Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes

 $\bf{Bonnie~L~Keeler}^{\bf 1*}, \bf{Spencer~A~Wood}^{\bf 2,3}, \bf{Stephen~Polasky}^{\bf 4}, \bf{Catherine~Kling}^{\bf 5}, \bf{Christopher~T~Filstrup}^{\bf 6},$ **and John A Downing⁶**

More than 41 000 water bodies are listed as impaired by the US Environmental Protection Agency under the Clean Water Act. Implementation and enforcement of regulations designed to address these impairments can be costly, raising questions about the value of the public benefits derived from improved surface water quality. Here, we assess the recreational value of changes in water quality using freely available geotagged photographs, taken by members of the public, as a proxy for recreational visits to lakes. We found that improved water clarity is associated with increased numbers of visits to lakes and that lake users were willing to incur greater costs to visit clearer lakes. Lake users were willing to travel 56 minutes farther (equivalent to US\$22 in travel costs) for every one-meter increase in water clarity in Minnesota and Iowa lakes, when controlling for other lake attributes. Our approach demonstrates the potential for social-media data to inform social–ecological research, including assessment of the recreational benefits of improvements in water quality.

Front Ecol Environ **2015; 13(2): 76–81, doi:10.1890/140124 (published online 29 Jan 2015)**

Lakes, rivers, and streams provide many benefits to the general public, but these services are not captured in **general public, but these services are not captured in** markets and have proven difficult to quantify (Brauman *et al*. 2007; Keeler *et al*. 2012). This is problematic because information on the value of water resources is needed in many policy and regulatory contexts. For example, the US Environmental Protection Agency is charged with estimating the benefits and costs associated with major rules and regulations designed to safeguard aquatic habitats (Griffiths *et al*. 2012). Cost–benefit assessments for water-quality changes are also considered in the design of payment and incentive programs, as well as in spatial planning decisions related to investments in conservation or habitat restoration (Olmstead 2010; Griffiths *et al*. 2012). Lack of information about the value of water-quality benefits can complicate justifying major spending on improved water quality.

Despite high demand, estimates of the value of clean water are often not available at the relevant scale for proposed interventions, are time- and resource-intensive to obtain, and are difficult to link to empirical measurements of water quality (Iovanna and Griffiths 2006; Keeler *et al*.

¹The Natural Capital Project, Institute on the Environment, University *of Minnesota, St Paul, MN * (keeler@umn.edu); 2 The Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA; ³ School for Environmental and Forest Sciences, University of Washington, Seattle, WA; ⁴ Department of Applied Economics and Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, MN; ⁵ Center for Agriculture and Rural Development, Department of Economics, Iowa State University, Ames, IA; ⁶ Ecology, Evolution, and Organismal Biology Department, Iowa State University, Ames, IA*

2012). The most common non-market methods for estimating the value of water-quality improvements typically require time-consuming and costly surveys, either to assess respondents' stated willingness-to-pay for improved water quality (Carson and Mitchell 1993) or to gather information on past recreational behavior (sites visited and distances traveled to sites; see Feather *et al*. 1995; Phaneuf and Smith 2005; Egan *et al*. 2009). In the first case, a survey is administered to ask respondents how much they would be willing to pay for a given change in water quality (stated preference approach). In the second case, users are asked information about their past behavior, and values are assessed based on how much (in terms of time and lost wages) they have conceded to obtain a higher-quality resource or experience (revealed preference approach). By design, survey data are often site-specific, limited in temporal and spatial scale, and not easily applicable to other decision-making contexts (Freeman *et al*. 2014). Estimates of benefits are also not typically expressed in terms of changes in water quality that can be linked to pollutant loads or land-use change, making it difficult to compare the costs and benefits associated with additional restoration or protection measures (Wilson and Carpenter 1999; Keeler *et al*. 2012).

We investigated how recreational lake users respond to variations in lake water quality, using data from an online social-media source as an alternative to survey data. Specifically, we used geotagged photographs (images associated with spatial-coordinate metadata) uploaded to the photo-sharing website Flickr (www.flickr.com) to estimate the number of visits to different lakes. We counted the number of uploaded photos taken by individual users on unique days that fell within selected boundaries of each lake

and used this "photo-user-days" measure as a proxy for lake visitation. We coupled these data on lake visits with information provided by Flickr users to estimate the distances that visitors traveled from their selfreported hometown to each photographed lake (assuming transport by passenger vehicle along known road networks). This approach offers some advantages over traditional survey tools in that behavioral data can be collected over longer time periods, across broad spatial scales, and at minimal cost. After extracting and processing the geotagged photographs taken at each lake, we used multiple regression analysis to determine which lake attributes and other factors best explain patterns of lake visitation and travel costs. We then applied the regression model to a scenario involving improved water quality to evaluate observed changes in the numbers of unique visits to lakes and the value that visitors associate with improved water quality.

n Methods

Lake attributes

Our study assessed the relationship between lake visitation and selected lake attributes for over 1000 lakes in the Midwestern US states of Minnesota and Iowa. We chose these states because of the availabil-

ity of water-quality data (for lakes in both states) and survey data for Iowa lake users; lakes within this region also represent a gradient of water-quality conditions, from relatively undisturbed oligotrophic lakes to lower-quality eutrophic ones. Water clarity in the study region's lakes typically ranges from depths greater than 10 m to less than 0.5 m, encompassing most of the range of lake water clarity observed worldwide (Watson *et al*. 1992) and thereby making this a good study system for our purposes. In addition to water clarity, we also assembled data for various other explanatory variables, including lake-water chemistry, lake depth and size, near-lake human populations, lake amenities such as boat ramps (from which watercraft may enter and be retrieved from the water; also known as slipways or launches) and fishing piers, proximity to state parks and the Boundary Waters Canoe Area Wilderness, and the presence of aquatic invasive species (see WebPanel 1 and WebTable 1 for more details on water quality and other lake attribute data).

Lake visitation

Our study takes advantage of the increase in spatially explicit voluntarily supplied content available online. These shared data are increasing in volume each year and allow researchers to rapidly and inexpensively study user behavior and preferences over space and time (Wood *et al*. 2013). Here we use the photo-sharing website Flickr because it represents one of the largest available datasets

Figure 1. Distribution of photo-visitations in Minnesota and Iowa lakes as measured by Flickr photographs. Photo-user-days per lake represent the sum of all unique daily lake and user combinations uploaded to Flickr between 2005 and 2012.

of geotagged images and has an application program interface (API) that facilitates data extraction. To assess visitation at Minnesota and Iowa lakes, we queried Flickr for all geotagged images (in our study, these were photographs with latitude and longitude data) taken from January 2005 to December 2012 within the boundaries of over 3000 lakes in Minnesota and over 100 lakes in Iowa. In the corresponding geographic information system (GIS) analysis, we established a 30-m buffer zone around each lake – measured outward from the water's edge – to account for photographs taken along the shoreline. Our search returned a total of 41 852 unique geotagged photographs for Minnesota and Iowa lakes.

For each lake associated with geotagged photographs, we estimated the number of unique photo-user-days per lake (the count of unique combinations of users and lake destinations within a 24-hour period). For instance, if an individual took multiple photos at the same lake on the same day, that would equate to a single photo-user-day. These data were averaged across the 8-year period for which photos were downloaded (2005 to 2012) to derive an average annual number of photo-user-days per lake. In Minnesota, 1079 lakes were visited and photographed by Flickr users; in Iowa, 72 lakes returned geotagged images (Figure 1).

To obtain home location information needed for estimating travel routes, we also downloaded publicly available user-profile information associated with individuals who uploaded the photographs in our sample set. We deleted all personally identifiable information and

Figure 2. Average visitor numbers per year to Iowa lakes and Minnesota state parks, measured as photo-user-days, as compared with the number of trips per year estimated with traditional surveys. Each observation is a lake in Iowa or a state park in Minnesota. Dotted line is a 1:1 relationship between photovisitation and surveyed visitation. Minnesota state park data are from Wood et al*. (2013). Trendline equations are non-linear fits of the untransformed data plotted on log-log axes. Corresponding* R*² values for each regression are 0.65 (Iowa) and 0.70 (Minnesota).*

assigned each user a numerical identification code that was associated with the number and location (not the content) of their geotagged lake photos and their userspecified home location. About 40% of Flickr users who uploaded pictures of lakes provided their home location in their public profile. Flickr users who visited Minnesota lakes came from 47 US states and 36 other countries, with 66% of visitors reporting a home location from Minnesota. There were significantly fewer Flickr users who visited Iowa lakes; these visits originated from 20 US states and there were no international visitors.

Are online photos a proxy for visitation?

Wood *et al*. (2013) compared surveyed data on the number of visits to various sites to visitation estimated by Flickr photo-user-days – using nine datasets consisting of 836 different natural and cultural attractions worldwide – and found this metric to be a good proxy for surveyed visitation rates. To evaluate the applicability of the photo-visitation method to lakes, we obtained data from a statewide survey of Iowa lake users conducted by Iowa State University (Evans *et al*. 2011). Survey information on lake visitation was reported over 5 years (2002–2005 and 2009) for 86 lakes in Iowa. We calculated average annual trips per lake over the 5 years for which data were available and plotted these values against photo-user-days for Iowa lakes estimated from Flickr (Figure 2). We found a significant positive relationship between 2005–2012 photo-user-days and surveyed visitation in Iowa lakes (R^2) = 0.65; Figure 2). This relationship is similar to one presented by Wood *et al*. (2013), between surveyed visits to Minnesota state parks and Flickr photo-user-days $(R^2 = 0.70;$ Figure 2).

Distance analysis

Photograph data can also be used to estimate the distance traveled or time spent traveling from a user's stated home location to a lake destination. For the subset of Flickr users who provided information on the location of their home in their online Flickr profile, we mapped each user's hometown to spatial coordinates in a database of populated places. We considered only users with hometowns in 12 nearby Midwestern US states (CO, IA, IL, IN, KS, MI, MN, MO, ND, NE, SD, WI). Users residing in other states were assumed to have used air travel or other modes of transportation to visit Minnesota and Iowa lakes and were excluded from the distance analysis.

To estimate the distance traveled to visit a lake, we performed a distance analysis in ArcGIS. We used the ESRI ArcGIS Business Analyst Desktop (Redlands, CA) and 2012

NAVTEQ Street Data (Greenwood Village, CO) to estimate the travel time (accounting for posted speed limits for vehicular traffic along known routes) from each home location to visited lakes (Figure 3). For consecutive-day trips of less than 80 km, we assumed that users returned home between each lake visit. For trips greater than 80 km, we deleted routes where the same lake or different lakes were visited on consecutive days, assuming that the visitor stayed at or near the lake overnight and did not return home between lake visits. After removing consecutive day trips, our database contained 6438 trips to Minnesota and Iowa lakes from 12 neighboring states. For each lake visited by a Flickr user with a known home location, we estimated the average time spent traveling to visit that lake. Of the over 3000 lakes in the dataset, 946 were visited by users with home location information and were assigned average travel-time values.

Regression modeling

We used multiple regression models to identify the relationships between lake attributes, lake visitation, and travel time. For the lake visitation data, we first applied a logistic regression model to identify factors that predicted whether or not a lake was visited (where visitation was defined as lakes being the subject of at least one photograph during the study period; see WebTable 4). We then used linear regression on the subset of lakes that were visited to identify how a hypothetical change in water quality would affect the number of additional visits, assuming that all other lake attributes remained constant. We also applied a multiple linear regression to the route data to estimate how changes in lake water quality would affect travel times to lakes (a proxy for travel cost). For all models we used backwards stepwise regression with Akaike information criterion (AIC) to select the best-fit model of the relationship between the response variables of lake visits and travel costs and the lake-specific explanatory variables (see WebTables 1–3 for parameter estimates, bivariate regressions, and pairwise correlations).

n Results

Which factors predict lake visitation?

We found that lake size, water clarity, near-lake population, presence of a boat ramp, and state (Iowa or Minnesota, represented by a dummy variable) were significant predictors

of annual average per-lake visitation. This set of predictors was significant both in the logistic regression model predicting the probability of a lake receiving at least one visit (WebTable 4) and in the linear regression model estimating per-lake visitation for the visited lakes (Table 1). The relationship between visitation and lake clarity was positive, such that lakes with greater water clarity were associated with higher numbers of visits. As expected, larger lakes received more visits than smaller ones, and lakes with a boat ramp attracted more visitors than lakes without one. Lakes in Iowa also had more average visits than those in Minnesota, presumably because there are fewer lakes to visit. Notably, we observed a bimodal distribution of lake visitation whereby lakes in both densely and sparsely populated areas received high numbers of visits (WebFigure 1).

To account for this distribution in our regression analysis, we centered the population variable in the multiple linear regression model (subtracted mean population from each lake population estimate) and included a squared term for population (WebPanel 2; Table 1).

Which factors affect the distance traveled to visit lakes?

Preferences for lake attributes can be not only inferred from the number of visits to each lake but also based on how far people are willing to travel to visit each lake (Parsons 2003; Egan *et al*. 2009). By using travel time as a proxy for the value individ-

Figure 3. Map of visited lakes, origins (user hometowns), and routes traveled by recreationists. We derived origins and lake destinations from Flickr user profiles and photographs, respectively, and estimated routes using ESRI ArcGIS Business Analyst.

uals place on various lake attributes, we can infer the amount that individuals are willing to trade off to visit lakes with better water quality, with all other lake attributes being equal. In the non-market valuation literature, travel cost estimates are typically based on site counts of visitors or mailed surveys asking respondents to recall the number of their visits to various destinations. Here we estimated the distance traveled from user-specified home locations to different lakes, based on the locations of their geotagged photographs.

We used multiple linear regression to construct a model of travel time as a function of lake attributes and found a significant positive relationship between lake water clarity and travel time (Table 2), indicating that people are willing to spend more time to travel to clearer lakes. A best-

Table 1. Multiple linear regression for lake attributes and photovisitation per lake

Notes: The response variable is visitation in units log(photo-user-days \times yr⁻¹). Each observation refers to a lake located in Minnesota or Iowa that was the subject of at least one recorded photograph during the study period (*n* = 1086 lakes). For explanation and justification of the centered population variable see WebPanel 2 and WebFigure 1, which plots visitation as a function of near-lake population. Effect tests refer to the significance level based on the probability of exceeding the *t* statistic (low values indicate the coefficient is significantly different than zero).

80

Table 2. Multiple linear regression for lake attributes and time spent traveling to each lake

Notes: The response variable is time spent traveling to each lake (one-way). Observations represent all
unique combinations of users and lake destinations on non-consecutive days (n = 6438 routes). Effect
tests refer to cate the coefficient is significantly different than zero).

fit model estimates that drivers were willing to spend an additional 56 minutes in round-trip travel time for each additional meter of depth visibility (ie water clarity). This translates to approximately US\$22.26 per trip via ground transportation that a given user is willing to trade off for improved water quality, assuming one-third the average hourly wage rate and a transportation cost of US\$0.30 per mile (WebTable 5; Parsons 2003). People are also willing to incur greater travel costs to visit larger lakes, lakes in the Boundary Waters Canoe Area Wilderness, and lakes with a boat ramp.

Would improvements to lake water quality increase the number of visits to lakes?

To estimate how a change in water quality would affect the number of visits to lakes, we applied the estimated relationship from the regression equation shown in Table 1 to lakes that received at least one visit over the study period. We estimated the change in photo-visitation between a scenario of baseline water clarity and a scenario where the water clarity of all lakes is increased by one meter. By assuming that the relationship between photo-visitation and surveyed visitation to Iowa lakes holds for all lakes in the sample region, we converted the model estimates of the change in photo-user-days into an estimate of annual trips per lake (based on correlation data presented in Figure 2). Using this approach, we calculated an average increase of 1389 annual trips for an average lake (1305 to 1481, lower and upper 95% mean confidence limits) per one-meter increase in water clarity, if all other variables remained constant. This result reflects the positive relationship between lake water quality and numbers of visitors.

Issues related to how well photo-takers represent all visitors, as well as potential biases in social-media use, contribute to uncertainty in these quantitative estimates. However, the qualitative relationship between visitor numbers and water quality is consistent with survey data and study expectations. That said, regional programs aimed at improving water quality might not generate increased visitation across all affected lakes because users may simply switch from visiting lower-quality lakes to visiting those with improved water quality. A net increase in the numbers of visitors could occur if visiting a clearer lake substituted for leisure time previously spent at a public pool or other attraction.

n Discussion and conclusions

We used photo-visitation data to understand the effects of improved water quality on the number of visits and the distance traveled to lakes. We found that recreational lake users visit clear lakes more often than less-clear lakes and are willing to incur increased travel costs to visit lakes

with better water quality. This conclusion is consistent with stated preference studies using contingent valuation surveys or choice experiments, which have found evidence for a positive relationship between water quality and willingness-to-pay (Carson and Mitchell 1993; Phaneuf 2002; Johnston *et al*. 2003; Banzhaf *et al*. 2006; Viscusi *et al*. 2008; Van Houtven *et al*. 2014). Far fewer studies have evaluated the benefits of improved lake water quality using revealed preference approaches based on the surveyed behavior of recreationists (eg Feather *et al*. 1995; Egan *et al*. 2009).

There are limitations to the use of social-media data to estimate recreational behavior. We recognize that Flickr users are not necessarily representative of all recreationists. We compared demographic characteristics of Flickr users worldwide (reported in Ignite Social Media 2012) with the demographics of Iowa Lakes Survey respondents (reported in Evans *et al*. 2011) and found that Flickr users were more likely to be female and of higher educational status than the Iowa lake users that responded to surveys. Reported income was comparable between the two groups. However, Flickr users were, on average, younger than the Iowa Lakes Survey respondents. At present, little is known about how the behavior and preferences of Flickr users differ from those of other lake users, including information about the lake-related activities that Flickr users participate in and how those activities differ from those of other lake users. Despite these limitations, our comparison of survey data from Iowa (Figure 2) and previous work across multiple sites worldwide (Wood *et al*. 2013) suggest that photo-visitation can be a reliable proxy for actual visitation.

We cannot fully account for the role of bias and representation in our model results due to our reliance on data from social media. Notably, however, similar issues must also be considered in "conventional" approaches to collecting data on preferences and behavior. For example, survey data can be subject to hypothetical bias, nonresponse and sample selection bias, inattentive or hasty responses to questions, recall errors, order effects, and framing effects, all of which can contribute to inaccurate reporting of magnitudes or frequencies (Hanemann 1994; BL Keeler *et al.* Studying lake visitation using Flickr

Kling *et al*. 2012). We hope that this paper stimulates further investigations into sources of bias, representation, and interpretation in social-media-related data, just as progress has been made over the past several decades in understanding the strengths and limitations of other revealed and stated preference methodologies (Hanemann 1994; Kling *et al*. 2012).

This paper contributes information on the benefits of improved water quality, which is needed to inform regulatory cost–benefit assessments: particularly where there is uncertainty surrounding the value generated by proposed investments in improving surface water quality (Griffiths *et al*. 2012). Similar methods could be used to evaluate the benefits of other changes in environmental quality, especially where resource and time constraints prevent the use of survey data or where data are required across broad temporal or spatial scales. The next steps for adapting this approach could include scaling up the analysis to link photo-visitation estimates to regional and national databases on lake water quality. These data can be overlaid with data on known impairments to evaluate the return on investments intended to improve surface water quality, from a single lake up to state or regional scales. In the future, we believe that use of social-media-derived geotagged data on recreational demand will help to inform spatial planning and resource investments, as well as to improve our understanding of the behavior and preferences of other users of surface waters.

n Acknowledgements

We thank M Wynia (U-Spatial, University of Minnesota) for assistance with route analysis, P Anderson and S Heisarky (Minnesota Pollution Control Agency) for data on Minnesota lake water quality, M Bezener (University of Minnesota) for assistance with statistical analysis, D Fisher (Stanford University) for assistance with Flickr photo processing, and M Rogers and J Johnson (University of Minnesota) for help with cost estimates. Funding for the Iowa Lakes Survey was provided by the Iowa Department of Natural Resources.

n References

- Banzhaf HS, Burtraw D, Evans D, and Krupnick A. 2006. Valuation of natural resource improvements in the Adirondacks. *Land Econ* **82**: 445–64.
- Brauman KA, Daily GC, Duarte TKE, and Mooney HA. 2007. The nature and value of ecosystem services: an overview highlighting hydrologic services. *Annu Rev Env Resour* **32**: 67–98.

Carson RT and Mitchell RC. 1993. The value of clean water: the

ble water quality. *Water Resour Res* **29**: 245–54. Egan KJ, Herriges JA, Kling CL, and Downing JA. 2009. Valuing water quality as a function of water quality measures. *Am J Agr Econ* **91**: 106–23.

public's willingness to pay for boatable, fishable, and swimma-

- Evans KS, Herriges JA, and Kling CL. 2011. Iowa Lakes Survey 2009. www.card.iastate.edu/lakes/iowa-lakes-report-2009.pdf. Viewed 25 Nov 2013.
- Feather P, Hellerstein D, and Tomasi T. 1995. A discrete-count model of recreational demand. *J Environ Econ Manag* **29**: 214–27.
- Freeman III AM, Herriges JA, and Kling CL. 2014. The measurement of environmental and resource values: theory and methods (3rd edn). Washington, DC: RFF Press.
- Griffiths C, Klemick H, Massey M, *et al*. 2012. Valuation of surface water quality improvements. *Rev Environ Econ Pol* **6**: 130–46.
- Hanemann WM. 1994. Valuing the environment through contingent valuation. *J Econ Perspect* **8**: 19–43.
- Ignite Social Media. 2012. Social network analysis report. www.ignitesocialmedia.com/social-media-stats/2012-socialnetwork-analysis-report/#Flickr. Viewed 24 Apr 2014.
- Iovanna R and Griffiths C. 2006. Clean water, ecological benefits, and benefits transfer: a work in progress at the US EPA. *Ecol Econ* **60**: 473–82.
- Johnston R, Besedin E, and Wardwell R. 2003. Modeling relationships between use and nonuse values for surface water quality: a meta-analysis. *Water Resour Res* **39**: 1363.
- Keeler BL, Polasky S, Brauman KA, *et al*. 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services. *P Natl Acad Sci USA* **109**: 18619–24.
- Kling C, Phaneuf D, and Zhao J. 2012. From Exxon to BP: has some number become better than no number? *J Econ Perspect* **26**: 3–26.
- Olmstead SM. 2010. The economics of water quality. *Rev Environ Econ Pol* **4**: 44–62.
- Parsons GR. 2003. The travel cost model. In: Champ PA, Boyle KJ, and Brown TC (Eds). A primer on nonmarket valuation. Dordrecht, The Netherlands: Kluwer.
- Phaneuf DJ and Smith VK. 2005. Recreation demand models. *Handbook Environ Econ* **2**: 671–761.
- Phaneuf DJ. 2002. A random utility model for total maximum daily loads: estimating the benefits of watershed-based ambient water quality improvements. *Water Resour Res* **38**: 1–11.
- Van Houtven G, Mansfield C, Phaneuf DJ, *et al*. 2014. Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. *Ecol Econ* **99**: 40–52.
- Viscusi WK, Huber J, and Bell J. 2008. The economic value of water quality. *Environ Resour Econ* **41**: 169–87.
- Watson S, McCauley E, and Downing JA. 1992. Sigmoid relationships between phosphorus, algal biomass, and algal community structure. *Can J Fish Aquat Sci* **49**: 2605–10.
- Wilson MA and Carpenter SR. 1999. Economic valuation of freshwater ecosystem services in the United States: 1971–1997. *Ecol Appl* **9**: 772–83.
- Wood SA, Guerry AD, Silver JM, and Lacayo M. 2013. Using social media to quantify nature-based tourism and recreation. *Scientific Reports* **3**: 2976.