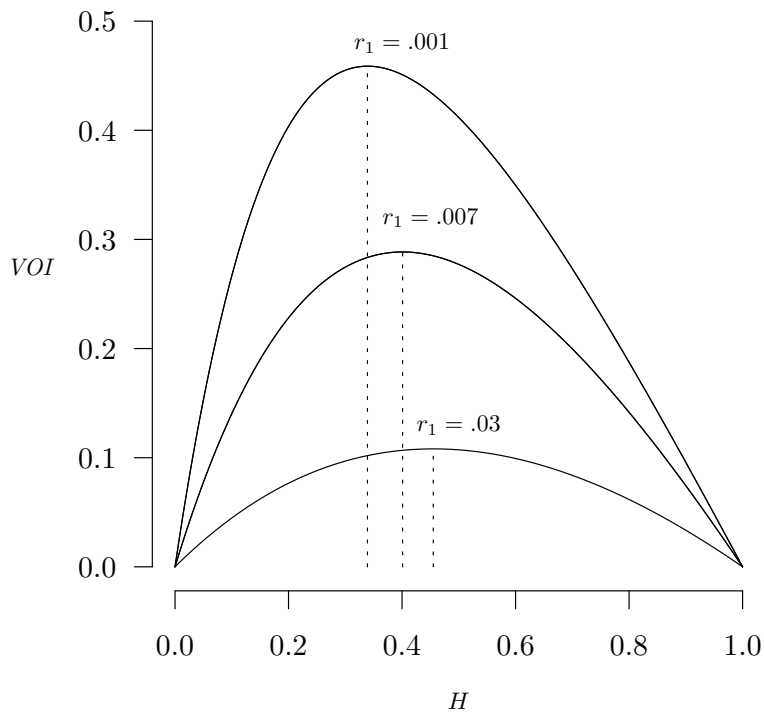

Figure 2. Value of a perfect early warning system (VOI) versus the probability of a HAB at each site (H) in a two-site recreation demand system.

The results of this illustrative exercise appear in Figure 2. The three curves show how VOI varies over the full range of H, the probability of a HAB at either site per choice occasion, for three different values of r_1. If $r_1 = 0.001$, VOI takes a maximum value at around $H = 0.35$ that is roughly 45% of the willingness to pay for access to the unimpaired sites. For higher values of r_1, which imply lower values of β and therefore lower damages from HABs, VOI is lower everywhere and the peaks of the curves shift to larger values of H. The two-site model illustrated here can be extended to any number of sites at any locations in the landscape, and the frequency and correlation of HAB occurrences at the sites can be calibrated to match historic data in the study area. This illustration is highly simplified yet is sufficient to demonstrate the key features of the satellite data valuation approach that we apply in our case study.

2.2 Model development

The foundation of our framework is a standard model of outdoor recreation site choices based on the discrete-choice random utility maximization (RUM) paradigm developed by McFadden and others [38, 39], which was described above. To estimate such a model for our study area, we used data on the outdoor recreation site choices made by a large sample of people who visited the lakes selected for analysis. We obtained mobility data from a private vendor, StreetLight Data, Inc.,[3], which indicates the number of people who visited each lake from each ZIP code in California for the months April–September 2019. Because the mobility data provide information on both origins and destinations based on a large sample of visitors (representing roughly 10% of all trips) for 100 large lakes throughout California, we were able to estimate a high-dimensional site-choice model using a unified dataset and a new estimation approach developed for this study. In the remainder of this subsection we describe our estimation approach, and in section 3.1 we present results from a set of simulation experiments that demonstrate the computational feasibility of the approach and illustrate the potential precision and accuracy of the estimator.

According to the discrete-choice random utility maximization (RUM) model used here, the probability that an individual who lives in zone z will choose option j, denoted as $y_{zj} = 1$, is given by the multinomial logit choice probability function,

$$\Pr\left[y_{zj} = 1\right] = \frac{e^{V_{zj}}}{e^{\alpha_z} + \sum_{k=1}^{J} e^{V_{zk}}}. \tag{7}$$

To estimate the model using the mobility data origin-to-destination visitation frequencies, we first decompose the indirect utility terms into site-specific components that are common to all visitors, representing fixed attributes of the recreation sites, θ_j—which is a lumped parameter including, among other features, the risk of a HAB at each site—and a travel cost component, c_{zj}, which depends on the travel distance between each visitor's origin and the site, i.e., $V_{zj} = \theta_j - \lambda c_{zj}$, where the z subscript indicates that the associated variables refer to average values within each zone (for which we use ZIP codes in this application), and λ is the travel cost parameter. We account for differences in zone-specific factors that affect recreation choices among zones by including a zone fixed effect, α_z, which represents the indirect utility of the outside option (not taking a water-based outdoor recreation trip) for a representative individual in the zone.[4] Denoting the number of people who live in zone z as N_z, and the number of choice occasions in a time period (e.g., month) as D, the expected number of trips by individuals from zone z to site j in a time period is

$$Y_{zj} = N_z D \frac{e^{\theta_j - \lambda c_{jz}}}{e^{\alpha_z} + \sum_{k=1}^{J} e^{\theta_k - \lambda c_{zk}}}. \tag{8}$$

Summing over all zones, the expected total number of visitors to site j in a time period is

$$Y_j = \sum_{z=1}^{Z} Y_{zj}, \tag{9}$$

3. `https://www.streetlightdata.com`

4. Zone-specific factors subsumed in the α_z's could include: the distributions of demographic characteristics among households within the zone, the availability of nearby substitute recreation sites not explicitly represented in the choice set, and unobserved differences in the avidity for outdoor recreation that may emerge due to locational sorting [e.g., 42, 43].

and summing over all sites, the expected total number of trips taken by residents of zone z in a time period is

$$Y_z = \sum\nolimits_{j=1}^{J} Y_{zj}. \tag{10}$$

We will refer to these equations below when we describe our estimation algorithm.

The key innovation in this set-up is the inclusion of a full set of fixed effects for both the sites (destinations) and zones (origins). It is common to estimate multi-site recreation demand models with fixed effects for all destinations (e.g., [44, 45, 46, 47]), but to our knowledge this effort represents the first attempt to include fixed effects for all origins as well. Typically, differences among origin zones are represented by a collection of demographic variables such as average income, education level, household size, and so on for each zone. But the risk of omitted variable bias remains if the list of demographic variables excludes one or more individual-level attributes that influence the demand for outdoor recreation. In particular, there may be unobservable differences in avidity for outdoor recreation that influence people's residential location choices that are not fully represented by observable demographic characteristics. One consequence of this possibility is that the cost of travel to the recreation sites would be endogenous, which would lead to biased estimates of willingness-to-pay [48]. The inclusion of a full set of origin fixed effects mitigates the risk of bias from this form of endogeneity. The inclusion of origin fixed effects also can control for any substitute recreation sites that are not explicitly represented in the choice set. This is important in our setting because we do not include the full universe of California lakes or other possible substitute sites in the model. Nevertheless, with the origin fixed effects included we can still compute unbiased estimates of willingness to pay for improvements at the included destination sites, assuming conditions do not also change at the omitted sites.

The mechanics of our estimation algorithm are outlined in the pseudocode presented in Table I. The algorithm comprises an iterative application of a line search over the travel cost coefficient and dual use of a contraction-mapping algorithm to compute the origin and destination fixed effects, based on the approach first introduced by Berry, Levinsohn, and Pakes [49].

The rationale for step 1 of the algorithm is as follows. First, we rearrange equation (8) and take natural logs of both sides to get:

$$\ln\left(\frac{Y_{zj}}{N_z D}\right) = \theta_j - \ln\left[e^{\alpha_z} + \sum\nolimits_{k=1}^{J} e^{\theta_k - \lambda c_{zk}}\right] - \lambda c_{zj}, \tag{11}$$

which can be written as a two-way fixed-effects regression model:

$$y_{zj} = \theta_j - \kappa_z - \lambda c_{zj}, \tag{12}$$

where $y_{zj} = \ln\left[Y_{zj}/\left(N_z D\right)\right]$ and $\kappa_z = \ln\left[e^{\alpha_z} + \sum_{k=1}^{J} e^{\theta_k - \lambda c_{zk}}\right]$. Subtracting averages over each dimension z and j gives the following estimating equation:

$$y_{zj} - \bar{y}_j - \bar{y}_z = \bar{\kappa} - \bar{\theta} - \lambda\left(c_{zj} - \bar{c}_j - \bar{c}_z\right). \tag{13}$$

Therefore, if we regeress the centered (de-meaned) y on a constant and the centered c, the slope coefficient will be an estimate of $-\lambda$. This estimate appears to be biased low by up to 10% in our simulations, so we use bisection to refine our estimate of λ in subsequent steps of the algorithm.

Table I. Pseudocode for estimation algorithm.

1. Regress de-meaned natural logs of trips per capita from each origin z to each destination j on the de-meaned travel costs between each origin and destination to obtain a preliminary estimate of the travel cost coefficient, $\hat{\lambda}_0$.

2. Set $\hat{\alpha}_z = 0$ for all origins z, $\hat{\theta}_j = 0$ for all destinations j, and $\hat{\lambda} = \hat{\lambda}_0$. Compute $SSE = \sum_{z=1}^{Z} \sum_{j=1}^{J} \left(Y_{zj} - \hat{Y}_{zj}\right)^2$.

3. Set bisection upper and lower bounds for the travel cost coefficient, e.g., $\underline{\lambda} = 0.5\hat{\lambda}_0$ and $\overline{\lambda} = 1.5\hat{\lambda}_0$.

4. Holding $\hat{\alpha}_z$'s fixed, contract $\hat{\theta}_j$'s—i.e., $\hat{\theta}_j \leftarrow \hat{\theta}_j + \ln\left(Y_j/\hat{Y}_j(\hat{\alpha}, \hat{\theta}, \hat{\lambda})\right)$—and bisect $\hat{\lambda}$ until $\sum_{j=1}^{J} \left(Y_j - \hat{Y}_j(\hat{\alpha}, \hat{\theta}, \hat{\lambda})\right)^2$ is smaller than a pre-defined tolerance.

5. Holding $\hat{\theta}_j$'s and $\hat{\lambda}$ fixed, contract $\hat{\alpha}_z$'s—i.e., $\hat{\alpha}_z \leftarrow \hat{\alpha}_j + \ln\left(Y_z/\hat{Y}_z(\hat{\alpha}, \hat{\theta}, \hat{\lambda})\right)$—until $\sum_{z=1}^{J} \left(Y_z - \hat{Y}_z(\hat{\alpha}, \hat{\theta}, \hat{\lambda})\right)^2$ is smaller than a pre-defined tolerance.

6. Compute $SSE' = \sum_{z=1}^{Z} \sum_{j=1}^{J} \left(Y_{zj} - \hat{Y}_{zj}\right)^2$. If $(SSE - SSE')$ is less than a pre-defined tolerance, then stop; otherwise, set $SSE = SSE'$ and return to step 3.

In step 2, we set all origin and destination fixed effects to zero, and we adopt the linear estimate of the travel cost coefficient as the initial value of $\hat{\lambda}$. Then we compute the sum of squared errors between the observed and predicted total visits from each zone to each site. In step 3, we set the upper and lower limits of $\hat{\lambda}$ equal to 0.5 and 1.5 times the initial value for the bisection algorithm (a range that, according to our simulations, safely covers the SSE-minimizing value of $\hat{\lambda}$). In step 4, the travel cost parameter is bisected and contraction mapping is used to update the destination fixed effects, which produces an (arbitrarily close to) exact match between the observed and predicted total visits to each site conditional on the travel cost parameter and the origin fixed effects. In step 5, contraction mapping is used to update the origin fixed effects, which produces an (arbitrarily close to) exact match between the obeserved and predicted total trips from each origin conditional on the travel cost parameter and the destination fixed effects. In step 6, we compute an updated sum of squared errors, SSE'. If the SSE is improved by less than a pre-defined tolerance, the algorithm stops; otherwise, we return to step 3 and repeat the process.

2.3 Model simulation

To test the computational feasibility and to gauge the potential statistical performance of the proposed estimation approach, we conducted a Monte Carlo simulation experiment. In each Monte Carlo iteration, we simulated data consistent with the assumed data generating process, applied the estimator to the simulated data, and recorded the resulting parameter estimates. We repeated this process 100 times and used the frequency distributions of the resulting parameter estimates to characterize the accuracy and precision of the estimator.

The Monte Carlo simulation experiment was specified as follows. First, we simulated a rectan-

gular study area of the approximate size of one half of California, with width $W = 250$ miles and height $H = 375$ miles.[5] Then we randomly distributed $J = 100$ recreation sites (trip destinations) and $Z = 1,000$ zones (trip origins) throughout the study area. We calculated travel costs by adding a fixed cost of \$50 to the 2-way (straight-line) distance cost between each origin-destination pair, which was computed using \$1 per mile of travel distance. Next, we randomly assigned population sizes to each zone, where $N_z \sim U[\bar{N}/10, \bar{N}]$ and $\sum_{z=1}^{Z} N_z = 20$ million people, about half the population of California. Next we randomly assigned origin and destination fixed effects to represent heterogeneity across space in people's latent demand for outdoor recreation and in the amenity and environmental quality levels across the sites.[6]

The experimental design parameters are shown in Table II. These values were chosen to provide a reasonably close match to the size of our study area and the dimensions of the estimation problem we will face in the application to our study area. The preference parameters in particular were chosen to produce distributions of simulated numbers of trips from the set of origins to the set of destinations that are within realistic ranges.

Table II. Parameters used for simulation experiments.

Parameters	Description	Values
w	Width of study area [miles]	250
h	Height of study area [miles]	375
J	Number of destinations	100
Z	Number or origins	1,000
N	Population in study area	20×10^6
λ	Travel cost parameter	0.070
θ_j	Destination fixed effects	$U\big[0, 4\big]$
α_z	Origin fixed effects	$N\big(0, 0.3\big)$

The main objective of the simulation exercise was to test the performance of the algorithm in recovering the preference heterogeneity across origins that we built into the data generating process. Can we consistently and efficiently estimate 100 destination fixed effects and 1,000 origin fixed effects? If so, then we will have a tool that can make full use of mobility data on origin-to-destination visitation frequencies. This simulation exercise serves as a proof-of-concept for our application to

5. At the time we conducted the simulation experiments, we expected to choose a study area that included only the southern half of California. We subsequently selected lakes throughout the entire state for our case study application, but we saw no good reason to modify and re-do our simulation experiments on this basis, so we left well enough alone.

6. Note that we do not explicitly account for heterogeneity *within* zones; here we focus exclusively on across-zone preference heterogeneity. This could emerge from sorting with respect to proximity to desirable outdoor recreation areas, or on other dimensions that happen to be correlated with outdoor recreation avidity. The very logic of spatial sorting suggests that across-zone heterogeneity will be greater than within-zone heterogeneity, so we believe that this is a defensible simplifying assumption.

California lakes, and for analogous applications involving many origins and destinations when full sets of fixed effects are to be estimated for each.

2.4 Valuing early warnings

The recreation demand model described above provides an important ingredient for calculating the value of information, but extra steps are required to compute the VOI from an early warning system for HABs. This section describes how we use the recreation demand model, plus extra assumptions, to calculate the value of perfect information.

The extra assumptions we make are that recreators behave rationally under uncertainty as described by Bayesian expected utility theory, and that recreators form their expectations about the likelihood of encountering a HAB at each lake based on the lake's recent historic frequency of HABs. If this description is sufficiently accurate, then we can use the parameterized recreation demand model to compute a recreator's willingness-to-pay for information about the presence or absence of a HAB at each candidate site.[7]

The conventional log-sum formula from the multinomial logit model is a measure of expected maximum utility, and dividing by λ (the travel cost coefficient and the marginal utility of income) converts the measure to dollars. Under the status quo, we assume that recreators use the long-run average frequency of HABs at each lake, H_j, to form their expectations about the occurrence of a HAB on any given choice occasion. Therefore, we can write the expected value (of access to all lakes in the dataset) with existing information, which here we denote as *EVEI*, as follows:

$$EVEI = \sum_{z=1}^{Z} \frac{1}{\lambda} \left\{ \ln \left[e^{\alpha_z} + \sum_{j=1}^{J} e^{\delta_j - \beta H_j - \lambda c_{zj}} \right] - \alpha_z \right\}. \tag{14}$$

Note that *EVEI* is the *maximized expected recreation value*. Under the assumptions of the model, on each choice occasion recreators choose the site with the largest realization of $\delta_j - \beta H_j - \lambda c_{zj} + \varepsilon_{zj}$. Recall that under the random utility maximization (RUM) framework we assume that recreators can observe ε_{zj} whereas the researcher cannot. Here we further assume that recreators cannot observe whether or not a HAB is present at the site, so they make their decision based on their expectation of the HAB occurrence, which corresponds to the long-run average frequency at the site, H_j. This is the logic for including H_j as an additional quality attribute of each site. H_j enters linearly due to our assumption of expected utility maximization (expected utility is linear in probabilities).

The expected recreation value if the recreators can obtain perfect information about the presence or absence of a HAB at all sites before they choose which site to visit—the expected value with perfect information, which here we denote as *EVPI*—can be computed using the same information used to compute *EVEI*. Note that *EVPI* is the *expected maximized recreation value*, where the expectation is taken over the possible states of the world, and where in this setting states of the world correspond to the possible combinations of lakes where a HAB is present. Formally, the

7. If other hypotheses about how recreators form expectations and make choices under uncertainty are thought to be more realistic—e.g., prospect theory or other theories of decision-making from behavioral economics—then these could be used in place of the standard Bayesian expected utility maximization assumption used here. This is one of several extensions are interested to pursue in follow-up work.

expected value with perfect information is:

$$EVPI = \sum_{s=1}^{S} P_s \frac{1}{\lambda} \left\{ \ln \left[e^{\alpha_z} + \sum_{j \in s} e^{\delta_j - \beta - \lambda c_{zj}} + \sum_{j \notin s} e^{\delta_j - \lambda c_{zj}} \right] - \alpha_z \right\}, \tag{15}$$

where $s = 1, 2, ..., S$ indexes all possible subsets of the J sites, and P_s is the probability that a HAB will occur at all sites in s and at no other sites on any given choice occasion. We calculate P_s from the H_j's by assuming independence among the site-specific probabilities of a HAB on each choice occasion, i.e.,

$$P_s = \prod_{j \in s} H_j \times \prod_{j \notin s} (1 - H_j), \tag{16}$$

which is the probability that all sites in set s have a HAB and all remaining sites do not.

Note the contrast between *EVEI* in equation (14) and *EVPI* in equation (15). The probabilities of a HAB occurrence at each lake, H_j, appear in both formulas but in different ways. They appear in the expected indirect utility functions in the formula for *EVEI*, and they appear as components of the probabilities of each possible state of the world, P_s, in the formula for *EVPI*. If a recreator will know which sites have a HAB and which do not, then the H_j's are no longer relevant to the recreator at the time of their choice. Site j will have indirect utility $\delta_j - \lambda c_{zj} + \varepsilon_{zj}$ if a HAB is not present, or $\delta_j - \beta - \lambda c_{zj} + \varepsilon_{zj}$ if a HAB is present. So given any possible pattern of HAB occurrences among the sites, the willingness-to-pay for access to the sites can be calculated using the standard log-sum formula with the appropriate indirect utilities, including β or not, at each site. This is the quantity inside the braces in equation (15). If there are $M(\leq J)$ sites with non-zero probabilities of HAB occurrences ($H_j > 0$), then there are 2^M possible patterns of HAB occurrences, including no HABs at any site, a HAB at all M sites, and in between these extremes many permutations with $1, 2, ..., M-1$ sites with HABs and the rest without.

The computational task is to numerically integrate the conditional (state-dependent) value of access to the sites over all possible states of the world. That is, we compute the probability of each state of the world, multiply it by the willingness-to-pay for access under each state of the world, and sum the products to get the expected value with perfect information, *EVPI*. The curse of dimensionality sets in quickly, so to keep the task manageable for this study we focused on the 10 lakes with the highest risk of HABs and we ignored the risk of HABs at all remaining lakes. Nevertheless, it is important to account for all lakes when estimating the value of information because the lakes with no risk of HABs can serve as substitutes for any lakes with a HAB on any given choice occasion.

2.5 Visitation data

We chose California as the site of our case study because it has a wide range of climate conditions, many water bodies used for outdoor recreation with varying susceptibilities to HABs, and a prototype early warning system for HABs currently under development, "My Water Quality Portal," from which we were able to obtain satellite data on the historic frequencies of algal blooms at each lake in our dataset [50, 51]. We calibrated the influence of HABs on the conditional indirect utilities at each site—β in equation (3)–(6) above—using summary results from several previously published recreation demand studies related to HABs, described in more detail below.

The main data we used to estimate our recreation demand model, which in turn forms the basis of our benefit transfer function to be used for calculating the value of satellite information for a HAB early warning system, was obtained from StreetLight Data, Inc. StreetLight is a private information technology company that uses smartphones as sensors to measure vehicle and pedestrian traffic patterns and trends across North America. StreetLight collects and processes billions of anonymized location records from electronic devices with GPS capabilities or apps that collect location data on an ongoing basis. StreetLight then processes, extrapolates, and validates these raw data using thousands of traffic counters and embedded sensors, to scale the data samples up to population estimates of traffic flows.[8] These data can then be used to estimate the number of people who visit any combination of points or areas of interest that can be spatially delineated in a GIS data layer and intersected with the location coordinates of the individual electronic devices that provide the raw location data records.

At least two previous outdoor recreation studies have used mobility data from personal electronic devices [52, 53]. Though potentially powerful, such data have several important limitations. They are based on a sample rather than a census of users, and it is not yet clear whether important forms of selection bias exist in the data that differentiate the average behavior of individuals who are counted versus those who are not. Also, the data indicate people's locations only—specifically, the location of their electronic devices—not their activities at the destination sites. Merrill et al. [52] provide a detailed discussion of these and other limitations of cell phone location data for modeling outdoor recreation demand.

To extract the data required for estimation from the StreetLight Data portal, we first created a GIS data layer containing polygons representing the location and spatial extent of all lakes to be included in the model. This task involved drawing polygons that encompass the complete shoreline and any adjacent infrastructure apparently used for lake recreation, such as parking lots, visitor centers, entry kiosks, campgrounds, boat ramps, publicly accessible beaches, etc. We began by obtaining a baseline data layer from the National Hydrography Dataset.[9]. Then, by visual inspection, we manually adjusted the polygon outlines to encompass the complete shoreline and adjacent infrastructure for each lake in turn. StreetLight Data then provided counts of visitors from each ZIP code in California to each lake polygon for each month April–September in 2019. This allowed us to estimate monthly recreation demand models and thereby account for seasonal visitation trends.

To estimate the value of a HAB early warning system, we must combine the mobility data described above with information on the frequency of HABs at the lakes. To meet this requirement, we obtained data from the San Francisco Estuary Institute (SFEI) on Cyanobacteria indicator levels in 201 of the largest lakes in California between 2002 and 2020. SFEI has developed an online Harmful Algal Bloom Analysis Tool[10], which currently uses a combination of satellite data and in situ monitoring at a sample of sites across the state to forecast HABs.[11] We used data collected

8. `https://www.streetlightdata.com/our-data/`
9. `https://www.usgs.gov/core-science-systems/ngp/national-hydrography`
10. `https://fhab.sfei.org/`
11. Prediction could be improved by expanding the in situ monitoring efforts and by using higher resolution remote sensing data from the Sentinel-3 OLCI instrument (300 m pixels at full resolution, return interval of ∼3 days, and level 2 data processing with calibrated estimates of chlorophyll-a concentration [mg (chl a) m^{-3}]) [54], and by combining MODIS and Landsat data [55]. These possibilities provide opportunities to estimate the value of incremental improvements in the accuracy and precision of the early warning system in follow-up work. To do so, it

by SFEI to compute the average probability of encountering a HAB at each lake if visitors have no early warnings of lake conditions from satellite data or other sources. These measures were used to construct the principal HAB attribute in our status quo site choice model assuming no early warning system is in place. We also assumed that before visiting a lake people cannot otherwise make a better prediction of a HAB occurrence than is implied by the daily frequency of HABs in the historic data.[12]

2.6 Literature review for benefit transfer

The recreation demand estimation approach described above will identify a full set of fixed effects for all lakes. These fixed effects are lumped parameters that represent the contribution to utility of all natural and constructed features at each recreation site, including parking and bathroom facilities, docks and piers, average weather conditions, the abundance of fish in the lake, average water quality conditions, and so on, including the prevalence of harmful algal blooms. To use the model for computing the value of information about HABs, we must decompose the destination fixed effects into a component related to algae and a residual that retains all other features of the sites, i.e., $\theta_j = \delta_j - \beta H_j$, where H_j is our measure of the prevalence of algae at lake j. We specified β using results from previous studies on outdoor recreators' aversion to the presence of algal blooms at the recreation sites they might visit. We identified eight published recreation demand studies whose results may provide suitable summary statistics to aid in the calibration of our benefit transfer function. In the following paragraphs we describe the relevant findings from each study, in alphabetical order by the lead author's last name.

Alvarez et al. [56] estimated a recreation demand site-choice model for 35 boat access points in a single county in southwest Florida, and they used the model to compute the potential welfare effects of boat ramp closures due to harmful algal blooms in the study area. Summarizing their quantitative results, the authors note "...the average individual who would have taken a boating trip to Lee County on a day when cyanobacterial HABs were occurring in 2018 is estimated to have lost $17.3 as a direct result of the algae blooms." Inflating from $2019 to $2021 gives $18.3.

Crase and Gillespie [57] examined recreational use values for Lake Hume in Australia, with a focus on the effect of water levels and algae contamination. The authors estimated a zonal travel cost model, which involves regressing visitation rates estimated for concentric zones surrounding the lake against the estimated travel cost for each zone. Demand curves estimated in this way were constructed using data on actual trips and hypothetical trips based on a sample of survey respondents' answers to a contingent behavior question that asked how many trips would be taken if an algae alert were in effect at the lake. The resulting estimates of consumer surplus per visit were $33 in the baseline scenario and $20 in the algae alert scenario. Inflating from 2008$ to 2021$ gives corresponding values of $41.6 and $25.2.

Egan et al. [58] conducted a large-scale study of water-based outdoor recreation activities in Iowa, and found that water clarity and nutrient levels were the most important water quality

will be necessary to estimate the rates of type I errors (false positives, predicting a HAB when none is present) and type II errors (false negatives, failing to predict a HAB when one is present) of the current early warning system, and the changes in these error rates if the satellite data and predictive models using those data are improved.

12. In reality, somewhat better predictions may be possible for several reasons. In particular, HABs are correlated over time: if a HAB is present today, it probably will be present tomorrow, so my experiences or those of other people I know on recent visits will be informative of conditions at those sites in the near future.

attributes for recreators who visit Iowa lakes. Combining results from their Tables 9 and 11, across 5 candidate models, welfare estimates for improving the quality of all lakes up to that of the best lake, were between $10 and $25 per visitor per trip. Egan et al. did not examine algal blooms specifically, but we conjecture that the value of eliminating a high risk of a HAB at a recreation site will be of the same order of magnitude as the large improvements examined in the Iowa lakes study. Inflating from 2009$ to 2021$ gives a range of $12.7–$31.8.

Piper et al. [59] used the travel cost method and a survey-based contingent behavior method to estimate recreational benefits from reduced algae growth at two popular lakes in South Dakota. Respondents were asked how many trips they took to the lakes in a recent time period and how many trips they would take if algae did not form thick mats on the lakes during their visits. The average number of visits to the lakes by those respondents who were interviewed in person and who took at least one trip was 9.6 trips per year. The same respondents indicated that they would take on average 14.1 trips if the lakes were cleaner. The associated consumer surplus for the average respondent was estimated to be $106. Dividing this by (9.6+14.1)/2 trips and inflating from 1987$ to 2021$ gives $21.4 per trip.

Roberts et al. [60] conducted a stated preference survey of environmental conditions at Tenkiller Ferry Reservoir, near Tulsa, Oklahoma, which is a popular recreation site for boating, fishing, and swimming. In the first of two survey treatments, a subset of respondents were asked which of two sets of lake conditions they preferred, where the conditions included algal bloom status (present or absent), water level (normal, or 2 or more feet below normal), and the level of the user fee ($0, $2, $4, $6, or $8). In the second survey treatment, a different subset of respondents were presented with contrasting lake conditions described probabilistically, e.g., "10% chance of algae bloom and 90% chance of no algae," "50% chance of normal water level and 50% chance of 4 feet below normal," etc. The authors used these contrasting treatments to investigate the influence of presenting environmental attributes as uncertain outcomes, for comparison to the standard approach of presenting all outcomes as deterministic. The authors estimated three models: a traditional model using deterministic attributes, an expected utility model using linear probability weights, and a prospect theory model using non-linearly transformed probability weights. All three models indicated that respondents viewed algal blooms as a dis-amenity. The expected utility model, which is linear in the probability of an algal bloom, indicated that on average recreators would be willing to pay an addition user fee in the amount of $13.02 to eliminate the risk of an algal bloom when visiting the lake. Inflating from 2008$ to 2021$ gives $16.4. Unlike the expected utility model, the prospect theory model implies a marginal willingness to pay function that varies with the probability of a bloom. Specifically, respondents were seen to under-weight low probabilities and over-weight high probabilities, such that the transformed probability weighting function looks like an 'S' rather than a straight line. Results from the prospect theory model imply that people are not willing to pay much to reduce the risk from around 30% to 0% or from 100% to 80%, but they are willing to pay a substantial amount to reduce the risk from 80% to 30%.[13]

Van Houtven et al. [61] conducted a stated preference survey of eutrophication conditions in North Carolina lakes. The survey instrument used a repeated dichotomous choice format. All respondents who had taken a trip to a lake in the last 12 months, or who expected to do so in

13. These findings raise interesting questions about the value of information in an expected utility framework versus a prospect theory framework, and the contrasting results reported by Roberts et al. provide an opportunity to compute VOI statistics using both frameworks and compare the results. We flag this as another task for follow-up work.

the coming 12 months, were presented with descriptions of two hypothetical lakes with differing environmental quality conditions and distances from their home. The quality attributes were: color, clarity, fish abundance, frequency of algal blooms, and odor. However, the experimental design involved perfectly correlated variations in each environmental attribute, so it was not possible to estimate the marginal effect of any single attribute alone. This means that we cannot use the estimated coefficients to help calibrate our benefit transfer function. Nevertheless, the results of this study are consistent with the hypothesis that recreators are averse to algal blooms, among other degraded environmental conditions.

Wolf et al. [4] used a revealed preference approach to examine the impacts of HABs on recreation behaviors in Lake Erie. Site visitation data were collected using a web based survey, and the authors constructed a "site-specific, summer-long mean algae measure" using satellite data provided by NOAA. Using a latent class modeling approach, the authors found heterogeneous preferences among different groups of users: in particular, beachgoers were more averse to *E. coli* and anglers were more averse to algae, and other users were more-or-less indifferent to algae. Across all respondents, the authors estimated an average marginal willingness to pay of −$0.25 per 10,000 cells/mL. "Considering the mean algae level at the recreation sites in our sample is 127,900 cells/mL,[14] this reflects a cost of $3.20 per visitor per trip associated with the presence of the mean level of algae at a recreation site relative to a site with no algae." Inflating from 2019$ to 2021$ gives $3.39 per visitor per trip.

Zhang and Sohngen [37] used a stated preference survey to study anglers' preferences for avoiding algal blooms in Lake Erie. The commodity description provided in the survey instrument was "distance of an algal bloom that the angler would need to boat through (from the shoreline) to get to the fishing site," and was presented as either 0, 4, or 8 miles. The authors found that "Ohio anglers are on average willing to pay $8 to $10 for one less mile of an algal bloom to boat through" per trip. Due to the unique commodity description used in this study, it is not obvious how these results can be transferred to our model, which uses the probability of a HAB at the lake as the quality attribute. For a very rough estimate, if the average size of an algal bloom is on the order of one extra mile of boat travel required to reach a desirable fishing location, then the average willingness to pay to eliminate the bloom should be within the $8–$10 range reported by the authors. Inflating from 2018$ to 2021$ gives a range of $8.64–$10.8.

3 Results

In the subsections that follow, we present the results of our model simulation experiments, our review of the literature, and our application to California lakes including recreation demand model estimation results using mobility data from StreetLight and value of information estimates using historic data on HAB frequencies from SFEI.

14. WHO guidelines indicate a threshold of 100,000 cells/mL for a "bloom" (WHO 2003), so according to this statistic the average state of Lake Erie is bloomed. However, the corresponding standard deviation was 217,000 cells/mL and the maximum observed value was 821,400 cells/mL, so while blooms are common they are also highly variable in Lake Erie.

3.1 Simulation experiments

Summary statistics based on the simulated data are shown in Table III. We view these as realistic figures for at least some outdoor recreation demand settings, but as we show below the trip frequencies simulated in our numerical experiments are substantially larger than the frequencies of trips to the subset of California lakes that we analyze in our case study. Nevertheless, we believe that the basic lessons from our simulation experiments should be sufficiently general to provide a reliable indication of the statistical performance of our estimation approach when applied to our case study data.

Table III. Summary statistics for simulated data. Average trips per person, and fraction of people who visited 0, 1, 2, or 3 or more sites.

Statistic	Average	St. dev.
Average trips per person by zone	3.14	3.52
$F(0)$ by zone	0.28	0.31
$F(1)$ by zone	0.34	0.22
$F(2)$ by zone	0.22	0.18
$F(3+)$ by zone	0.17	0.25

Table IV shows summary results from the Monte Carlo simulation experiments, based on 100 repetitions of the data generation and estimation process described in section 2.3. The estimated biases are $\ll$0.1% of the magnitude of the travel cost parameter, $\approx$1.0% of the average magnitude of the destination fixed effects, and $\approx$14% of the average magnitude of the origin fixed effects. The first two of these are small enough to suggest that the estimator is consistent for these parameters, while the third suggests there is a measurable bias in the estimates of the origin fixed effects. Overall, these results suggest the estimator is very precise and highly, but not perfectly, accurate.

Table IV. Estimated accuracy (Bias) and precision (St. dev.) of the proposed recreation demand estimation approach, based on 100 Monte Carlo repetitions.

Estimate(s)	True	Bias	St. dev.
Travel cost coefficient, $\hat{\lambda}$	0.07	-4.21e-6	2.32e-5
Destination fixed effects, $\hat{\theta}_j$	$\sim U(0,4)$	0.0326	0.0212
Origin fixed effects, $\hat{\alpha}_j$	$\sim N(0,0.3^2)$	0.0333	0.0184

A visual indication of the performance of the estimator is provided in Figure 3. The top left plot shows the association between the true fixed effects (α_z) and the averages of the estimated fixed effects ($\hat{\alpha}_z$) across all 100 Monte Carlo iterations for the 1,000 origins. The top right plot shows

analogous association between the true (θ_j) and estimated fixed effects ($\hat{\theta}_j$) for the 100 destinations. The bottom plot shows the association between the true (Y_{zj}) and predicted number of visits ($\hat{Y}_{zj}$) from each origin to each destination (of which only 1,000 out of the total 100,000 pairs are shown for readability) for the final Monte Carlo iteration. Looking closely at the α_z scatter plot, it is possible to see that the mass of points is shifted slightly above the 1:1 dashed line, where the intercept ($a = 0.0333$) corresponds to the magnitude of the estimated bias reported in Table IV.

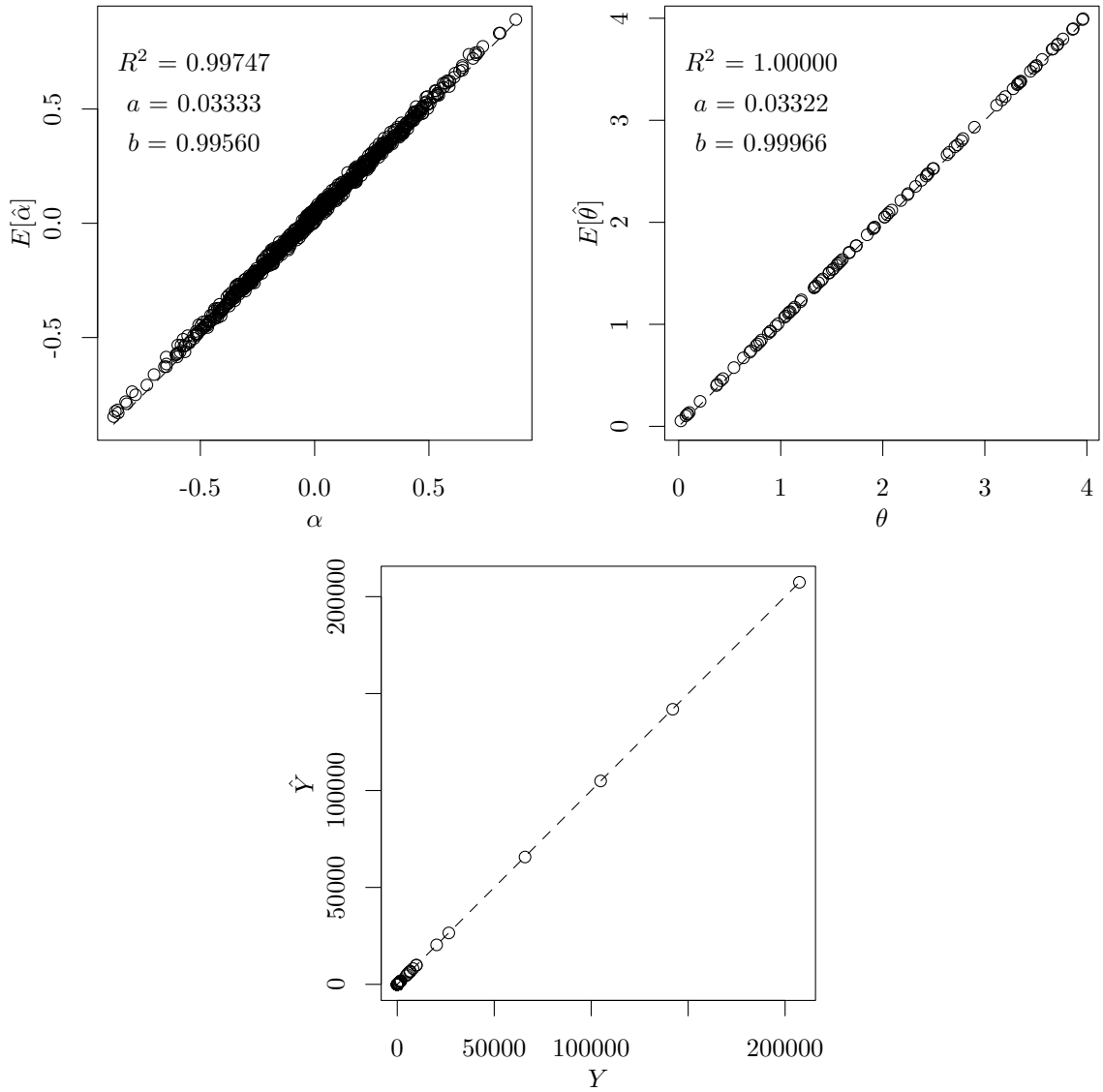

Figure 3. Scatter plots of the average estimates of the origin fixed effects across all Monte Carlo iterations vs. the true origin fixed effects (top left), estimated vs. true destination fixed effects (top right), and predicted vs. true trips from each origin to each destination in the final Monte Carlo iteration (bottom).

The strong performance of the estimator is a consequence of the high information content in the simulated data. These simulations assume that we have data on the total number of trips from each origin to each destination—importantly, encompassing all trips, not just trips made by a subset of visitors as would be the case in a traditional recreation demand study using survey data from a random sample of residents or an on-site survey of visitors. The simulations also assume that we know the correct model structure—or, more pragmatically, that the multinomial logit model can provide a good fit to the data—and that all variables are measured without appreciable error. These are standard assumptions for a data simulation and estimation exercise like this one. What we have shown here is: If we have a good model and can collect good data, our proposed estimation approach can produce very precise and highly (though not perfectly) accurate estimates of the model parameters.

3.2 Benefit transfer assumptions

Summary estimates from the seven studies reviewed in section 2.6 are shown in Table V. This compilation suggests that recreators' preferences related to the presence of algae in surface water bodies are heterogeneous. In the current study we use a simple model of recreation site choices that stipulates a uniform value for eliminating a HAB at a visited lake, as if based on a representative visitor. It is not immediately clear under what conditions and how severely this simplification might bias aggregate welfare measures or estimates of the value of information, so we highlight this issue for further study in follow-up work. It may be possible to investigate the potential importance of preference heterogeneity through counterfactual simulations using homogeneous versus heterogeneous preference functions. In the meantime, we chose a benchmark central value for β, the utility loss from the presence of a HAB at a visited site, near the center of the range of estimated values extracted from the seven studies summarized in Table V. The estimates reported in the table differ along multiple dimensions—the specific definition of the ecological endpoint, the exact welfare measure used, context including number and quality of substitute sites, and others—all of which will influence the suitability of the reported estimates for benefit transfer to our study area. Nevertheless, we believe it is reasonable to infer that the average willingness-to-pay per visitor per trip will be in a similar range as the assembled estimates. Therefore, as a benchmark value we assumed that the average willingness-to-pay to eliminate a HAB at a visited lake on a trip is $wtp =$ \$15. We also re-estimated the model using a low value of \$5 and a high value of \$25, to determine whether our results scale with β linearly or otherwise. In our linear indirect utility function, the willingness to pay to eliminate a HAB at a visited site is $wtp = \beta/\lambda$, so we can calibrate the algae coefficient using $\beta = \hat{\lambda} \times wtp$.

3.3 Application to California lakes

We obtained data on 100 of the 201 California lakes in the SFEI online HAB analysis tool. To select lakes for inclusion in our case study, we sorted the lakes by their surface area and historic risk of HABs. We chose the 100 lakes with the highest sum of standardized scores on these two measures for analysis. We reasoned that these lakes are likely to be among the most important

Table V. Summary estimates of willingess-to-pay [2021$] to avoid algal blooms at surface water recreation sites extracted from seven outdoor recreation demand studies.

Study	Estimates	Description
Alvarez et al. (2019)	$18.3	Surplus loss for the average individual who would have taken a boating trip to Lee County on a day when a cyanobacterial HAB was occurring.
Crase & Gillespie (2008)	$16.4	Difference between estimated average consumer surplus per visit with versus without an algal alert in effect.
Egan et al. (2009)	$12.7–$31.8	Range of welfare estimates across 5 candidate models for improving the quality of all lakes in Iowa up to the quality of the best lake.
Piper et al. (1987)	$21.4	Estimated annual increase in consumer surplus per capita per visit to current users of Oakwood Lakes/Lake Poinsett from eliminating algal growth in the lakes.
Roberts et al. (2008)	$16.4	Average willingness-to-pay by recreators in the form of an additional user fee to eliminate [the risk of] an algal bloom when visiting the study lake.
Van Houtven et al. (2014)	NA	Not possible to estimate the marginal effect of algae due to a perfectly correlated experimental design.
Wolf et al. (2019)	$3.39	Cost per visitor per trip associated with the presence of the mean level of algae (127,000 cells/mL) relative to a site with no algae.
Zhang & Sohngen (2018)	$8.64–$10.8	Average willingness-to-pay by Ohio anglers for one less mile of an algal bloom to boat through per trip.

for outdoor recreation activities in the state where early warnings would be most valuable.[15] Figure 4 shows the total number of trips to all 100 California lakes in our case study during the time period of April–September 2019, as indicated by the electronic device mobility data obtained from StreetLight Data, Inc. This plot shows an expected pattern, with total visits lowest in the shoulder months of April and September and peaking in the middle of the summer in the month of July, during which the number of trips to all lakes is between two to three times greater than in April or September.

The ten most visited lakes are listed in decreasing order in Table VI, with visitation totals for each. The top ten most visited lakes account for nearly two thirds of the visits to all 100 lakes. The total number of visits to the set of 100 lakes was 17.3 million, with the most visited lake among them, Lake Tahoe, attracting 1.95 million visitors between April and September of 2019. The frequency distribution of total visits among the 100 lakes is shown in Figure 5. This plot also

15. Ideally we would have selected the lakes with a combination of the largest visitation rates and HAB risks, but visitation rates could not be estimated ex ante without the StreetLight data.

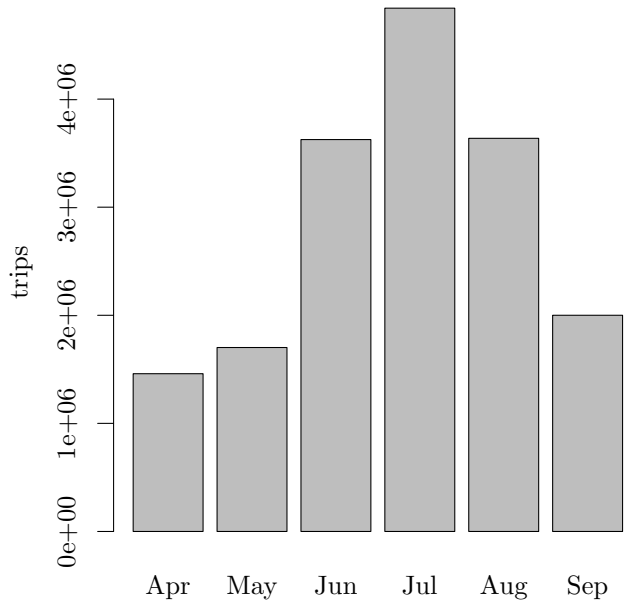

Figure 4. Total number of trips taken from ZIP codes in California to 100 lakes in California in each month April–September 2019.

shows a common pattern: many lakes had relatively few visitors (more than half below 100,000) and a few lakes had many visitors (500,000 or more).

Summary statistics related to the frequency of HABs in California lakes computed using the SFEI satellite data are presented in Figures 6 – 8. Figure 6 shows the frequency distribution of lakes where the fraction of summer days (inclusive of April through September) when the 'Modified Cyano Index' (hereafter C.I.) was greater than 100 [62]. This level of the C.I. is near the middle of the transition from 'low' (green) to 'high' (red) levels on the SFEI online HAB analysis tool color scale, and is believed to be near the center of the range within which most people clearly recognize the presence of excessive green algae in the water column. Figure 6 shows that at a majority of lakes, less than 1 percent of summer days have C.I. levels above 100, but there is a fairly long thin tail of lakes with up to 35 percent of days when C.I. levels exceed this threshold. The most severely impacted lakes are listed in decreasing order of severity in Figure 7, and the same lakes are shown on a map in Figure 8.

For the purposes of estimating expenditures on visits to the lakes in our case study and for estimation of our recreation demand model, we computed travel costs between each origin (ZIP code) and destination (lake) as follows:

$$c_{zj} = \underbrace{P\$}_{\substack{\text{fixed cost}\\ \text{per trip}}} + 2\left(\underbrace{p\frac{\$}{\text{mi}}}_{\substack{\text{variable cost}\\ \text{of travel}}} + \underbrace{\frac{\frac{1}{2}\bar{Y}\frac{\$}{\text{yr}}}{50\frac{\text{wk}}{\text{yr}} \times 40\frac{\text{hr}}{\text{wk}} \times 45\frac{\text{mi}}{\text{hr}}}}_{\text{opportunity cost of time}} \right) \times d_{zj} \text{ mi.} \tag{17}$$

We multiply the second term inside the parentheses by $\frac{1}{2}$ because we assume that the opportunity

Table VI. Top ten most visited lakes among the 100 California lakes in our dataset during the period April–September 2019. Based on electronic device location data from StreetLight, Inc.

Lake	Visits
Lake Tahoe	1,950,000
Lake Shasta	1,540,000
Isabella Lake	1,090,000
Folsom Lake	1,090,000
Lake Berryessa	1,050,000
Lake Almanor	745,000
Big Bear Lake	697,000
Clear Lake	602,000
Lake Oroville	574,000
New Melones Lake	571,000
All other lakes	7,340,000
Total	17,300,000

cost of time is 1/2 of the average hourly wage, $\bar{Y} \div 50 \div 40$, where $\bar{Y}$ is average annual income. We use $P = \$50$ per trip and $p = \$0.25$ per mile [63], and using data from StreetLight we compute $\bar{Y} = \$77{,}506$ per year, so the term in parentheses computes to 1.36 \$/mi. For example, if $c_{zj} = \$100$ then the distance traveled is (100-50)/1.36/2 = 18.4 miles.

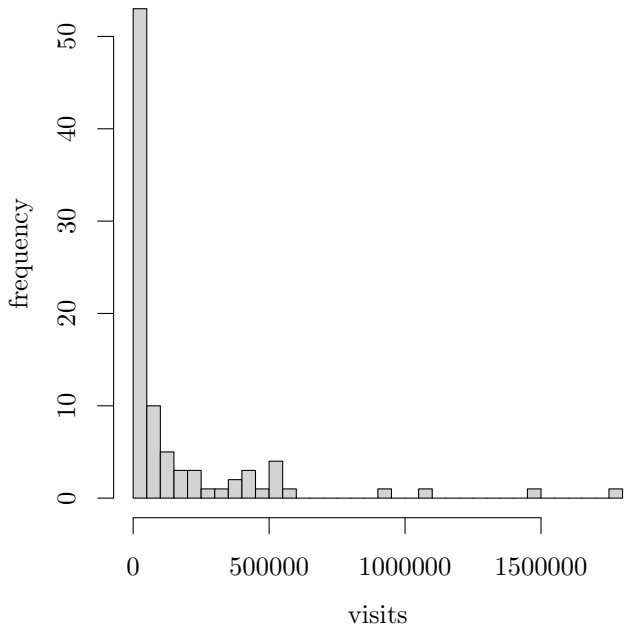

Figure 5. Frequency distribution of total visits to 100 California lakes included in our case study.

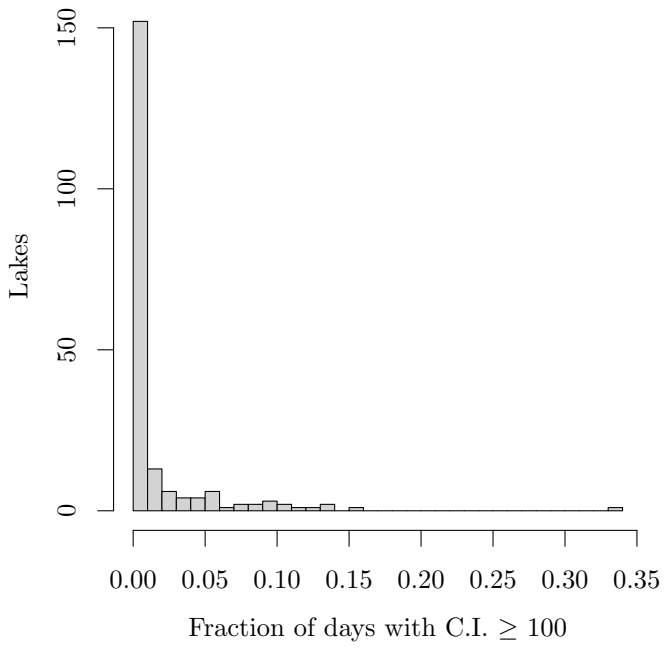

Figure 6. Frequency distribution of lakes with fraction of summer days (April–September) between 2002–2020 when the SFEI 'Modified Cyano Index' (C.I.) > 100.

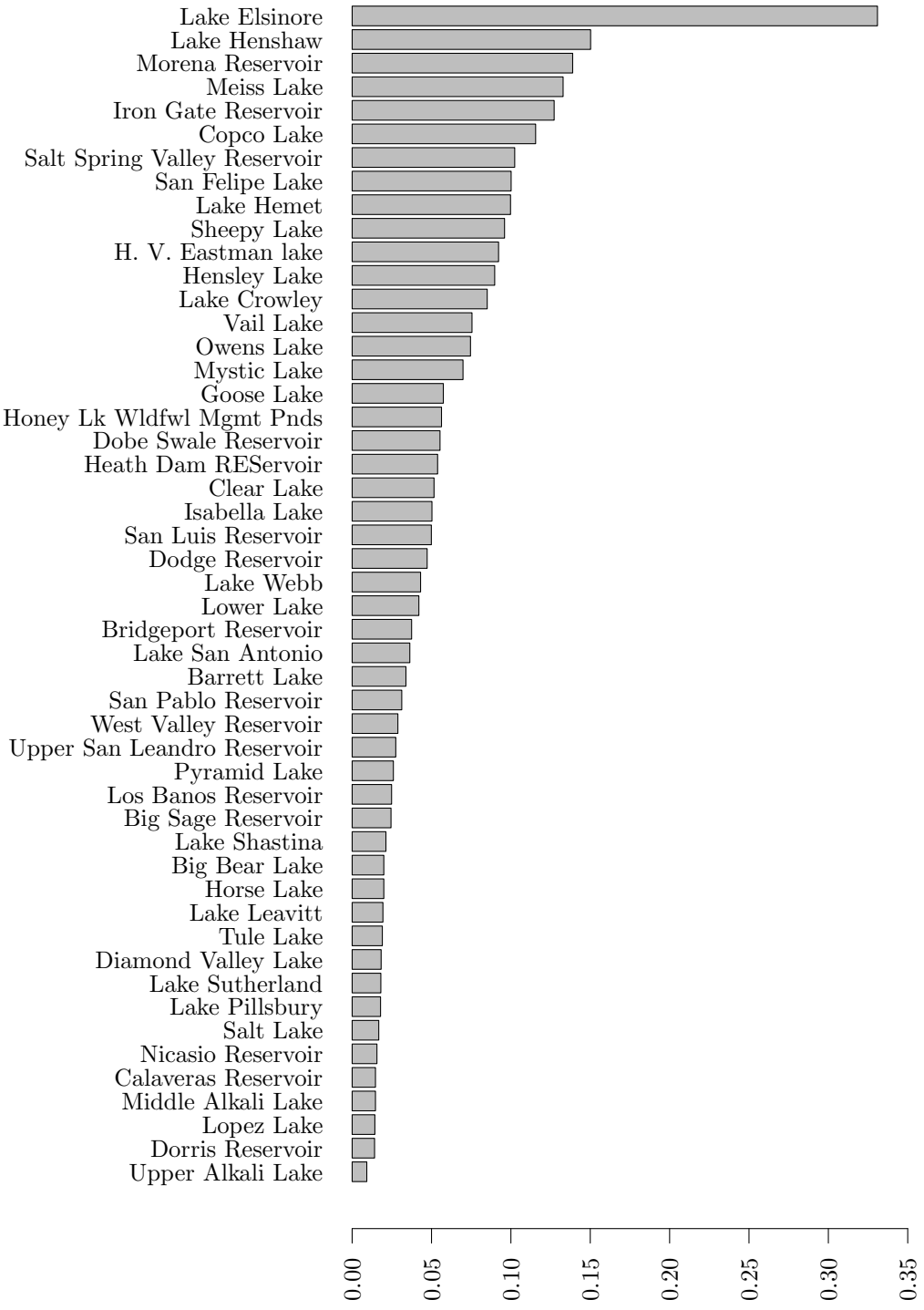

Figure 7. Frequency of summer days (April–September) between 2002–2020 when the SFEI 'Modified Cyano Index' (C.I.) > 100, for the top 50 (most severely impacted) lakes.

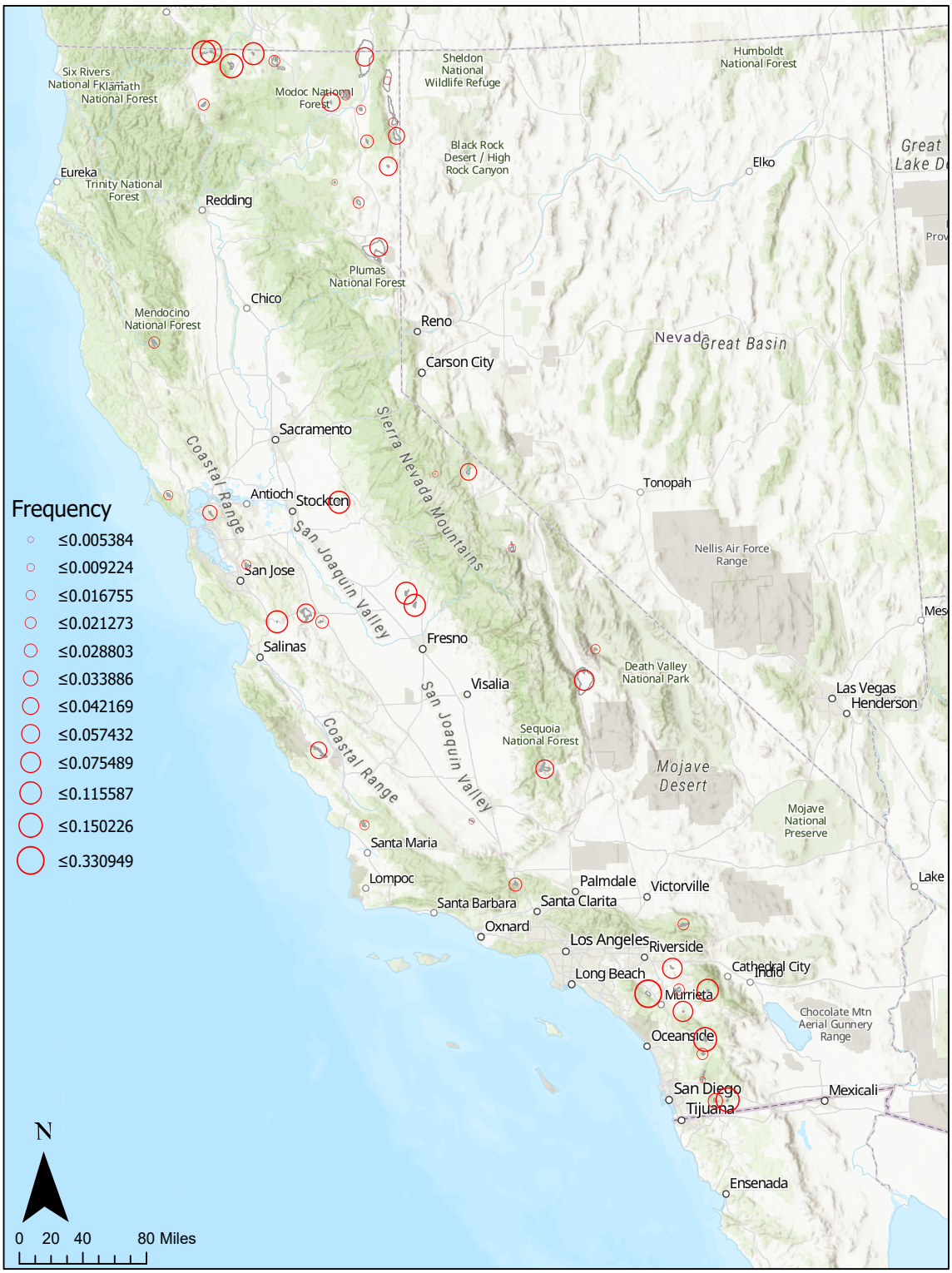

Figure 8. Locations of 50 large lakes in California with the highest frequency of summer days (April–September) between 2002–2020 when the SFEI 'Modified Cyano Index' (C.I.) > 100.

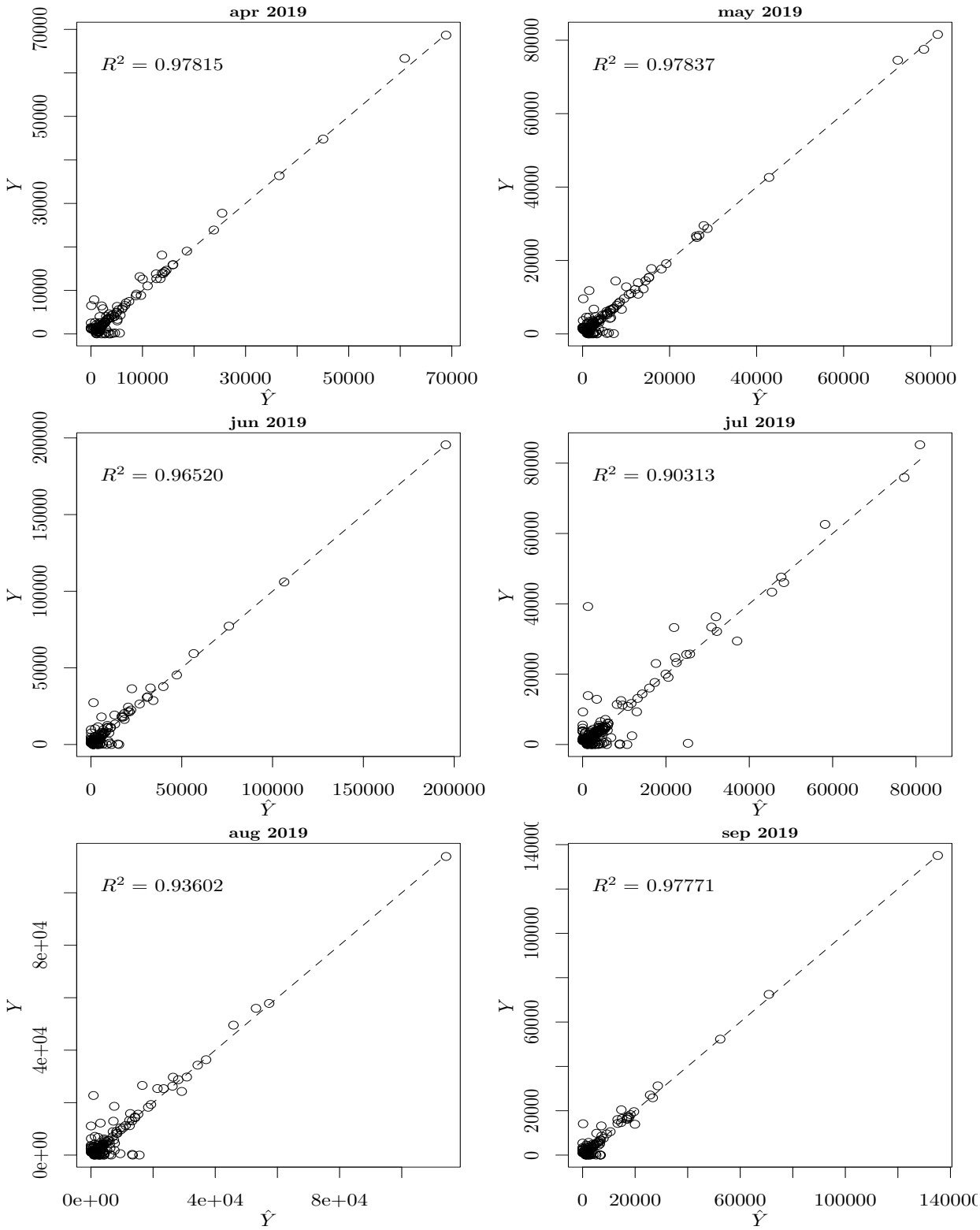

Figure 9. Scatterplots of observed vs. predicted trips from each origin to each destination during the months April–September of 2019.

Estimation results for the recreation demand model are shown in Figure 9, Table VII, and

Figure 10. Figure 9 shows scatterplots of predicted and observed numbers of visits from each origin to each destination in each month between April and September of 2019. Note that there are many observations clustered in the lower left corner of each plot. This is because there are many origin-destination combinations with very few visits (e.g., people who live in ZIP codes in the southern half of the state will rarely visit lakes in the northern half of the state, and vice versa). Nevertheless, we observe that the fit is not entirely driven by the origin-destination pairs with the highest number of visits. The clustering of circles around the 45 degree line is evident even at relatively low numbers of trips.

Table VII contains estimates of the travel cost coefficient in each month. The estimates are remarkably stable, all between 0.10 and 0.13. As explained by MacNair et al. [64], when the frequency of trips to the modeled destinations is not a large fraction of all choice occasions, the inverse of the travel cost coefficient is a close approximation of the value per trip (specifically, the willingness-to-pay for access to all modeled destinations divided by the expected number of trips to any destination). So in this case the values per trip are between roughly $8 and $10.

Table VII. Estimated travel cost coefficients for April–September 2019.

Month	$\hat{\lambda}$
April	0.116
May	0.136
June	0.111
July	0.123
August	0.117
September	0.117

Figure 10 shows histograms of the estimated origin and destination fixed effects. The origin fixed effects ($\hat{\alpha}$'s) are centered near zero but span a wide range from -10 to more than 15. The destination fixed effects ($\hat{\theta}$'s) are centered between 0 and 5 and span a similar range as the $\hat{\alpha}$'s. These estimates of origin fixed effects suggest that there is substantial heterogeneity among ZIP codes in preferences for lake recreation activities relative to other activies, or the quantity and quality of nearby substitute sites, or both. The estimates of destination fixed effects suggest there is substantial heterogeneity among lakes in the suite of attributes that make them attractive (or unattractive) for outdoor recreation activities. It should be possible to model this heterogeneity in a second-stage regression with estimated fixed effects as the dependent variable and a vector of measured lake attributes as explanatory variables [65].[16]

Table VIII shows the means and quartiles of the predicted number of trips per person per month to the 100 California lakes in our dataset among all California ZIP codes April–September

16. This provides an opportunity for estimating β using revealed preference data rather than transferring a value from prior studies, as we have done here. This extension is among our top priorities for follow-up work.

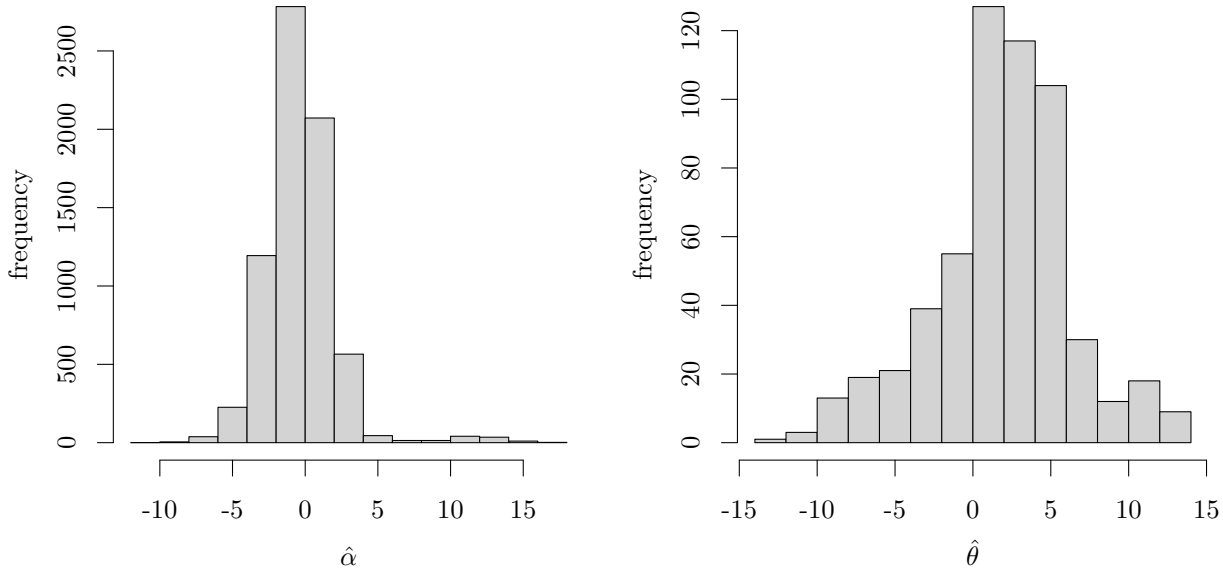

Figure 10. Histograms of estimated origin (left) and destination (right) fixed effects for all months April–September 2019.

Table VIII. Quartiles of predicted **average trips per capita** among ZIP codes in each month April–September 2019.

Month	Mean	0%	25%	50%	75%	100%
April	0.119	0.000284	0.00277	0.0108	0.0428	4.15
May	0.154	0.000294	0.00309	0.0103	0.0446	5.47
June	0.152	0.000284	0.00343	0.0138	0.0613	4.94
July	0.147	0.000366	0.00359	0.0160	0.0601	4.26
August	0.150	0.000294	0.00314	0.0121	0.0542	5.88
September	0.112	0.000309	0.00259	0.00917	0.0380	4.01

2019. The shoulder months of April and September have the fewest trips, while the peak months of May through August each have roughly 25% more trips than April and September. We also see that trips are concentrated among a relatively few ZIP codes, as indicated by the means greater than the 75th percentile in each month. This suggests that ZIP codes closest to one or more lakes account for most of the trips, which in turn suggests the high importance of travel costs to people's lake visitation decisions. The means are between 0.1 and 0.15, which suggests that the average California resident visits at least one of the lakes in our dataset approximately once every 8 to 10 (summer) months. The 100th percentiles are between 3.9 and 6.6, so the average number of trips per capita in the ZIP codes with the highest rates of visitation can be more than once per week.

Table IX shows the means and quartiles of average cost per trip to the 100 California lakes in

our dataset in the months April–September 2019. Costs per trip do not vary much across months but they vary by a factor of three among ZIP codes. Lake visitors from ZIP codes in the first quartile spend around $65 per trip while those in the highest quartile spend up to $165 per trip on average, so travel about 8 times as far to visit a lake ((165-50)/(65-50) ≈ 8). Average costs per trip are between $78 and $88. Note that these cost estimates assume a fixed cost per trip of $50, uniform among all ZIP codes.[17] These results suggest that many trips are close to home, i.e., most visits to the lakes in our case study are from nearby ZIP codes. The full distribution of travel costs for visits in all months April–September 2019 is shown in Figure 11.

Table IX. Quartiles of **average cost per trip** [$] among ZIP codes in each month April–September 2019.

Month	Mean	0%	25%	50%	75%	100%
April	88.1	50	73.7	87.4	102.3	163
May	85.3	50	71.0	84.0	99.7	168
June	82.3	50	69.3	80.4	94.8	161
July	78.3	50	65.1	75.9	90.3	159
August	81.6	50	66.6	80.1	95.0	165
September	80.0	50	68.1	79.8	91.9	154

Table X shows the distribution of willingness-to-pay for access to the full set of lakes in the choice set among all ZIP codes in each month. As with trips per capita, these distributions are highly skewed, with most ZIP codes having very low values (less than $1 per month) and a few ZIP codes having high values up to $66 in the month of August.[18]

Table XI shows the distributions of per capita willingness-to-pay for complete mitigation at all lakes in each month April–September 2019. These estimates follow from our benchmark assumed value of willingness pay to eliminate a HAB at a visited site of $15, based on the literature review results summarized in section 3.2, and the historic frequency of HABs at the 100 lakes in our case study, summarized for the top 50 most severely impacted lakes in Figure 7. The average HAB risk among the 100 lakes included in our case study is 0.0241 per day, which explains why the estimated values for complete mitigation shown in Table XI are two orders of magnitude smaller than the estimated willingness-to-pay for access shown in Table X. Encountering a HAB at our case study lakes is still a relatively infrequent event, and the value of eliminating the risk is a product of the probability and the magnitude of the consequence.

17. This is an ad hoc guesstimate. It is intended to put us in the right ballpark but should be refined using supplemental data in follow-up work.

18. If these average values seem low, consider that the average number trips per month is around 0.15, so the average expenditures per month is around $12. As a cross-check, we conducted a side calculation using a model with three sites to confirm that these magnitudes are reasonable. Specifically, we assigned travel costs for the three sites of $75, $100, and $125. Site fixed effects were 2, 3, and 4, and the travel cost coefficient = 0.1. These assumptions give monthly trip probability of 0.155, monthly average expenditures of $12.6 (excluding fixed costs per trip), and monthly WTP for access of $1.56, all of which are in line with our average results shown in Tables VIII – X.

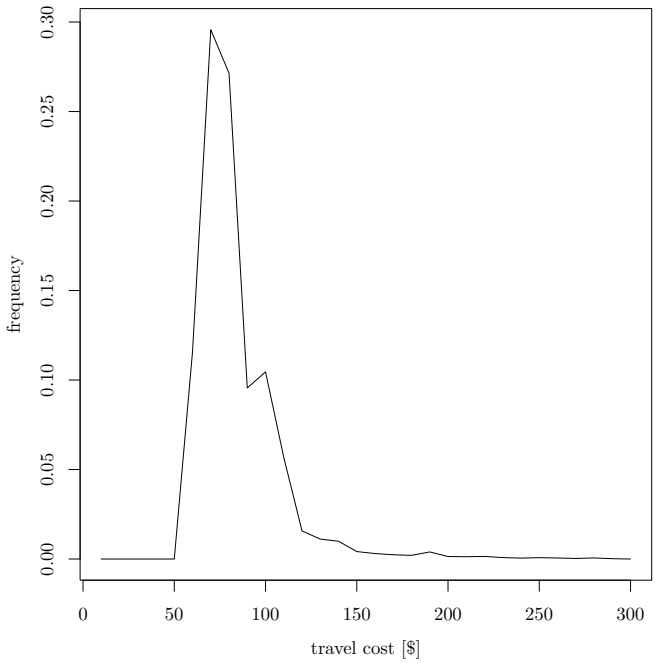

Figure 11. Frequency distribution of travel costs for all visits from ZIP codes in California to 100 California lakes in our dataset April–September 2019.

Table X. Quartiles of **average WTP for access** [$] among ZIP codes in each month April–September 2019.

Month	Mean	0%	25%	50%	75%	100%
April	1.05	0.00245	0.0238	0.0927	0.369	38.4
May	1.17	0.00217	0.0228	0.0758	0.329	44.4
June	1.41	0.00257	0.0309	0.125	0.553	48.7
July	1.22	0.00297	0.0291	0.130	0.488	37.1
August	1.31	0.00251	0.0267	0.103	0.462	55.6
September	0.981	0.00265	0.0222	0.0787	0.326	36.9

TableXII shows the total values of complete mitigation of HAB risks, aggregated over all individuals in all ZIP codes, at three assumed levels of willingness-to-pay to eliminate a HAB at a visited site: $5, $15, and $25. The middle column of results are based on our benchmark value of $15, which is the basis for the results shown in Table XI. The left and right columns in Table XII show how the estimates scale with β. The total value of mitigation when $wtp = 5$ is $2.34 million, which is less than 1/3 of the total value when $wtp = 15$, and the total value when $wtp = 25$ is $13.1 million, which is more than 2/3 larger than the total value when $wtp = 15$. These results

Table XI. Quartiles of **WTP for mitigation** [$] among ZIP codes in each month April–September 2019 (assuming $wtp = \$15$).

Month	Mean	0%	25%	50%	75%	100%
April	0.0377	1.35e-06	4.94e-04	1.42e-03	5.05e-03	2.26
May	0.0499	2.88e-08	4.14e-04	1.20e-03	4.29e-03	3.87
June	0.0426	1.52e-06	5.09e-04	1.61e-03	5.18e-03	3.83
July	0.0361	3.71e-08	4.28e-04	1.35e-03	4.79e-03	4.48
August	0.0429	2.43e-07	4.47e-04	1.33e-03	4.91e-03	4.45
September	0.0389	5.06e-06	4.13e-04	1.11e-03	3.98e-03	4.72

Table XII. The total value of complete mitigation of HAB risks [$] conditional on willingness-to-pay to eliminate a HAB at a visited lake (wtp) equal to $5, $15, or $25.

Month in 2019	$wtp = \$5$	$wtp = \$15$	$wtp = \$25$
April	334,000	1,010,000	1,850,000
May	524,000	1,360,000	2,890,000
June	442,000	1,420,000	2,560,000
July	299,000	750,000	1,670,000
August	355,000	1,140,000	2,010,000
September	387,000	1,380,000	2,090,000
Total	2,340,000	7,410,000	13,100,000

indicate that the value of mitigating HABs is not directly proportional to β but rather increases at an increasing rate (over the range considered here). Therefore, we cannot simply scale our value estimates with β if new evidence about this quantity is published in future studies, or if we are able to produce a better estimate of β in future work on this case study.

Table XIII shows our final estimates for the main quantities of interest in this study: the total value of perfect information about the presence or absence of HABs among the 100 California lakes in our case study. As in Table XII, here we show the estimates at three levels of willingness-to-pay to eliminate a HAB at a visited lake, which indicates a similar convex relationship to β. At our benchmark value of $wtp = 15$, the total value of a perfect early warning system is $2.46 million for the months April–September 2019. The ratio of this value to the corresponding total value of mitigation shown in Table XII is 0.332. This ratio also increases with β over the range considered here. It is 0.314 at $wtp = 5$ and 0.353 at $wtp = 25$. This is consistent with the nonlinear nature of

the relationship shown by our minimal model in section 2.1.

Table XIII. The total value of a perfect early warning system [$] conditional on willingness-to-pay to eliminate a HAB at a visited lake (wtp) equal to $1, $10, or $20.

Month in 2019	wtp $5	$wtp = \$15$	$wtp = \$25$
April	89,600	290,000	551,000
May	121,000	333,000	726,000
June	140,000	470,000	887,000
July	178,000	504,000	1,170,000
August	127,000	432,000	813,000
September	77,200	294,000	486,000
Total	734,000	2,460,000	4,630,000

Table XIV assembles summary results for each month on the total number of trips taken from ZIP codes in California to our 100 case study lakes, the total willingness-to-pay for access and for mitigation, and the total value of perfect information about the occurrence of HABs. About 17.3 million trips were taken to the 100 lakes during the time period of our case study, the total value of access was $191 million, the total value of complete mitigation of HAB risks was $7.41 million, and the total value of perfect information was $2.46 million.

Table XIV. Trips, willingness-to-pay for access, willingness-to-pay for mitigation, and the value of perfect information about HABs in each month April–September 2019 (assuming *wtp* to eliminate a HAB at a visited lake = $15).

Month	Trips	WTP for access [$]	WTP for mitigation [$]	Value of perfect information [$]
April	1,460,000	25,300,000	1,010,000	290,000
May	1,700,000	29,700,000	1,360,000	333,000
June	3,630,000	38,200,000	1,420,000	470,000
July	4,840,000	22,300,000	750,000	504,000
August	3,640,000	27,400,000	1,140,000	432,000
September	2,000,000	26,700,000	1,380,000	294,000
Total	17,300,000	170,000,000	7,060,000	2,320,000

4 Discussion

In this study, we developed and applied a new approach for estimating an outdoor recreation site-choice model using mobility data. We also developed an algorithm that leverages the estimated recreation demand model to compute the value of perfect information about recreation site conditions that are not readily observable by recreators before they visit a site. We applied our framework to a case study of an early warning system for harmful algal blooms (HABs) in 100 lakes used for water-based recreation activities in California. We reviewed eight previously-published non-market valuation studies focused on HABs or excess algae in surface water bodies used for outdoor recreation. We extracted summary estimates of willingness-to-pay to eliminate a HAB at a recreation site from these studies, and found that most central estimates were between $3 and $30 per visitor per visit. We used a figure near the middle of this range to calibrate the algae coefficient in the indirect utility function of our recreation demand model, β. We also performed a sensitivity analysis to examine how the value of mitigation and early warnings scales with β. We found that both the value of mitigation and the value of a perfect early warning system increases at an increasing rate with β over the range of values suggested in our review of the literature. Under our benchmark assumptions, our estimates of the total value of a perfect early warning system covering the 100

lakes in our dataset for the months April–September 2019 is $2.46 million, and the total value of complete mitigation at the same set of lakes is $7.41 million.

The value attributed to an early warning system by the framework used in this study stems from the role of such a system in allowing recreators to choose to visit a different site if they want to avoid a HAB at their otherwise preferred site. That is, such a system would have no effect on the water quality conditions in the lakes themselves, but would allow people to better adapt to the worsening conditions over time. Mitigation, on the other hand, would completely eliminate the adverse effects of HABS on recreation opportunities and other harmful effects on ecosystems or human health. In follow-up research we intend to build on the results reported here towards a more comprehensive analysis of the benefits and costs of these two strategies—mitigation and adaptation through early warnings—applied both separately and jointly at varying levels of coverage or stringency.

The recreation demand model and estimation approach developed in this study provides another tool for non-market valuation using revealed preference data. The approach is designed to use origin-destination data collected from cell phones or other personal electronic devices that record location information using GPS or wireless network technology. We were able to use these data in place of administrative data on visitation rates, which were not readily available for a large share of the sites in our study area, and we suspect that other large-scale recreation demand studies may face similar challenges in the future.

Inclusion of both destination and origin fixed effects in our estimating equation reduces the risk of omitted variable bias (at both ends of a trip, unlike previous approaches that include fixed effects for destinations only), and provides opportunities for consistent second-stage estimation using instrumental variables [65] or possibly other approaches. The simulation experiments reported here suggest that the estimation algorithm is consistent or nearly so under the maintained assumptions of the model.

The value of information estimates presented in this report are estimates of an upper bound, because perfect information is an unattainable ideal for any data collection endeavor. A real-world early warning system close to this ideal might be approached by a satellite data collection system with a good predictive algorithm to translate sensed colors into predictions of HAB occurrences. If predictions from the early warning system are easily accessible and widely used, then we would anticipate a high HAB avoidance rate by those recreators who have strong preferences, either because a HAB would disrupt their plans for water-based recreation activities or because of concerns about the health risks to themselves or their children or pets. The temporal correlation in HAB occurrences is beneficial in this respect. Because HABs typically persist for some days, predictions a few days ahead should be sufficient to provide good forecasts of current conditions in most cases.

It also is possible to estimate the value of improved but imperfect information using an extension of the methods demonstrated in this study. This would require simulating site choices by recreators under intermediate information regimes in which the HAB forecasts at the time of choice are not perfectly accurate but are more accurate than long run historic averages. Under such intermediate regimes, recreators will encounter HABs less often than under the reference scenario examined in this study but more often than under the counterfactual scenario of a perfect early warning system. Recreators' type I and type II errors would be reduced by not eliminated under such intermediate information regimes. The value of imperfect information would necessarily be between zero and the value of perfect information reported here, but we do not know how the value would scale with the

increasing precision of an early warning system. Extending our approach to cover these intermediate cases will be important for evaluating incremental expansions of satellite data collection activities or incremental improvements in predictive models of HABs that use satellite data as input.

Due to the curse of dimensionality, we computed the value of information under the assumption that early warnings can be provided only at the 10 lakes with the highest HAB risks (that is, we examined $2^{10} = 1,024$ combinations of lakes where recreators would know with certainty that a HAB is present or not). Another task for follow-up research is to examine the sensitivity of our VOI estimates to this constraint. This would reveal how rapidly the returns to expanding the scope of an early warning system diminish as additional lakes are covered by a network of remote sensors.

Another high-priority task for follow-up work is to attempt to estimate β, the dis-utility of encountering a HAB at a visited site, using revealed preference data. In this study we transferred a central estimate of β from a set of previously-published non-market valuation studies that are more-or-less similar to our application. At the most basic level, estimation of β using primary data would rely on the correlation between total visits and estimated HAB risks at each destination, controlling for other important factors. However, what drives visitation patterns are people's perceived risks of HABs. An open question is whether people's perceived risks are similar enough to the measures of HAB risks we are able to construct using available data to support estimation of β with reasonable accuracy and precision. If people's perceptions are unbiased on average but noisy, then it would seem we have a classical measurement error problem, which we would expect to bias the estimate of β towards zero. If we can find a secondary proxy measure of perceived risks—one with measurement errors independent of the errors in our primary measure—then it may be possible to develop a consistent estimator for β using the covariance between the two proxy measures in lieu of the variance of a single direct measure with no error, which may be unattainable (e.g., see Kennedy [66, p 168]).

Another important assumption of our framework is that of classically rational agents, i.e., expected utility-maximizers who assimilate new information using Bayes' rule. Roberts et al. [60] found that a prospect theory model fit their contingent valuation survey data better than an expected utility model, so another important task for follow-up research will be to develop approaches to estimate the value of mitigation and early warning systems when individuals follow non-expected-utility decision rules.

To summarize, we developed a new recreation demand modeling approach using mobility data on the number of visits from many origin zones to many destination sites, and we developed a companion algorithm to estimate the value of perfect information about uncertain recreation site conditions. We applied the framework to a case study of harmful algal blooms at 100 large lakes in California. Our results suggest that the value of complete HAB mitigation at the case study lakes in the months April – September 2019 was \$7.41 million, and the value of a perfect early warning system was \$2.46 million. Our top priorities for follow-up research include estimating the value of incremental improvements in the predictive accuracy of early warning systems, estimating the dis-utility of encountering a HAB at a visited site using the mobility data, and estimating the value of information under alternative models of decision-making under uncertainty for comparison to the conventional expected utility model used in this study.

References

[1] BW Brooks, JM Lazorchak, MDA Howard, M-VV Johnson, SL Morton, DAK Perkins, ED Reavie, GI Scott, SA Smith, and JA Stevens. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environmental Toxicology and Chemistry*, 35(1):6–13, 2016.

[2] SK Moore, MR Cline, K Blair, T Klinger, A Varney, and K Norman. An index of fisheries closures due to harmful algal blooms and a framework for identifying vulnerable fishing communities on the US West Coast. *Marine Policy*, 2019.

[3] D Wolf, W Georgic, and HA Klaiber. Reeling in the damages: harmful algal blooms' impact on Lake Erie's recreational fishing industry. *Journal of Environmental Management*, 199:148–157, 2017.

[4] David Wolf, Wei Chen, Sathya Gopalakrishnan, Timothy Haab, and H Allen Klaiber. The impacts of harmful algal blooms and E. coli on recreational behavior in Lake Erie. *Land Economics*, 95(4):455–472, 2019.

[5] D Wolf and HA Klaiber. Bloom and bust: toxic algae's impact on nearby property values. *Ecological Economics*, 135:209–221, 2017.

[6] Congressional Research Service. Freshwater harmful algal blooms: causes, challenges, and policy considerations, 2019. `https://fas.org/sgp/crs/misc/R44871.pdf`.

[7] BA Jones. Infant health impacts of freshwater algal blooms: Evidence from an invasive species natural experiment. *Journal of Environmental Economics and Management*, 96:36–59, 2019.

[8] CRC Kouakou and TG Poder. Economic impact of harmful algal blooms on human health: a systematic review. *Journal of Water and Health*, 17(4):499–516, 2019.

[9] J Heisler, PM Glibert, JM Burkholder, DM Anderson, W Cochlan, WC Dennison, Q Dortch, CJ Gobler, CA Heil, E Humphries, et al. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae*, 8(1):3–13, 2008.

[10] AM Michalak, EJ Anderson, D Beletsky, S Boland, NS Bosch, TB Bridgeman, JD Chaffin, K Cho, R Confesor, and I Daloğlu. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, 110(16):6448–6452, 2013.

[11] PM Glibert and MA Burford. Globally changing nutrient loads and harmful algal blooms: recent advances, new paradigms, and continuing challenges. *Oceanography*, 30(1):58–69, 2017.

[12] HW Paerl, TG Otten, and R Kudela. Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum, 2018.

[13] SC Moser. Adaptation, mitigation, and their disharmonious discontents: an essay. *Climatic Change*, 111(2):165–175, 2012.

[14] DB Lobell, ULC Baldos, and TW Hertel. Climate adaptation as mitigation: the case of agricultural investments. *Environmental Research Letters*, 8(1):015012, 2013.

[15] AK Magnan and T Ribera. Global adaptation after Paris. *Science*, 352(6291):1280–1282, 2016.

[16] J Martinich and A Crimmins. Climate damages and adaptation potential across diverse sectors of the united states. *Nature Climate Change*, 9(5):397, 2019.

[17] State of Oregon. Recreation Use Values Bibliography: `http://recvaluation.forestry.oregonstate.edu/sites/default/files/RUVD_biblio_2016.pdf`, 2020.

[18] DJ Phaneuf and VK Smith. Recreation demand models. *Handbook of environmental economics*, 2:671–761, 2005.

[19] G Glasgow and K Train. Lost use-value from environmental injury when visitation drops at undamaged sites. *Land Economics*, 94(1):87–96, 2018.

[20] Jon M Conrad. Quasi-option value and the expected value of information. *The Quarterly Journal of Economics*, 94(4):813–820, 1980.

[21] W Michael Hanemann. Information and the concept of option value. *Journal of Environmental Economics and management*, 16(1):23–37, 1989.

[22] Paul Mensink and Till Requate. The dixit–pindyck and the arrow–fisher–hanemann–henry option values are not equivalent: a note on fisher (2000). *Resource and Energy Economics*, 27(1):83–88, 2005.

[23] Fumie Yokota and Kimberly M Thompson. Value of information analysis in environmental health risk management decisions: past, present, and future. *Risk Analysis*, 24(3):635–650, 2004.

[24] Ramanan Laxminarayan and Molly K Macauley. *The Value of Information: Methodological Frontiers and new Applications in Environment and Health*. Springer Science & Business Media, 2012.

[25] Roger Cooke, Bruce A Wielicki, David F Young, and Martin G Mlynczak. Value of information for climate observing systems. *Environment Systems and Decisions*, 34(1):98–109, 2014.

[26] SC Newbold and AL Marten. The value of information for integrated assessment models of climate change. *Journal of Environmental Economics and Management*, 68(1):111–123, 2014.

[27] DR Petrolia, J Penn, R Quainoo, RH Caffey, and JM Fannin. Know thy beach: values of beach condition information. *Marine Resource Economics*, 34(4):331–359, 2019.

[28] D Jin and P Hoagland. The value of harmful algal bloom predictions to the nearshore commercial shellfish fishery in the Gulf of Maine. *Harmful Algae*, 7(6):772–781, 2008.

[29] M Bloudoff-Indelicato. Satellites help keep communities safe from toxic algal blooms. *Resources Magazine*, 2019. `https://www.resourcesmag.org/archives/satellites-help-keep-communities-safe-toxic-algal-blooms/`.

[30] Y Kuwayama and B Mabee. How do we measure the value of satellite data? *Resources Magazine*, 2018. https://www.resourcesmag.org/archives/how-do-we-measure-the-value-of-satellite-data/.

[31] M Dickie. Defensive behavior and damage cost methods. In *A Primer on Nonmarket Valuation*, pages 395–444. Springer, 2003.

[32] Y Konishi and K Adachi. A framework for estimating willingness-to-pay to avoid endogenous environmental risks. *Resource and Energy Economics*, 33(1):130–154, 2011.

[33] Michael J Messner and Philip Berger. Cryptosporidium infection risk: results of new dose-response modeling. *Risk Analysis*, 36(10):1969–1982, 2016.

[34] Aksana Chyzheuskaya, Martin Cormican, Raghavendra Srivinas, Diarmuid O'Donovan, Martina Prendergast, Cathal O'Donoghue, and Dearbháile Morris. Economic assessment of waterborne outbreak of cryptosporidiosis. *Emerging Infectious Diseases*, 23(10):1650, 2017.

[35] Patricia Kocagil, Nadia Demarteau, Ann Fisher, and James S Shortle. The value of preventing cryptosporidium contamination. *Risk*, 9:175, 1998.

[36] LH Palm-Forster, F Lupi, and M Chen. Valuing Lake Erie beaches using value and function transfers. *Agricultural and Resource Economics Review*, 45(2):270–292, 2016.

[37] W Zhang and B Sohngen. Do US anglers care about harmful algal blooms? A discrete choice experiment of Lake Erie recreational anglers. *American Journal of Agricultural Economics*, 100(3):868–888, 2018.

[38] D McFadden. Econometric models of probabilistic choice. In C Manski and D McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*. MIT press Cambridge, MA, 1981.

[39] Daniel McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.

[40] TC Haab and KE McConnell. *Valuing Environmental and Natural Resources: the Econometrics of Non-market Valuation*. Edward Elgar Publishing, 2002.

[41] Giovanni Parmigiani and Lurdes Inoue. *Decision Theory: Principles and Approaches*, volume 812. John Wiley & Sons, 2009.

[42] Patrick Bayer and Christopher Timmins. Estimating equilibrium models of sorting across locations. *The Economic Journal*, 117(518):353–374, 2007.

[43] H Allen Klaiber and Daniel J Phaneuf. Valuing open space in a residential sorting model of the twin cities. *Journal of Environmental Economics and Management*, 60(2):57–77, 2010.

[44] Handling unobserved site characteristicds in random utility models of recreation demand. *Journal of Environmental Economics and Management*, 51:1–25, 2006.

[45] Microeconometric streategies for dealing with unobservables and endogenous variables in recreation demand models, 2011.

[46] Babatunde O Abidoye, Joseph A Herriges, and Justin L Tobias. Controlling for observed and unobserved site characteristics in rum models of recreation demand. *American Journal of Agricultural Economics*, 94(5):1070–1093, 2012.

[47] Do random coefficients and alternative specific constants improve policy analysis? An empirical investigation of model fit and prediction. *Environmental and Resource Economics*, 73:75–91, 2019.

[48] George R Parsons. A note on choice of residential location in travel cost demand models. *Land Economics*, 67(3):360–364, 1991.

[49] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[50] State of California. California My Water Quality HAB Data Viewer: `https://mywaterquality.ca.gov/habs/data_viewer/`, 2020.

[51] SFEI. San Francisco Estuary Institute HAB Satellite Analysis Tool: `https://fhab.sfei.org/`, 2020.

[52] Nathaniel H Merrill, Sarina F Atkinson, Kate K Mulvaney, Marisa J Mazzotta, and Justin Bousquin. Using data derived from cellular phone locations to estimate visitation to natural areas: an application to water recreation in New England, USA. *PloS one*, 15(4):e0231863, 2020.

[53] Christopher Monz, Milan Mitrovich, Ashley D'Antonio, Abigail Sisneros-Kidd, et al. Using mobile device data to estimate visitation in Parks and protected areas: an example from the nature reserve of orange County, California. *Journal of Park and Recreation Administration*, 37(4):92–109, 2019.

[54] European Space Agency. Sentinel User Guide: `https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/product-types/level-2-water`, 2020.

[55] SJ Goetz, N Gardiner, and JH Viers. Monitoring freshwater, estuarine and near-short benthic ecosystems with multi-sensor remote sensing: an introduction to the special issue. *Remote Sensing of Environment*, 112(11):3993–3995, 2008.

[56] Sergio Alvarez, Frank Lupi, Daniel Solís, and Michael Thomas. Valuing provision scenarios of coastal ecosystem services: the case of boat ramp closures due to harmful algae blooms in florida. *Water*, 11(6):1250, 2019.

[57] Lin Crase and Rob Gillespie. The impact of water quality and water level on the recreation values of lake hume. *Australasian Journal of Environmental Management*, 15(1):21–29, 2008.

[58] KJ Egan, JA Herriges, CL Kling, and JA Downing. Valuing water quality as a function of water quality measures. *American Journal of Agricultural Economics*, 91(1):106–123, 2009.

[59] Steven Piper, Marc O Ribaudo, and Ardelle Lundeen. The recreational benefits from an improvement in water quality at oakwood lakes and lake poinsett, south dakota. *North Central Journal of Agricultural Economics*, pages 279–287, 1987.

[60] David C Roberts, Tracy A Boyer, and Jayson L Lusk. Preferences for environmental quality under uncertainty. *Ecological Economics*, 66(4):584–593, 2008.

[61] G Van Houtven, C Mansfield, DJ Phaneuf, R von Haefen, B Milstead, MA Kenney, and KH Reckhow. Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. *Ecological Economics*, 99:40–52, 2014.

[62] Randy Turner and Pete Kauhanen. Visualizing cyanobacteria from space. 2019. `https://www.nalms.org/wp-content/uploads/2019/07/39-2-7.pdf`.

[63] Derrick Hang, Daniel McFadden, Kenneth Train, and Ken Wise. Is vehicle depreciation a component of marginal travel cost? *Journal of Transport Economics and Policy*, 50(2):1–19, 2016.

[64] Doug MacNair, George Parsons, Theodore Tomasi, and Heath Byrd. Trip equivalency for economic valuation in recreation demand models: implications for compensatory restoration and benefits transfer. *Marine Resource Economics*, 37(1):91–107, 2022.

[65] Christopher Timmins and Jennifer Murdock. A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and Management*, 53(2):230–249, 2007.

[66] Peter Kennedy. *A Guide to Econometrics*. Blackwell Publishing, 6 edition, 2008.