**Long-Term Citizen-Collected Data Reveal Geographical Patterns and Temporal Trends in Lake Water Clarity**

The Faculty of Oregon State University has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DOI</td>
<td>10.1371/journal.pone.0095769</td>
</tr>
<tr>
<td>Publisher</td>
<td>Public Library of Science</td>
</tr>
<tr>
<td>Version</td>
<td>Version of Record</td>
</tr>
<tr>
<td>Terms of Use</td>
<td><a href="http://cdss.library.oregonstate.edu/sa-termsofuse">http://cdss.library.oregonstate.edu/sa-termsofuse</a></td>
</tr>
</tbody>
</table>
Long-Term Citizen-Collected Data Reveal Geographical Patterns and Temporal Trends in Lake Water Clarity

Noah R. Lottig1*, Tyler Wagner2, Emily Norton Henry3,4, Kendra Spence Cheruvelil3,5, Katherine E. Webster6, John A. Downing7, Craig A. Stow8

1 Center for Limnology Trout Lake Station, University of Wisconsin Madison, Boulder Junction, Wisconsin, United States of America, 2 U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, University Park, Pennsylvania, United States of America, 3 Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, United States of America, 4 Department of Outreach and Engagement, Oregon State University, Tillamook, Oregon, United States of America, 5 Lyman Briggs College, Michigan State University, East Lansing, Michigan, United States of America, 6 School of Natural Sciences, Trinity College, Dublin, Ireland, 7 Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, United States of America, 8 Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, Michigan, United States of America

Abstract

We compiled a lake-water clarity database using publically available, citizen volunteer observations made between 1938 and 2012 across eight states in the Upper Midwest, USA. Our objectives were to determine (1) whether temporal trends in lake-water clarity existed across this large geographic area and (2) whether trends were related to the lake-specific characteristics of latitude, lake size, or time period the lake was monitored. Our database consisted of >140,000 individual Secchi observations from 3,251 lakes that we summarized per lake-year, resulting in 21,020 summer averages. Using Bayesian hierarchical modeling, we found approximately a 1% per year increase in water clarity (quantified as Secchi depth) for the entire population of lakes. On an individual lake basis, 7% of lakes showed increased water clarity and 4% showed decreased clarity. Trend direction and strength were related to latitude and median sample date. Lakes in the southern part of our study-region had lower average annual summer water clarity, more negative long-term trends, and greater intra-annual variability in water clarity compared to northern lakes. Increasing trends were strongest for lakes with median sample dates earlier in the period of record (1938–2012). Our ability to identify specific mechanisms for these trends is currently hampered by the lack of a large, multi-thematic database of variables that drive water clarity (e.g., climate, land use/cover). Our results demonstrate, however, that citizen science can provide the critical monitoring data needed to address environmental questions at large spatial and long temporal scales. Collaborations among citizens, research scientists, and government agencies may be important for developing the data sources and analytical tools necessary to move toward an understanding of the factors influencing macro-scale patterns such as those shown here for lake water clarity.


Editor: Peter S. Petraitis, University of Pennsylvania, United States of America

Received January 21, 2014; Accepted March 29, 2014; Published April 30, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Funding for this research was provided by the National Science Foundation (MSB-1065786, EF-1065818, EF-1065649) and NTL-LTER (DEB-0822700). K.E.W. also thanks STRIVE grant 2011-W-FS-7 from the Environmental Protection Agency, Ireland for support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nrlottig@wisc.edu

Introduction

Macrosystems ecology has emerged as a new ecology sub-discipline aimed at understanding ecosystem patterns and processes resulting from broad-scale environmental changes such as land use and climate change. This sub-discipline studies multi-scaled patterns and processes in biological, geophysical, and social components, and their interactions with each other and processes operating at finer and coarser scales [1]. One of the approaches that macrosystems ecologists are taking is integrating large datasets from a variety of sources. There may be many inherent challenges in assembling appropriate datasets for analysis in a macrosystems ecology context [2], however, we feel that well-established citizen monitoring programs can play an important role in macrosystems ecology, especially those that collect data across large spatial and temporal extents with standardized measurement, data entry, and quality control protocols [3]. The Christmas Bird Count is a well-known example of a citizen-based program [http://birds.audubon.org/about-christmas-bird-count], long recognized as a successful model of how to integrate citizens into meaningful science and conservation [4,5].

In freshwater ecosystems, the indicator of water quality with the longest history of standardized measurement by scientists and citizens alike is Secchi depth [6-8]. The first transparency measurements were made in the ocean during the early 1800s, followed by more careful investigations by Pietro Angelo Secchi in the mid 1800s, which led to his recognition as the method’s founder (Fig. 1) [8]. Water clarity is easy and inexpensive to measure using a Secchi disk, and comparisons have shown that citizen-collected Secchi depth readings are nearly identical to those taken by professionals [9,10]. Consequently, the frequent, citizen-collected Secchi depth readings that are publically available online (e.g., Secchi Dip-In; http://www.secchidipin.org) often
span multiple decades for individual lakes and can be used to complement data collected by state and federal agencies.

The U.S. Clean Water Act of 1972 regulates anthropogenic inputs of pollutants to receiving waters [11]. Therefore, Secchi depth records from citizen monitoring programs provide a potentially valuable resource for examining broad-scale responses of lakes to nutrient control efforts instituted in the subsequent 40 years. However, there have been few attempts at compiling and using these data to concurrently analyze spatial and temporal trends at broad scales. Previous studies generally have focused at the scale of a single U.S. state [7,12,13].

We compiled citizen-collected, publically available Secchi depth measurements to answer two questions: (1) what are the long-term trends in lake-water clarity across a broad geographic region?; and (2) how do trends differ as a function of spatial location, lake size, and when Secchi depth records were collected? To our knowledge, this is the first study to look at long-term water clarity trends using (1) a broad spatial extent (eight states with a total land area greater than 1.1 million km² in the Upper Midwest region of the United States), (2) a long temporal window (1938–2012), (3) a large number of lakes (3,251), and (4) publically-available citizen-collected Secchi depth measurements (148,991). We answer these questions using Bayesian hierarchical modeling, an approach that is useful because it can incorporate disparate data from multiple sources to examine relationships across multiple spatial and temporal extents [14,15]. Our results can be used as a basis for forming hypotheses to test the mechanisms driving large-scale spatial patterns and temporal trends in water clarity as well as informing local, regional, or federal policies and management actions.

Methods

Data Acquisition

We obtained 239,741 citizen-collected Secchi depth measurements from monitoring networks in eight states in the Upper Midwest region of the United States (Fig. 2; Table 1) [16]. Observations spanned the time period from 1938–2012 with the oldest records from the state of Minnesota; the other citizen monitoring networks began collecting data during the early 1970s, and the majority (99%) of measurements were taken post-1972. The number of waterbodies monitored varied widely among states, ranging from 5 in Iowa to 1,507 in Wisconsin (Table 1).

Prior to analysis, Secchi data were put through a series of QA/QC checks, limited to a consistent seasonal window, and summarized as annual averages. First, we converted readings to meters and removed 31 Secchi depth observations that exceeded 30 m. Given the distribution of all Secchi observations (data not shown), these observations appeared to be unrealistic for the region and were likely the result of transcription errors. Second, we limited our analysis to Secchi observations collected during June through August, which removed a source of inter-seasonal variation [17] and generally was the season when most of the monitoring networks collected data (e.g., Secchi Dip-In), providing consistency across sampling networks. Finally, the response variable used for all analyses was average summer Secchi depth calculated for each lake-year, estimated using raw-untransformed Secchi depth values. We also calculated the median sampling date of each lake for use as a predictor variable. For example, for a lake that was continuously monitored by citizens from 1980 through 2000, the median date of the lake record was 1990. Therefore, median sample date represents a surrogate for the time period a given lake was sampled over the entire 1938–2012 span of this study.

Using the latitude, longitude, and unique lake identifiers provided by the individual sampling networks, we merged the study lakes with the high-resolution National Hydrography Dataset (NHD; http://nhd.usgs.gov). We then verified the identity and location (latitude and longitude) of lakes in the dataset and obtained lake surface areas (ha) from the NHD. The final dataset included 148,991 Secchi observations, from which we calculated 21,020 summer lake specific average Secchi depth values for 3,251 lakes that spanned monitoring record lengths from 1 to 34 years (Fig. 2).

Statistical Analysis

To estimate the overall temporal trend (i.e., trend across all lakes) and assess the variability among lakes in temporal trends, we first fit a hierarchical model with log-transformed Secchi depth as a linear function of year allowing the intercepts, slopes, and model error variances to differ among lakes. The form of the hierarchical model was as follows:

\[
y_j \sim N\left(\beta_{j(i)} + \beta_{j(j)}x_i, \sigma_j^2\right), \text{ for } i = 1, \ldots, n
\]

\[
\left(\beta_{j}\right) \sim N\left(\left(\begin{array}{c}
\mu_{\beta} \\
\sigma_{\beta}
\end{array}\right), \left(\begin{array}{cc}
\sigma_{\beta}^2 & \rho_{\beta} \sigma_{\beta} \sigma_{\beta}
\end{array}\right)\right), \text{ for } j = 1, \ldots, J
\]

\[
\log(\sigma_j) \sim N\left(\mu_{\sigma}, \sigma_{\sigma}^2\right), \text{ for } j = 1, \ldots, J
\]
Where $y_i$ is observation $i$ of loge-transformed Secchi depth from lake $j$; $a_j$ and $b_j$ are the intercept and slope (temporal trend) for lake $j$, respectively; $x$ is the year covariate (standardized; $(x_i - \bar{x})/SD(x_i)$) corresponding to the year Secchi observation $i$ was measured; $\mu_a$ is the population-average intercept and $\mu_b$ is the population-average slope; $\sigma_a^2$ and $\sigma_b^2$ are the variances of the intercepts and slopes, respectively; and $\rho_{\sigma_a\sigma_b}$ describes the covariance between $a_j$ and $b_j$, with $\rho$ describing the correlation between $a_j$ and $b_j$. We assumed a normal probability distribution for loge-transformed $s_j$ (lake-specific residual variances), with mean $\mu_s$ and variance $\sigma_s^2$. Non-informative normal priors were used for $\mu_a$, $\mu_b$, and $\mu_s$ (i.e., $\mu_x \sim N(0,1000)$). Non-informative uniform priors were used for $\sigma_a$, $\sigma_b$, $\omega_x$ (i.e., uniform distribution on $[0,100]$), and $\rho$ (i.e., uniform distribution on $[-1,1]$).

Secondly, we fit models that were similar to Eq. 1, but that included predictor variables (i.e., latitude, lake surface area, median sample date) to model variation among lake-specific intercepts, slopes, and residual standard deviations. For example, using latitude ($lat$) as a lake-specific predictor, the model was as follows:

$$y_i \sim N\left(a_{(i)} + b_{(j)} x_i, \sigma_j^2\right), \text{for } i = 1, \ldots n$$

$$\begin{pmatrix} a_j \\ b_j \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_{a0}^j + \gamma_{a1}^j \times lat_j \\ \gamma_{b0}^j + \gamma_{b1}^j \times lat_j \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho_{\sigma_a\sigma_b} \\ \rho_{\sigma_a\sigma_b} & \sigma_b^2 \end{pmatrix}\right), \text{for } j = 1, \ldots J$$

$$\log(\sigma_j) \sim N\left(\gamma_{\sigma0}^j + \gamma_{\sigma1}^j \times lat_j, \omega_s^2\right), \text{for } j = 1, \ldots J$$

(2)

Where $\gamma_{a0}^j$ and $\gamma_{a1}^j$ are intercepts and slopes, respectively. Non-informative normal priors were used for $\gamma_{a0}^j$ and $\gamma_{a1}^j$ and priors on other parameters were as described above. All covariates were standardized prior to analysis. Using an assessment method similar to classical significance testing, we examined 95% credible intervals (CRI}s) for $\gamma_{a1}^j$ to determine the importance of predictor variables (i.e., examined if 95% CRIs overlapped zero). The models were estimated using Bayesian estimation and the program JAGS was used for all analyses [18]. Three parallel chains were run with different initial values. After discarding the first 10,000 burn-in samples and thinning to every 500th iteration, we obtained the following results:

- For the intercept of $a_j$, the 95% credible interval was (-0.05, 0.01)
- For the slope of $a_j$, the 95% credible interval was (-0.02, -0.005)
- For the residual standard deviation of $\sigma_j$, the 95% credible interval was (0.01, 0.03)

Figure 2. Upper Midwest states in the US and Secchi record lengths (years) for all citizen monitored lakes included in the study. Size and color of symbol designate record lengths. See Table 1 for information on citizen monitoring networks by state.

doi:10.1371/journal.pone.0095769.g002

Trends and Patterns in Citizen Secchi Data
samples, we examined the scale reduction factor ($R$), a convergence statistic, for each parameter, trace plots, and plots of posterior distributions to assess convergence.

### Results

Average lake specific summer Secchi depth for the 3,251 study lake population covered a wide range from 0.2 to 16 m, with a sample mean of 2.4 m (median 2.1 m; Fig 3A). For a majority of lakes (75%) the year-to-year variability of Secchi depth values within a specific lake was small (coefficient of variation <30%). However, at the population-level, the range was large (coefficient of variation between 6.5 and 184%).

For the entire lake population, temporal trends in Secchi depth from 1938 to 2012 were positive (average of all lake specific trend slope (Fig. 4B) increased and inter-annual variability in Secchi depth, long-term Secchi depth trend, or inter-annual variability between 6.5 and 184%).

Numbers represent totals through 2012.
1. Number of lakes determined from 2007 National Hydrography Dataset (NHD) with surface area equal to or greater than 4 ha.
2. Includes all citizen Secchi observations available from designated monitoring network.

### Discussion

This study relied solely on citizen monitoring data to assess long-term water clarity (quantified as Secchi depth) trends and their relationships with spatial location, size of lake monitored, and time period monitoring took place. Overall, water clarity has been increasing slightly for the entire population of monitored lakes (average trend slope $[\mu_\text{trend}] > 0$). Eleven percent of lakes have strong trends in water clarity (90% probability slope is different than zero), with two-thirds corresponding to increased water clarity. We observed compelling spatial and temporal patterns with respect to the water clarity, long-term trends, and inter-annual variability of water clarity. Other studies have observed similar patterns in citizen monitored and remotely sensed Secchi depth [12,13] but over no more than two decades. We are unaware of any studies conducted with citizen-collected Secchi data at the spatial and temporal scales considered in this study.

Previous analyses of satellite-derived long-term Secchi depth patterns in the study region found that on average water clarity
Figure 3. Distribution of lake-specific overall average Secchi depths ($\delta_j$ from Eq. 1; A), long-term trends ($\beta_j$ from Eq. 1; B), and probability of lake specific long-term trend (probability of $\beta_j$ estimated in Eq. 1 of being either $<0$ or $>0$; C) based on a hierarchical Bayesian analysis. Long-term trend (B) positive values indicate long-term increases in water clarity (i.e., increased water clarity) and negative values indicate long-term declines in water clarity (i.e., decreased water clarity) for individual lakes. Probability of lake specific long-term trends (C) left of the dotted red line indicate the number of lakes with long-term annual declines in water clarity (Secchi depth becoming shallower) and associated probability of those long-term declines; values to the right indicate the number of lakes with long-term increases in annual water clarity (Secchi depth becoming deeper) and associated probability of those long-term increases.

DOI:10.1371/journal.pone.0095769.g003
not representative of all lakes within the study extent. Although a substantial number of water clarity measurements have been made, only a fraction of all lakes in each state had water clarity actively monitored by citizen scientists (Table 1). Additionally, the distribution of surface areas of citizen-monitored lakes diverged from that of the entire distribution of lakes in our study extent (Fig. 6). The lakes that citizens monitor are disproportionately larger than would be expected based on the size distribution of lakes in general [25]. Similar biases towards overrepresentation of large lakes have been observed when examining the types of lakes monitored by state agencies [26]. We believe that this is the largest dataset of citizen-monitored Secchi data to be analyzed to date and suspect this bias may also exist for datasets from beyond our study region. The patterns presented here are characteristic of the suite of lakes included in our dataset and, thus, they may not be representative of all lakes in our study region or those beyond it. Consequently, careful consideration needs to be made should these results be extrapolated to lake populations other than the those considered in this study [26]. Third, water clarity is affected by a number of natural features that alter light attenuation including dissolved organic carbon, chlorophyll, and suspended sediments [17] and is altered by human-induced nutrient inputs [27], invasion by non-native species (e.g., zebra mussels) [28], acidification [29] and climate change (browning of waters) [23]. The complex set of factors affecting lake water clarity, many of which are correlated, makes it difficult to interpret the causes of trends in water clarity and has, perhaps, led to Secchi depth being considered a relatively coarse indicator of overall water quality. However, its use as an approximation of water quality status [30] and as a measure of the reliability of satellite-derived water clarity data [13], combined with ease of collection by citizen scientists, suggests this 150-year measure of water clarity will continue to provide valuable insights into the condition of lake ecosystems.

Conclusions

Although there has been a long history of examining water clarity trends and patterns augmented with citizen collected data, the results presented here incorporate over 140,000 citizen monitoring water clarity values, extending back to 1938, collected at greater spatial and temporal scales than any previous study we are aware of. It is impossible for aquatic management organizations to monitor all lakes all the time; citizen water clarity data have been critical in augmenting traditional agency monitoring efforts and provide vital data for identifying and understanding water quality patterns at broad spatial and temporal scales (e.g.,
calibration of satellite-derived water clarity estimates). Here, we demonstrate the effectiveness of citizen data for quantifying long-term water clarity trends in individual lakes and identifying compelling population-level patterns in lake water clarity at the macro-scale. In an era of limited resources and during a push toward applying data-intensive science to environmental management and policy, the importance of citizen data such as these is likely to increase. Therefore, there is a clear need to continue to learn how to effectively leverage these types of data in the future.

These data needs coincide with the increasing recognition that stressors occurring at local, regional, and global scales interact to influence ecological patterns and processes from local to macroscales [24]. At the scale of an individual lake, our data indicate that a majority of lakes in the Upper Midwest region have had relatively stable long-term water clarity patterns. However at the regional scale, we observe emerging macro-scale patterns in long-term water clarity trends and inter-annual variability. Understanding the linkages between the multi-scale water clarity patterns presented here and the factors like climate, land use/cover, and biotic interactions that have been shown to influence water clarity [7,23,31] have been elusive. Although citizen science affords a unique opportunity to obtain data such as water clarity at unprecedented spatial and temporal scales, we generally lack the large, multi-thematic data necessary to explain these patterns and fully apply a macrosystems ecology approach. Consequently, there is a need to compile multi-thematic data across multiple spatial and temporal scales in order to link patterns observed at local and regional scales to processes occurring across the landscape. We
believe that collaborations among citizens, research scientists, and local, state and national agencies are important for developing the data sources and analytical tools necessary to move beyond identifying patterns toward an understanding of the critical factors influencing macro-scale patterns such as those shown here for lake water clarity.

Acknowledgments

We are grateful to the countless numbers of citizens who contributed Secchi data and the organizations that managed and provided us with data. It is only with the dedication of each citizen volunteer who diligently participated in lake monitoring at broad spatial extents that these collected observations, state coordinators who trained volunteers and maintained databases, and high profile programs that have stimulated participation in lake monitoring at broad spatial extents that these invaluable and unique long-term data records have become available. We would also like to acknowledge the helpful contributions and suggestions from the individuals that reviewed this manuscript. We also thank the Macrosystems CSI Limnology research group for their contributions to the effort, especially Ed Bissell for all his assistance managing the data and GIS analyses. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. GLERL contribution number 1703.

Author Contributions

Conceived and designed the experiments: NRL TW ENH KSC KEW JAD CAS. Analyzed the data: TW NRC. Wrote the paper: NRL TW ENH KSC KEW. Collect, compiled, and QAQGd data: NRL JAD ENH. Conducted Bayesian hierarchical modeling: TW. Created tables and figures: NRL TW. Contributed critical reviews and edited manuscript prior to submission: NRL TW ENH KSC KEW JAD CAS.

References