Color, chlorophyll $a$, and suspended solids effects on Secchi depth in lakes: implications for trophic state assessment

Patrick L. Brezonik,1,5 R. William Bouchard Jr.,2 Jacques C. Finlay,3 Claire G. Griffin,3 Leif G. Olmanson,4 Jesse P. Anderson,2 William A. Arnold,1 and Raymond Hozalski1

1Department of Civil, Environmental, and Geo Engineering, University of Minnesota, Minneapolis, Minnesota 55455 USA
2Environmental Analysis and Outcomes Division, Minnesota Pollution Control Agency, 520 Lafayette Road, Saint Paul, Minnesota 55155-4194 USA
3Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, Minnesota 55455 USA
4Department of Forest Resources, University of Minnesota, Saint Paul, Minnesota 55455 USA


Abstract. Secchi depth (SD), a primary metric to assess trophic state, is controlled in many lakes by algal densities, measured as chlorophyll-$a$ (chl-$a$) concentration. Two other optically related water quality variables also directly affect SD: non-algal suspended solids (SS$_{Na}$) and colored dissolved organic matter (CDOM), expressed as the absorption coefficient at 440 nm, $a_{440}$. Using a database of ~1,460 samples from ~625 inland lake basins in Minnesota and two other Upper Midwest states, Wisconsin and Michigan, we analyzed relationships among these variables, with special focus on CDOM levels that influence SD values and the Minnesota SD standards used to assess eutrophication impairment of lakes. Log-transformed chl-$a$, total suspended solids (TSS), and SD were strongly correlated with each other; log($a_{440}$) had major effects on log(SD) but was only weakly correlated with log(chl-$a$) and log(TSS). Multiple regression models for log(SD) and 1/SD based on the three driving variables (chl-$a$, SS$_{Na}$, and CDOM) explained ~80% of the variance in SD in the whole data set, but substantial differences in the form of the best-fit relationships were found between major ecoregions. High chl-$a$ concentrations (>50 µg/L) and TSS (>20 mg/L) rarely occurred in lakes with high CDOM ($a_{440} > 4$ m$^{-1}$), and all lakes with $a_{440} > 8$ m$^{-1}$ had SD ≤2.0 m despite low chl-$a$ values (<10 µg/L) in most lakes. Further statistical analyses revealed that CDOM has significant effects on SD at $a_{440}$ values > 4 m$^{-1}$. Thus, SD is not an accurate trophic state metric in moderately to highly colored lakes, and Minnesota’s 2-m SD criterion should not be the sole metric to assess eutrophication impairment in warm/cool-water lakes of the Northern Lakes and Forest ecoregion. More generally, trophic state assessments using SD in regions with large landscape sources of CDOM need to account for effects of CDOM on SD.

Key words: chlorophyll $a$; colored dissolved organic matter; dissolved colored organic matter; ecoregion; lakes; Secchi depth; total suspended solids; trophic state; Upper Midwest.

INTRODUCTION

Secchi depth (SD), the most common indicator of lake water clarity and quality, has long played an important role in defining lake trophic state (Nürnberg 1996, Heiskary and Wilson 2008, Lottig et al. 2014). Along with chlorophyll and total phosphorus, SD is one of the three metrics in Carlson’s (1977) trophic state indices. The State of Minnesota uses SD in numeric standards to determine whether or not a water body meets the water quality conditions that support its designated beneficial uses, e.g., aquatic life and recreational uses such as swimming (Heiskary and Wilson 2008). SD is used extensively by citizen monitoring programs to track trends in lake trophic status, owing to its simplicity and low measurement costs (e.g., Lottig et al. 2014, Heiskary and Egge 2016), and it also is useful to estimate other optical properties, such as the diffuse attenuation coefficient for photosynthetically active radiation, $K_{PAR}$ (Lee et al. 2018).

As an integrative measure of water clarity, SD is determined primarily by three variables. Algal biomass, usually measured as chlorophyll $a$ (chl-$a$), has a major influence of SD, and SD is often used as a proxy for algal levels. A second factor is non-algal suspended solids (SS$_{Na}$), including clays and other suspended minerals, that often are affected by storm events, particularly in rivers and reservoirs. The third is colored
dissolved organic matter (CDOM), which is composed largely of humic substances. In many lakes affected by eutrophication, algal biomass (chl-a) is the primary determinant of SD. Several studies have reported hyperbolic relationships between SD and chl-a and linear relationships between log SD and log chl-a (e.g., Carlson 1977). SSNA can be a key determinant of water clarity in surface waters where watershed conditions promote soil erosion but is not commonly a major factor in natural recreational lakes of the Upper Midwest (UMW) states. Colored dissolved organic matter similarly is not important in many lakes, but high levels are common in regions dominated by forests and wetlands, such as the Northern Lakes and Forests (NLF) ecoregion (Fig. 1) in northeastern Minnesota, northern Wisconsin, Upper Michigan and the northern half of Lower Michigan (Griffin et al. 2018) and in similar ecoregions in the northeastern and southeastern U.S. (Omernik 1987, Omernik and Griffith 2014).

Colored dissolved organic matter levels are increasing in some temperate and boreal regions, e.g., Scandinavia (Haaland et al. 2010), for reasons still not fully understood. Lakes where CDOM affects SD levels thus may become more common in the future as a result of increasing precipitation (Leech et al. 2018), declining acidification (Monteith et al. 2007), or both. Evidence for the widespread occurrence of this “browning” phenomenon in the Upper Midwest is inconclusive (Brezonik et al. 2015), however, and several studies have shown that CDOM temporal trends in individual lakes are not monotonic but driven by variations in climate and hydrologic conditions (Pace and Cole 2002, Jane et al. 2017, Carpenter and Pace 2018, Corman et al. 2018, Leech et al. 2018).

Limnologists have long recognized that SD can be controlled by CDOM. Based on data from 470 northeastern Wisconsin lakes, Juday and Birge (1933) concluded that color had more important effects on lake SD than plankton did, and they found an inverse hyperbolic relationship between SD and lake color. Brezonik (1978) found a linear relationship between inverse SD (1/SD) and color using in situ mesocosms to which a concentrated source of humic color was added. Data from a Florida lake survey also showed a strong regression relationship between 1/SD and color and turbidity (Brezonik 1978). These studies showed that CDOM strongly influences SD at moderate to high levels of CDOM and low levels of algal biomass.

Less well known are the relationships between CDOM, SD, chl-a, and total suspended solids (TSS) in lakes with higher and more variable algal biomass and mineral turbidity. When both CDOM and algae co-occur at levels that affect SD, failure to consider the influence of CDOM will bias interpretations of lake impairment and trophic state based solely on SD. Consequently, the U.S. EPA (2000) recommended that development of water quality standards to assess lake eutrophication should consider CDOM as a potential confounding factor of trophic status measurements.

![Map of Upper Midwest states showing Minnesota ecoregions delineated by Omernik and Griffith (2014). Database includes lakes from all ecoregions except for DLA, where only a few small lakes occur.](image-url)
Despite the now broadly recognized importance of CDOM in lake ecosystems, e.g., the nutrient-color paradigm (Williamson et al. 1999, 2014, Webster et al. 2008), its quantitative influence on SD is poorly known, partly because CDOM is measured much less often than SD and other trophic indicators (chl-\(\alpha\) and total phosphorus). A recent study on coastal waters suggested that CDOM should be routinely monitored to help interpret water clarity monitoring (Harvey et al. 2015), but CDOM impacts have not yet been formally integrated with lake assessments. Here we use a large database on lakes of the UMW to examine relationships among four common optical water quality parameters, chl-\(\alpha\), TSS, CDOM, and SD, and evaluate the CDOM levels that affect interpretation of SD as a trophic state metric.

**Methods**

*Study area and data sources*

This study focused on the large, lake-rich NLF ecoregion that extends across the UMW states of Minnesota, Wisconsin, and Michigan (Fig. 1) and comparisons with two ecoregions to its south: the North Central Hardwood Forest (NCHF) and Western Corn Belt Plains (WCBP; Omernik and Griffith 2014). The NLF is heavily forested (approximately 50%, mixed conifers and hardwoods). About one-third of its area is wetlands and lakes (Homer et al. 2015), and the high proportion of forest and wetlands in the NLF leads to an abundance of lakes with high CDOM. The ecoregion has little urban and agricultural land (4% and 7%, respectively). In contrast, nearly one-half (48%) of the NCHF is used for agriculture and 9% is urban; forests account for one-quarter of the ecoregion, and wetlands constitute 10% of land cover. Lakes are abundant throughout the NCHF, but high CDOM levels are uncommon (Griffin et al. 2018). The WCBP, southernmost ecoregion in Minnesota, is dominated by agricultural land (83%) and has only ~4% forest. It tends to have higher chlorophyll and smaller SD values than the NLF and NCHF.

We have been studying CDOM and mapping its abundance in UMW lakes since 2014 using a combination of field campaigns and satellite imagery (e.g., Brezonik et al. 2015, Olmanson et al. 2016, Griffin et al. 2018). Ground-based sampling in 2014–2015 was focused in the NLF and NCHF in northern Minnesota (Fig. 1A). In 2016 sampling was expanded to include NLF and NCHF portions that extend across Wisconsin and Michigan, as well as the Northern Minnesota Wetlands (NMW), an ecoregion in north-central Minnesota that has only ~100 lakes, very few of which are road accessible (Fig. 1B). Sample collection in 2017 was extended to the WCBP and other ecoregions in central, western, and southern Minnesota.

The Minnesota Pollution Control Agency (MPCA) routinely monitors ~150 lakes across the state each year for water quality assessments. Since 2015 they have included CDOM in their measurements. We combined the 2014–2017 UMN data from ground-based sampling (708 site-date measurement sets) with 754 sets of 2015–2017 measurements on ground-based samples by the MPCA to produce a data set of 1,462 site-date measurements with little overlap between the two data sources. Many lakes were sampled more than once in both studies, and all observations (site-date combinations) were treated separately; i.e., multiple samples from a lake were not averaged. The final data set includes data from 251 MPCA lake basins and 382 UMN basins.

*Sampling and analysis methods*

Sampling procedures and field and laboratory analyses followed standard limnological practices. Detailed methods are described elsewhere (Egge et al. 2018, Griffin et al. 2018). In brief, UMN water samples were collected from ~0.25 m below the lake surface, and the MPCA collected a 0–2 m integrated sample of the epilimnion. Samples were stored in acid-washed and triple-rinsed polycarbonate or high-density polyethylene bottles and filtered for chl-\(\alpha\) and dissolved constituents within 24 h of collection. Chl-\(\alpha\) was filtered from water with 0.22 \(\mu\)m cellulose nitrate filters (0.45 \(\mu\)m glass fiber filters for MPCA) and stored frozen until analysis by fluorometry after 90% acetone extraction. Total suspended solids (TSS) was measured as the additional dry weight after filtration and drying at 105°C, normalized by volume. Water for CDOM analysis was filtered through 0.45 \(\mu\)m Geotech High Capacity filters and stored in the dark at 4°C in pre-ashed 40-mL amber glass bottles until analysis within 1 month of collection. Absorbance at 440 nm, measured using a Shimadzu (Columbia, Maryland, USA) 1601UV-PC dual beam spectrophotometer through 1- or 5-cm quartz cuvettes against a nanopure water blank, was converted to Napierian absorption coefficients (Kirk 1994) using:

\[
a_{440} = \frac{2.303 A_{440}}{l}
\]

where \(a_{440}\) is the absorption coefficient at 440 nm, \(A_{440}\) is absorbance at 440 nm, and \(l\) is cell path length (m). Absorbance scans were blank-corrected before conversion. CDOM values are reported as \(a_{440}\).

*Statistical analyses*

The data were assembled into an Excel 2016 spreadsheet. Distributional statistics and principal components analysis based on inter-parameter correlation coefficients were analyzed using JMP Pro 13.1 (SAS Institute, Inc., Cary, North Carolina, USA). Generalized simple and multiple regression analyses were conducted using the Akaike information criterion (AIC) to select the best models, and least-squares regressions then were run on these models. Concentrations of non-algal suspended solids (SSNA) and their log transforms were estimated as the chl-\(\alpha\) detrended values of TSS, calculated as the
Physically based predictive equation for SD

A physically based relationship between SD and its controlling constituents (chl-a, CDOM, and TSS) is derived as follows for use in evaluating the effects of the three controlling variables on SD. Lorenzen (1980) and Megard et al. (1980) demonstrated that SD can be expressed in terms of an equation of the form

$$SD = \frac{\ln I_0/I_z}{k_{NA} + k_A^* \text{chl-a}}$$

(2)

where SD is the depth (in m) at which the Secchi disk disappears, $I_0$ is the light intensity at the depth of disappearance, $I_z$ is the light intensity at the water surface, $k_{NA}$ is the light attenuation coefficient for all non-algal constituents causing either light absorption or scattering, and $k_A^*$ is the light attenuation coefficient caused by algal biomass, expressed as chl-a concentration. The usual assumption is that $I_z = 0.1 I_0$ at the depth of Secchi disk disappearance (Brezonik 1978, Megard et al. 1980). The term $\ln(I_0/I_z)$ thus becomes a constant, 2.303, that is subsumed into the attenuation coefficients, and Eq. 2 becomes

$$1/SD = k_{NA} + k_A^* \text{chl-a}$$

(3)

where $k_{NA} = k_{NA}^*/2.303$ and $k_A^* = k_A^*/2.303$.

The coefficient for light attenuation by non-algal constituents, $k_{NA}^*$, is a composite coefficient that can be expanded into terms for the specific constituents causing light attenuation; e.g., for most natural waters

$$k_{NA} = k_w + k_{NAP} \times \text{NAP} + k_{CDOM} \times a_{440}$$

(4)

where $k_w$ is the light attenuation coefficient for water itself; $k_{NAP}$ is the light attenuation coefficient for non-algal particles (clays, other minerals, and organic particles not derived from or associated with algal activity); NAP is the concentration (mg/L) of these particles; $k_{CDOM}$ is the attenuation coefficient for light absorption by CDOM; and $a_{440}$ is the measure of CDOM. The attenuation coefficient for water itself is small compared to the other factors and not important in most lakes of interest here. Exceptions in Minnesota might be Lake Superior and the deep, ultra-oligotrophic lakes found in abandoned iron mine pits of the Mesabi Range, where SD values of 15 m or more can occur. For simplicity, we assumed $k_w$ is negligible in the following analysis. NAP was calculated as the chl-a detrended TSS, i.e., SSNA. Eq. 3 thus becomes

$$1/SD = k_A^* \times \text{chl-a} + k_{NAP} \times \text{SSNA} + k_{CDOM} \times a_{440}$$

(5)

Eq. 5 was used to evaluate the effects of chl-a, SSNA, and $a_{440}$ on SD in UMW lakes.
A principal components analysis on log-transformed values of SD, chl-\(a\), TSS, and \(a_{440}\) showed that TSS and chl-\(a\) had similar and high loadings on the first principal component (PC1; Fig. 3). SD also had a high loading on PC1 but in a negative direction. PC1 thus can be viewed essentially as a composite trophic state variable. In contrast, PC2 had high loading from \(a_{440}\) and much lower loadings from the other three variables; PC2 thus appears to represent a (humic) color component that is largely orthogonal to the trophic state component. The essentially orthogonal relationship between TSS and \(a_{440}\) in the principal components analysis (Fig. 3) is similar to that found by Olson et al. (2018) between turbidity (an optical property related to TSS) and \(a_{440}\) in glacially fed alpine lakes of the Canadian Rocky Mountains.

Given the above findings, the negative linear relationship between log(SD) and log(chl-\(a\)) for all the data (Fig. 4A) was expected, but the fit (\(R^2 = 0.61\)) was lower than others have reported for similar relationships (e.g., \(R^2 = 0.86\), Carlson 1977; \(R^2 = 0.82\) and 0.69; Nürnberg 1996). Many outliers in Fig. 4A are below the line of best fit; in these cases, SD was less than expected based on chl-\(a\) concentration, implying that some other factor(s) also affected SD. Removal of samples with \(a_{440} > 3.0 \text{ m}^{-1}\), which was found to be a limiting value for CDOM domination by allochthonous sources (Griffin et al. 2018), improved the fit to \(R^2 = 0.76\). A few sites, mostly with high TSS, however, still fell far from the regression line (Fig. 4B).

The three highly correlated variables, chl-\(a\), TSS, and SD, all had complicated relationships with CDOM (\(a_{440}\)). High values of the first two (Fig. 5A, B) were essentially orthogonal to the \(a_{440}\) distribution, with chl-\(a\) concentrations > \(50 \mu g/L\) and TSS > 20 mg/L occupying narrow ranges of CDOM (\(a_{440}\) generally < 3.5 \text{ m}^{-1}\) and 4.5 \text{ m}^{-1}, respectively), supporting the nutrient-color paradigm (Williamson et al. 1999). Of the 203 samples with both measured \(a_{440}\) and chl-\(a\) values, only two samples with \(a_{440} > 5 \text{ m}^{-1}\) had chl-\(a\) > 50 \(\mu g/L\). One value (57 \(\mu g/L\); Fox Lake, Minnesota, USA) was from a shallow, bog-stained NLF lake with both wetlands and agricultural activity in its riparian zone. The other value (64 \(\mu g/L\), shallow, wetland-dominated Turner Lake, near
Brainerd, Minnesota, USA) is an anomaly insofar as the next highest chl-a value in two seasons of monthly sampling was only 27 \( \mu \text{g/L} \). Similarly, of the 192 samples with both measured \( a_{440} \) and TSS, only two with \( a_{440} > 5 \text{ m}^{-1} \) had TSS > 15 mg/L. Both were from sites in the St. Louis River Estuary of Lake Superior that are influenced by runoff from the Pokegama River, which drains a region with highly erodible clay soils (Roesler et al. 2018).

The distribution of SD vs. \( a_{440} \) was somewhat broader than that of chl-a or TSS (Fig. 5C), but almost no SD values > 3.0 m occurred for \( a_{440} > 3 \text{ m}^{-1} \), and no SD values > 2 m were found for \( a_{440} > 8 \text{ m}^{-1} \) (Fig. 5D). High CDOM levels apparently are antithetical to production of high levels of algal biomass and related organic suspended solids in UMW lakes, as Thrane et al. (2014) found for boreal lakes. Despite the low chl-a concentrations associated with high-CDOM waters, their SD values are small because of light absorption by CDOM. Similarly, catchment factors that promote high export of CDOM into UMW lakes, such as wetlands and other poorly drained landscape features, apparently are not favorable for high export of mineral or non-algal organic suspended solids into lakes.

**SD predictive relationships**

We evaluated relationships between SD and its controlling variables singly and in combination using log-transformed values for the whole data set and separately for the NLF, NCHF, and WCBP ecoregions using generalized regression analysis and the AIC to select the best models (Table 3). For the whole database, the best-fit relationship (\( R^2 = 0.80 \)) included all three predictor variables, but a two-variable relationship with chl-a and \( a_{440} \) was found to be best (\( R^2 = 0.76 \)) for the NLF. Although the three-variable model had slightly higher \( R^2 \), addition of TSS added little explanatory power and increased the AIC, probably because TSS concentrations in this ecoregion are low and generally associated with chl-a. In contrast, a two-variable model using chl-a and TSS was the

**Table 1. Statistical summaries of optical water quality variables for all data and for data separated by ecoregion.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( a_{440} ) (m&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Chl-a (( \mu \text{g/L} ))</th>
<th>SD (m)</th>
<th>TSS (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.68</td>
<td>16.74</td>
<td>2.31</td>
<td>7.26</td>
</tr>
<tr>
<td>Median</td>
<td>1.31</td>
<td>6.33</td>
<td>1.80</td>
<td>3.20</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5.45</td>
<td>36.28</td>
<td>1.72</td>
<td>12.94</td>
</tr>
<tr>
<td>SEE</td>
<td>0.16</td>
<td>1.06</td>
<td>0.05</td>
<td>0.41</td>
</tr>
<tr>
<td>Maximum</td>
<td>32.5</td>
<td>721</td>
<td>19.5</td>
<td>120</td>
</tr>
<tr>
<td>75th percentile</td>
<td>4.01</td>
<td>15.36</td>
<td>3.3</td>
<td>6.58</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0.67</td>
<td>3.14</td>
<td>1.0</td>
<td>2.00</td>
</tr>
<tr>
<td>N</td>
<td>1,193</td>
<td>1,177</td>
<td>1,238</td>
<td>991</td>
</tr>
<tr>
<td>NLF ecoregion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.79</td>
<td>8.6</td>
<td>2.60</td>
<td>3.77</td>
</tr>
<tr>
<td>Median</td>
<td>1.93</td>
<td>5.1</td>
<td>2.13</td>
<td>2.80</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>6.22</td>
<td>10.6</td>
<td>1.76</td>
<td>4.67</td>
</tr>
<tr>
<td>SEE</td>
<td>0.22</td>
<td>0.4</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td>Maximum</td>
<td>32.5</td>
<td>124.8</td>
<td>19.5</td>
<td>93.4</td>
</tr>
<tr>
<td>75th percentile</td>
<td>6.7</td>
<td>10.0</td>
<td>3.6</td>
<td>4.4</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0.7</td>
<td>2.8</td>
<td>1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>N</td>
<td>793</td>
<td>749</td>
<td>823</td>
<td>655</td>
</tr>
<tr>
<td>NCHF ecoregion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.23</td>
<td>26.11</td>
<td>2.12</td>
<td>8.77</td>
</tr>
<tr>
<td>Median</td>
<td>0.97</td>
<td>8.19</td>
<td>1.57</td>
<td>4.80</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.00</td>
<td>29.25</td>
<td>1.65</td>
<td>14.09</td>
</tr>
<tr>
<td>SEE</td>
<td>0.06</td>
<td>1.76</td>
<td>0.1</td>
<td>0.98</td>
</tr>
<tr>
<td>Maximum</td>
<td>8.27</td>
<td>264</td>
<td>7.4</td>
<td>120</td>
</tr>
<tr>
<td>75th percentile</td>
<td>1.51</td>
<td>25.4</td>
<td>2.8</td>
<td>9.2</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0.57</td>
<td>3.7</td>
<td>0.9</td>
<td>2.4</td>
</tr>
<tr>
<td>N</td>
<td>268</td>
<td>275</td>
<td>248</td>
<td>205</td>
</tr>
<tr>
<td>WCBP ecoregion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.48</td>
<td>73.9</td>
<td>0.77</td>
<td>29.9</td>
</tr>
<tr>
<td>Median</td>
<td>1.20</td>
<td>42.3</td>
<td>0.52</td>
<td>21.5</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.88</td>
<td>97.5</td>
<td>0.64</td>
<td>24.6</td>
</tr>
<tr>
<td>SEE</td>
<td>0.10</td>
<td>105.2</td>
<td>0.07</td>
<td>2.8</td>
</tr>
<tr>
<td>Maximum</td>
<td>4.42</td>
<td>721</td>
<td>3.4</td>
<td>111</td>
</tr>
<tr>
<td>75th percentile</td>
<td>2.13</td>
<td>111</td>
<td>1.1</td>
<td>39.2</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0.84</td>
<td>17.3</td>
<td>0.3</td>
<td>11.2</td>
</tr>
<tr>
<td>N</td>
<td>73</td>
<td>86</td>
<td>92</td>
<td>76</td>
</tr>
</tbody>
</table>

**Note:** Ecoregions are NLF, northern lakes and forests; NCHF, north central hardwood forest; and WCBP, western corn belt plains. SEE, standard error of estimate.

**Table 2. Correlation matrix of Pearson r values for the four water quality variables.**

<table>
<thead>
<tr>
<th></th>
<th>( \log(a_{440}) )</th>
<th>( \log(\text{chl-a}) )</th>
<th>( \log(\text{SD}) )</th>
<th>( \log(\text{TSS}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(a_{440}) )</td>
<td>1.000</td>
<td>0.253</td>
<td>-0.528</td>
<td>0.147</td>
</tr>
<tr>
<td>( \log(\text{chl-a}) )</td>
<td>1.000</td>
<td>-0.761</td>
<td>0.778</td>
<td></td>
</tr>
<tr>
<td>( \log(\text{SD}) )</td>
<td>1.000</td>
<td>-0.759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(\text{TSS}) )</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Brainerd, Minnesota, USA** is an anomaly insofar as the next highest chl-a value in two seasons of monthly sampling was only 27 \( \mu \text{g/L} \). Similarly, of the 192 samples with both measured \( a_{440} \) and TSS, only two with \( a_{440} > 5 \text{ m}^{-1} \) had TSS > 15 mg/L. Both were from sites in the St. Louis River Estuary of Lake Superior that are influenced by runoff from the Pokegama River, which drains a region with highly erodible clay soils (Roesler et al. 2018).

The distribution of SD vs. \( a_{440} \) was somewhat broader than that of chl-a or TSS (Fig. 5C), but almost no SD values > 3.0 m occurred for \( a_{440} > 3 \text{ m}^{-1} \), and no SD values > 2 m were found for \( a_{440} > 8 \text{ m}^{-1} \) (Fig. 5D). High CDOM levels apparently are antithetical to production of high levels of algal biomass and related organic suspended solids in UMW lakes, as Thrane et al. (2014) found for boreal lakes. Despite the low chl-a concentrations associated with high-CDOM waters, their SD values are small because of light absorption by CDOM. Similarly, catchment factors that promote high export of CDOM into UMW lakes, such as wetlands and other poorly drained landscape features, apparently are not favorable for high export of mineral or non-algal organic suspended solids into lakes.

**SD predictive relationships**

We evaluated relationships between SD and its controlling variables singly and in combination using log-transformed values for the whole data set and separately for the NLF, NCHF, and WCBP ecoregions using generalized regression analysis and the AIC to select the best models (Table 3). For the whole database, the best-fit relationship (\( R^2 = 0.80 \)) included all three predictor variables, but a two-variable relationship with chl-a and \( a_{440} \) was found to be best (\( R^2 = 0.76 \)) for the NLF. Although the three-variable model had slightly higher \( R^2 \), addition of TSS added little explanatory power and increased the AIC, probably because TSS concentrations in this ecoregion are low and generally associated with chl-a. In contrast, a two-variable model using chl-a and TSS was the
best predictive model for the NCHF, and CDOM was not an important predictor of SD in the NCHF owing to the relative homogeneity of lake color in the ecoregion. More ambiguous results were found for the WCBP, with two two-variable models yielding similar $R^2$ and AIC values (Table 3). $R^2$ values for both models were lower than those for the best models for the whole database and NLF and NCHF ecoregions, reflecting more compressed variable ranges in this nutrient-rich ecoregion (Table 1).

In terms of simple bivariate relationships, log(TSS) and log(chl-$a$) had nearly the same predictive values of log(SD) for the whole database (Table 3), which likely reflects the important role of algae as a TSS source in lakes of this study. In contrast, log(SSNA) was a poor predictor ($R^2 = 0.08$), supporting the idea that SSNA is not an important control on SD in most UMW lakes. Log($a_{440}$) by itself also was a poor predictor of log(SD) for the whole database, no doubt because of the L-shaped distribution of the SD-$a_{440}$ relationship (Fig. 5C).

Because of the strong correlation between chl-$a$ and TSS, we were concerned that including both variables in the regression could “double-count” the influence of chl-$a$. A three-term SD regression in which TSS was replaced by SSNA gave the same fit ($R^2$ and RMSE), but increased the equation coefficient for log(chl-$a$) (Table 3). The RMSE for the three-term log(SD) equation (0.80, Table 3), the RMSE (0.54) translates to larger uncertainty in SD (1.85 m). Moreover, the large intercept value implies that the equation applies only to relatively low SD values; an SD > 2.8 m would require negative values for one or more of the predictor variables, which clearly is not possible.

Reasons for the limitations of the 1/SD equation are found in the predictor variable distributions (Fig. 5A–C). High chl-$a$ and TSS concentrations occurred nearly exclusively at low $a_{440}$, and these high values caused low SD (Fig. 4, Table 2); low SD is equivalent to high 1/SD. High chl-$a$ and TSS at some sites with low $a_{440}$ thus promoted high 1/SD values, which distorted the 1/SD vs. $a_{440}$ relationship, effectively decreasing the slope of the best-fit line between 1/SD and $a_{440}$. This distortion can be seen in a plot of 1/SD vs. $a_{440}$ (Fig. 6A), which includes only sites with chl-$a$ < 10 µg/L. High 1/SD values (~1 m$^{-1}$) for sites with $a_{440}$ < 3 m$^{-1}$ were associated with high TSS (typically 15–20 mg/L). Given the low chl-$a$ concentrations at these sites, the TSS likely was mostly SSNA (e.g., clay minerals or non-algal organic matter from allochthonous sources or macrophytes). These high values pulled the regression line upward at low $a_{440}$ and decreased the slope and fit ($R^2$) of the relationship, indicating that the relationship should not be used to quantify CDOM effects on SD. A plot of 1/SD vs. $a_{440}$ for sites with chl-$a$ < 10 µg/L and $a_{440}$ > 3 m$^{-1}$ effectively eliminated the sites affected by high SSNA (Fig. 6B).

**DISCUSSION**

**Relationships of SD with other optical variables**

SD often is used as a surrogate for estimating algal biomass and trophic state and, in many lakes, SD alone...
is a good estimator of trophic state (Carlson 1977, Nürnberg 1996, Heiskary and Wilson 2008). SD is also a good measure for evaluating recreational suitability, as users often respond to water clarity when deciding whether a water body is suitable for recreation (Smeltzer and Heiskary 1990, Heiskary and Wilson 2005). For these reasons and the widespread availability of data, the U.S. EPA recommended the use of SD to develop lake and reservoir nutrient standards (U.S. EPA 2000).

SD thus is now widely accepted for trophic state assessment, but our results indicate that it cannot be applied uniformly. Moreover, there is a long history of studies demonstrating that algal production (chl-a concentration) does not always control SD. Juday and Birge (1933) found an inverse curvilinear relationship between SD and lake color similar to the trend in our SD-$a_{440}$ data (Fig. 5C). They obtained the relationship by dividing data on 470 lakes from northeastern Wisconsin into 11 groups with SD ranges of 0.9 m to 7.5–9.4 m and plotting mean SD for the groups vs. mean color determined visually and reported in platinum-cobalt units (PCU) for the lakes in each group. Differences in measurement methods and reporting units between our results and those of Juday and Birge make quantitative comparisons difficult, but the inverse curvilinear trends in the two data sets are similar.

Brezonik (1978) found close-fitting, straight-line relationships between 1/SD and color in experiments where a concentrated humic color source was added to mesocosm enclosures in two Florida lakes. Color was determined colorimetrically and reported in PCU. The slope of the 1/SD-color (PCU) relationship was 0.0040, which translated to a slope of $\sim 0.074$ when the absorption coefficient of chloroplatinate at 440 nm was used to convert color in PCU to $a_{440}$ in m$^{-1}$. This is similar to the slope (0.069) for the 1/SD vs. $a_{440}$ relationship of colored lakes ($a_{440} > 3$ m$^{-1}$) in Fig. 6B but lower than that for the three-variable model based on all UMW data (first
Bivariate relationships

Brezenik (1972), Brezonik (1978) found a strong relationship between log SD or 1/SD and optical properties of water, which raised the regression line at low color and turbidity. A regression for light attenuation by all suspended particles, both algal and non-algal, and color accounts for light attenuation for light-absorbing CDOM. A regression of 1/SD vs. color and chl-α was not as strong ($R^2 = 0.63$). Canfield and Hodgson (1983) reported an $R^2$ of 0.79 for a regression of ln(SD) vs. ln(chl-α) and ln(color) based on data from 165 Florida lakes. Nürnberg (1996) found an $R^2$ of 0.88 for a regression of log(SD) vs. log(chl-α) and log(color) for 33 lakes from northeastern North America and an $R^2$ of 0.79 for another data set of 91 lakes worldwide.

Although the three-term regression models (both for log(SD) and 1/SD) explained a high proportion (~80%) of the variance in SD for the whole database, they still did not account for ~20% of the SD variance, and the models had high RMSE values (1.45–1.85 m$^{-1}$) that limit

equation in Table 3). As described above, the latter was influenced by high TSS and chl-α values at low $a_{440}$, which raised the regression line at low color and decreased the slope of the relationship.

Several previous studies also have reported regression relationships between log SD or 1/SD and optical water quality variables. Using data from a survey on 55 lakes in north and central Florida (Shannon and Brezonik 1972), Brezonik (1978) found a strong relationship ($R^2 = 0.89$) between 1/SD and color and turbidity: $1/SD = 0.106 + 0.128(T) + 0.0025(C)$, where $T$ was laboratory-measured turbidity in standard formazin (nephelometric) turbidity units and $C$ was color measured in PCU by colorimetry. Turbidity accounts for light attenuation by all suspended particles, both algal and non-algal, and color accounts for light attenuation for light-absorbing CDOM. A regression of 1/SD vs. color and chl-α was not as strong ($R^2 = 0.63$). Canfield and Hodgson (1983) reported an $R^2$ of 0.79 for a regression of ln(SD) vs. ln(chl-α) and ln(color) based on data from 165 Florida lakes. Nürnberg (1996) found an $R^2$ of 0.88 for a regression of log(SD) vs. log(chl-α) and log(color) for 33 lakes from northeastern North America and an $R^2$ of 0.79 for another data set of 91 lakes worldwide.

Although the three-term regression models (both for log(SD) and 1/SD) explained a high proportion (~80%) of the variance in SD for the whole database, they still did not account for ~20% of the SD variance, and the models had high RMSE values (1.45–1.85 m$^{-1}$) that limit

Table 3. Regression equations to predict log(SD) and 1/SD†.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Variables‡</th>
<th>AIC</th>
<th>Adjusted R²</th>
<th>RMSE</th>
<th>N</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best equation based on the Akaike information criterion (AIC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>chl-α, $a_{440}$, TSS</td>
<td>−560</td>
<td>0.80</td>
<td>0.16</td>
<td>718</td>
<td>log(SD) = 0.722 − 0.202 × log(chl-α) − 0.240 × log($a_{440}$) − 0.446 × log(TSS)</td>
</tr>
<tr>
<td>All</td>
<td>chl-α, $a_{440}$, TSS</td>
<td>1162</td>
<td>0.79</td>
<td>0.54</td>
<td>719</td>
<td>1/SD = 0.790 + 0.0095 × Chl-α + 0.052 × $a_{440}$ + 0.054 × TSS</td>
</tr>
<tr>
<td>NLF</td>
<td>chl-α, $a_{440}$, TSS</td>
<td>−531</td>
<td>0.75</td>
<td>0.15</td>
<td>581</td>
<td>log(SD) = 0.619 − 0.283 × log(chl-α) − 0.334 × log($a_{440}$)</td>
</tr>
<tr>
<td>NCHF</td>
<td>chl-α, TSS</td>
<td>−148</td>
<td>0.87</td>
<td>0.14</td>
<td>150</td>
<td>log(SD) = 0.88 − 0.26 × log(chl-α) − 0.334 × log(TSS)</td>
</tr>
<tr>
<td>WCBP</td>
<td>chl-α, TSS</td>
<td>−47</td>
<td>0.73</td>
<td>0.17</td>
<td>72</td>
<td>log(SD) = 0.687 − 0.153 × log(chl-α) − 0.594 × log(TSS)‡</td>
</tr>
<tr>
<td>WCBP</td>
<td>TSS, $a_{440}$</td>
<td>−49</td>
<td>0.74</td>
<td>0.15</td>
<td>58</td>
<td>log(SD) = 0.535 − 0.655 × log(TSS) − 0.184 × log($a_{440}$)‡</td>
</tr>
<tr>
<td>Bivariate relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>chl-α</td>
<td>−132</td>
<td>0.60</td>
<td>0.229</td>
<td>1081</td>
<td>log(SD) = 0.711 − 0.545 × log(chl-α)</td>
</tr>
<tr>
<td>All</td>
<td>TSS</td>
<td>−56</td>
<td>0.60</td>
<td>0.234</td>
<td>925</td>
<td>log(SD) = 0.628 − 0.677 × log(TSS)</td>
</tr>
<tr>
<td>All</td>
<td>$a_{440}$</td>
<td>457</td>
<td>0.27</td>
<td>0.305</td>
<td>977</td>
<td>log(SD) = 0.292 − 0.341 × log($a_{440}$)</td>
</tr>
<tr>
<td>All</td>
<td>SS$_{NA}$</td>
<td>690</td>
<td>0.08</td>
<td>0.357</td>
<td>879</td>
<td>log(SD) = 0.231 − 0.403 × log(SS$_{NA}$)</td>
</tr>
</tbody>
</table>

Note: SS$_{NA}$

† All regressions and coefficients significant at $P < 0.0001$ except as noted in footnote ‡.

‡ All independent variables were log-transformed except in row 2 (regression vs. 1/SD).

§ $P = 0.009$ for chl-α and 0.04 for $a_{440}$ in the WCBP regression equations.

Fig. 6. (A) The 1/SD vs. $a_{440}$ and regression line for all sites in the database with chl-α ≤ 10 μg/L; (B) same plot and regression but for sites with chl-α ≤ 10 μg/L and $a_{440} > 3.0$ m$^{-1}$. 
their usefulness for SD predictions. The unexplained variance in the models likely reflects measurement uncertainties in the variables, especially SD, values for which depend on such factors as sunlight intensity/angle, wave activity, and observer biases. Uncertainties from model assumptions also cannot be ruled out. For example, chl-\( a \) is an imperfect measure of the effect of algae on SD because it does not take into account differences in “packaging”; i.e., a given chl-\( a \) concentration associated with algal cells clumped into visibly large particles will have less effect on SD than the same concentration associated with discrete free-floating cells. Moreover, chl-\( a \) to cell–volume relationships differ between algal taxa, and chl-\( a \) concentrations within algal cells vary with environmental conditions (Reynolds 1984, Felip and Catalan 2000). Similarly, TSS is a mass-related variable, but the amount of light scattered by suspended particles depends on particle numbers, shapes, and surface properties more than on mass itself. Thus, the multiple regression results should be viewed as tools to understand how limnological characteristics affect SD, rather than as equations that exactly predict SD given certain physical/optical measurements.

The PCA results (Fig. 3) support the ideas of the nutrient–color paradigm (Williamson et al. 1999, Webster et al. 2008, Fergus et al. 2016), which views trophic (food web) processes in lakes as defined by two orthogonal drivers: CDOM and TP. The former promotes heterotrophy and the latter autotrophy with attendant effects on algal (chl-\( a \)) production. Although the influence of CDOM on algal biomass and production is less clear at large (continental) scales (Havens and Nurnberg 2004, Yuan and Pollard 2014), more consistent negative impacts have been observed at local to regional scales (e.g., Karlsson et al. 2009, Fergus et al. 2016), mediated by the strong reduction of light availability. Recent research shows that despite modest positive impacts on nutrient availability, CDOM suppresses primary production and food web production in colored lakes (Karlsson et al. 2009, Thrane et al. 2014, Creed et al. 2018).

**Implications for SD trophic state standards for CDOM-rich lakes**

We used our results to evaluate the influence of CDOM on SD and specifically the levels of CDOM that interfere with interpreting SD as a measure of algal abundance and lake trophic state. The State of Minnesota has adopted eutrophication standards for warm- and cold- (i.e., trout) water NLF lakes. For example, the standards for warm-water NLF lakes include water quality criteria of 30 \( \mu \)g/L for total phosphorus (TP), 9 \( \mu \)g/L for chl-\( a \), and 2.0 m for SD as June–September averages (Heiskary and Wilson 2005). Nonattainment of this standard occurs when the TP criterion is exceeded and either or both chl-\( a \) and SD are in non-attainment. In evaluating whether warm-water NLF lakes satisfy water quality conditions for their designated beneficial uses, use of the SD criterion is based on the assumption that SD is controlled by algal abundance. Inspection of the SD vs. \( a_{440} \) plots in Fig. 5C, D shows that no site with \( a_{440} > 8 \text{ m}^{-1} \) had an SD > 2.0 m. Of the 128 sites having \( a_{440} > 8 \text{ m}^{-1} \) and also having chl-\( a \) data, chl-\( a \) was < 10 \( \mu \)g/L in 100 cases; 20 had chl-\( a \) of 10–20 \( \mu \)g/L, and only 8 had chl-\( a \) > 20 \( \mu \)g/L. It thus is apparent that SD is limited to values < 2.0 m primarily by CDOM for sites with \( a_{440} > 8 \text{ m}^{-1} \) and that an SD limit of 2.0 m is unlikely to be a realistic trophic state criterion for such waters.

The above result can be considered an “upper limit,” however, insofar as it is likely highly that \( a_{440} \) values < 8 \text{ m}^{-1} also affect SD. To evaluate the \( a_{440} \) and\( L \) level at which this begins to occur (relative to the 2-m SD criterion), we used the regression results described above. Insertion of the MPCA’s chl-\( a \) criterion of 9 \( \mu \)g/L for NLF lakes and SD = 2 m into the 1/SD predictive equation (Table 3) led to a negative \( a_{440} \), even at SS\( _{NA} = 0 \), however, and clearly that is not possible. As described above, high chlorophyll and/or TSS concentrations and associated low SD that occurred at some low-CDOM sites distorted the 1/SD vs. \( a_{440} \) relationship (Fig. 6A). The plot of 1/SD vs. \( a_{440} \) for sites with chl-\( a \) < 10 \( \mu \)g/L and \( a_{440} > 3 \text{ m}^{-1} \) (Fig. 6B), however, eliminated sites affected by high SS\( _{NA} \). The best-fit line for this data set yielded \( a_{440} = 4.2 \text{ m}^{-1} \) for an SD of 2 m. We suggest that this is a reasonable value for limiting use of the 2-m SD water quality standard as an indicator of eutrophication impairment in NLF lakes. As a minimum, it should serve as a threshold where additional response data (e.g., chl-\( a \)) are needed to assess lake eutrophication. Because SD standards vary by state and ecoregion, the specific CDOM threshold value may be different in other settings.

The above analysis is not to suggest that lakes with high CDOM (\( a_{440} > ~ 4 \text{ m}^{-1} \)) cannot have eutrophication problems, although high levels of CDOM tend to promote heterotrophy rather than autotrophy and may suppress algal growth, thus counteracting some effects of eutrophication (e.g., Williamson et al. 1999, Webster et al. 2008). Instead, these results indicate that SD is not an effective predictor of such problems in lakes with color higher than \( a_{440} \sim 4 \text{ m}^{-1} \).

Restricting the use of SD as a water quality standard based on a threshold level of CDOM indicates that greater reliance on indicators such as TP and chl-\( a \) is needed to assess impacts of eutrophication in the NLF. TP, TN, and chl-\( a \) are commonly used trophic state parameters that are linked to key ecosystem services (Keeler et al. 2012) and water quality standards (Heiskary and Wilson 2008). Colored dissolved organic matter has been shown to influence all of these parameters in oligotrophic lakes (e.g., Karlsson et al. 2009). Increasing CDOM levels in many lakes across northern Europe and parts of North America (Monteith et al. 2007, Williamson et al. 2015, Corman et al. 2018) also may be
increasing the effects of CDOM on trophic state metrics in these regions. Although these influences may be somewhat subtle compared to human influences on nutrients, they may require further modification of trophic state standards in moderate-to-high CDOM lakes, where availability and cycling of nutrients contrast strongly with those of low-CDOM waters.

Acknowledgments

This work was supported in part by National Science Foundation grant (CBET 1510332), the Minnesota Environmental and Natural Resources Trust Fund, as recommended by the Legislative-Citizen Commission on Minnesota Resources, the Minnesota Pollution Control Agency’s lake water quality assessment program, and the University of Minnesota’s Office of the VP for Research and Retirees Association, U-Spatial Program, Sea Grant Program, and Agricultural Experiment Station. We thank numerous collaborators, research staff, and students for assistance in sample collection and analysis and colleague Marvin Bauer for advice and encouragement throughout the larger project, of which this paper is a component.

Literature Cited


SUPPORTING INFORMATION

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.1871/full

DATA AVAILABILITY

The data upon which the analyses reported are based are available from the Data Repository for U of MN (DRUM) at: https://doi.org/10.13020/01wt-ij66.