# Verification of a simple band ratio algorithm for retrieving Great Lakes open water surface chlorophyll concentrations from satellite observations

Barry M. Lesht[a,*], Richard P. Barbiero[b], Glenn J. Warren[c]

[a] *CSC and Department of Earth and Environmental Sciences, University of Illinois at Chicago, 845 W. Taylor St., Chicago, IL 60607, USA*
[b] *CSC and Loyola University of Chicago, 1359 W. Elmdale Ave. Suite 2, Chicago, IL 60660, USA*
[c] *USEPA Great Lakes National Program Office, 77 W. Jackson Boulevard, Chicago, IL 60604, USA*

## Abstract

We compared *in situ* surface chlorophyll concentration values measured between 2012-2015 as part of the U.S. Environmental Protection Agency's Great Lakes National Program Office (GLNPO) annual monitoring program with corresponding concentration estimates obtained by applying our previously published (Lesht et al., 2013) Great Lakes Fit (GLF) band ratio algorithm to data from the Moderate-resolution Imaging Spectroradiometer (MODIS) sensor. Coefficients used in the original GLF algorithm were derived from similarly matched GLNPO and satellite observations collected between 2002 and 2011. The Model II linear relationship between the original GLF-predicted log-transformed values and the new set (2012-2015) of field observations yielded intercept = 0.036, slope = 1.063, and $r^2$ = 0.830. Residuals for modeled chlorophyll concentrations below ~ 8.0 mg $m^{-3}$ were unbiased and normally distributed, but positively biased at higher modeled

*Corresponding author
*Email addresses:* blesht@gmail.com (Barry M. Lesht), gloeotri@sbcglobal.net (Richard P. Barbiero), warren.glenn@epa.gov (Glenn J. Warren)

concentrations. When applied to the entire data set (2002-2015), the linear relationship between the GLF-modeled and the observed values had intercept = -0.000, slope=0.999, and $r^2$=0.820. New model coefficients derived from the entire (2002-2015) data set were very similar those obtained from the 2002-2011 data. Continual testing and assessment of any empirical model is desirable especially when the model is designed to be employed by a broad community. We conclude that this comparison of the GLF algorithm with the additional four years of independent data further validates its use for estimating surface chlorophyll concentrations from satellite observations of the open waters of the Great Lakes.

*Keywords:* Remote sensing, ocean color, chlorophyll *a*, satellite observations, algorithm verification

---

## Introduction

Several papers have described new algorithms for estimating chlorophyll concentrations in the Great Lakes from satellite observations (see Lesht et al. (2012) for a review of work prior to 2012 and more recently Lesht et al. (2013) and Shuchman et al. (2013)). However, in contrast to the continual assessment of algorithms developed by the major space agencies (*e.g.* NASA and ESA) which are provided to users via the agencies' web sites, to our knowledge no work has been reported in which the performance of these Great Lakes-specific algorithms has been tested with new, independent observations. Because all empirical or semi-empirical algorithms depend on the data from which their numerical coefficients are derived, testing with new data is essential both to assess the models' success and to examine their limitations (Augusiak et al., 2014). In some cases it might be desirable simply to update the model coefficients as more data are acquired (Werdell

et al., 2003), in other cases models may have to be discarded or structurally modified if systematic failure is observed.

In Lesht et al. (2013) we demonstrated that a simple band-ratio algorithm could be used to estimate surface chlorophyll values in the offshore waters of the Great Lakes from ocean color satellite observations (SeaWiFS from 1998-2010 and MODIS from 2002-2011). In that work we tuned the MODIS retrieval model to field data collected in all five lakes as part of the annual GLNPO monitoring program that we pooled across ten years (2002-2011). Although we presented an uncertainly analysis based on Monte-Carlo and sub-sampling methods in that paper, we were unable to test the algorithm with new, independent data. The primary purpose of this note is to report the results of our assessment of the GLF algorithm's performance based on comparison of its predictions with new MODIS observations collected from 2012-2015. Such a test is necessary to demonstrate that the algorithm is capable of representing data that were not available when it was developed and so could not have affected its structure or coefficients (Augusiak et al., 2014). For completeness, our secondary objective was to determine how much the original model coefficients change when the model is fit to the full (2002-2015) data set.

## Methods

*Matching field observations and satellite data*

Our original methods for matching the GLNPO field measurements with the satellite observations (Lesht et al., 2013) were similar to the procedures described by Bailey and Werdell (2006). Field sampling and determinations of *in situ* chlorophyll concentrations followed standard GLNPO protocols (U.S.E.P.A., 2010) and were consistent

throughout the study period. For the satellite data, we processed Level 1A (digital counts) data to Level 2 (geophysical values, L2) using the *l2gen* (v8.1.4) processing code in SeaDAS (Baith et al., 2001). We limited comparisons to satellite images that were collected within one day of the field sampling. This temporal window is larger than the ±3-hr window used by Bailey and Werdell (2006); inconsistencies in the recorded times of the GLNPO field sampling made it impossible for us to resolve them to intervals finer than one day. Images that satisfied the temporal matching criterion next were checked for viewing geometry and for overall cloud cover. Images for which the solar zenith angle at the scene center exceeded 70° were rejected as were those for which more than 20% of the pixels exceeded a satellite zenith angle of 60°. We also rejected images in which less than 20% of the lake surface was cloud free. For the images that passed this screening, we then checked the native-resolution L2 pixels within 5x5-pixel boxes centered on the field sampling locations to ensure that none of the NASA quality flags (http://oceancolor.gsfc.nasa.gov/VALIDATION/flags.html) we used (ATMFAIL, LAND, CLDICE, HIGLINT, HILIT, STRAYLIGHT, CHLFAIL, NAVFAIL) were set. This quality test is more stringent than the one used by Bailey and Werdell (2006) who accepted matches for which 50% of the pixels (rather than all) in the 5x5 box were unflagged. The CHLFAIL test, which checks the basic shape of the input reflectance spectrum, is based on the standard NASA OC3M wavelengths; because we used these same wavelengths in GLF algorithm (described below) CHLFAIL also screened pixels for which the input spectra were inappropriate for use with the GLF algorithm. To further limit the sampling area, we used the arithmetic mean value of the retrieved values within a 3x3 pixel box centered on the sample location as the representative satellite value for that location. Fi-

nally, we rejected matches for which any of the values of remote sensing reflectance ($Rrs$) were less than zero, indicating possible overcorrection by the atmospheric radiation model used to remove the contributions of atmospheric scattering and reflectance from the signal received at the satellite. We used the same SeaDAS default 2-band NIR iteration model (Bailey et al., 2010) that Shuchman et al. (2013) showed performs well over the Great Lakes in our processing.

We did not include either a check for spatial homogeneity or a pixel-by-pixel check of the satellite viewing angle in our original (Lesht et al., 2013) analysis; both were used by Bailey and Werdell (2006) in their study. To make our approach in the present work even more comparable to theirs, we added the pixel-based validation flag HISATZEN to the suite of flags checked in the 5x5 pixel box and also used a spatial homogeneity test, similar to that presented by Kahru et al. (2014) in the new calculations presented here. For the homogeneity test we rejected matches if the retrieved chlorophyll concentration ($C$) values in the 5x5 pixel region surrounding the sampling location were such that $(C_{max} - C_{min})/C_{min} > 1$. We include comparisons between the results of the original analysis (Lesht et al., 2013) and of the updated analysis where appropriate below.

*Chlorophyll retrieval*

The general form of the band-ratio retrieval algorithm is $log_{10}(Chl_{mod}) = c_0 + \sum_{i=1}^{n} c_i X^i$, in which $X$ is $log_{10}(MBR)$, the $c_i$ are the model coefficients, and $MBR$ represents the maximum band ratio, which for MODIS is calculated as $Max\{Rrs_{443}, Rrs_{488}\}/Rrs_{547}$ in which $Rrs_{nnn}$ is the remote sensing reflectance at nominal wavelength $nnn$. In our original investigation of the GLF we found that a third-order polynomial was adequate for use in the Great Lakes and noted that the primary effect of adding the fourth-order term to the

polynomial was to change the shape of the relationship at larger values of $log_{10}(MBR)$ (smaller values of chlorophyll). The fourth-order term is used in the standard NASA band-ratio algorithms because it is necessary for retrievals in the ocean where the lowest chlorophyll concentrations can be over an order of magnitude smaller than they are in the Great Lakes (Lesht et al., 2013). Although we did test a fourth-order version of the GLF as part of this work, we noted only a slight improvement over the third-order model and report only the most basic results of this test.

The coefficients used in the original GLF model were obtained by using an iterative method based on successive applications of a reduced major axis (Model II) regression, the appropriate regression approach when both variables are uncertain (Press and Teukolsky, 1992). This procedure, which yields the model coefficients that result in a 1:1 relationship between the retrieved and observed log-transformed chlorophyll values, is the same method used by NASA in their development of the standard SeaWiFS and MODIS retrieval algorithms (O'Reilly et al., 2000). In the present study, we used a maximum likelihood fitting method based on direct minimization of the chi-square function appropriate when both variables are subject to experimental error to determine the GLF coefficients. The chi-square function is written as (Press and Teukolsky, 1992)

$$\chi^2(a,b) = \sum_{i=1}^{N} \frac{(y_i - a - bx_i)^2}{\sigma_{yi}^2 + b^2\sigma_{xi}^2} \tag{1}$$

in which $y$ is $\log_{10}(C_{mod})$ and $x$ is $\log_{10}(C_{obs})$. Because the retrieval algorithm is in the form of a polynomial ($\log_{10}(C_{mod}) = c_0 + \sum_{j=1}^{n} c_j X^j$, in which $X$ is $log_{10}(MBR)$, $n$ is the degree of the polynomial ($n = 3$ in the original GLF and $n = 4$ in the standard NASA retrieval models) and the $c_j$ values are the model coefficients) we can substitute this polynomial for

$y_i$, apply the desired constraints that the intercept ($a$) equals zero and slope ($b$) equals 1, and re-write Eq. 1 as a function only of the model coefficients ($c_j$).

$$\chi^2(\mathbf{c}) = \sum_{i=1}^{N} \frac{((c_0 + \sum_{j=1}^{n} c_j X_i^j) - x_i)^2}{\sigma^2_{c_{0+\sum_{j=1}^{n} c_j X^j}} + \sigma_x^2}. \tag{2}$$

We used the R (R Core Team, 2014) package optimx (Nash and Varadhan, 2011) to minimize Eq. 2 with respect to the parameter vector $\mathbf{c}$. The initial parameter vector needed to begin the minimization was obtained by using the function *monpol* from R package MonoPoly (Murray et al., 2013) for the third-order model and with the *lm* function from the R base stats package (R Core Team, 2014) for the fourth-order model (fourth-order models are not necessarily monotone). Our experience comparing the original and the maximum likelihood fit methods showed that the maximum likelihood method yields nearly identical results to the original method (see Table 2) but is more intuitive and computationally efficient.

*Comparison statistics*

Our analysis generally follows the procedures outlined by Bailey and Werdell (2006) and Campbell and O'Reilly (2006) both of which were developed specifically for evaluation of satellite retrieval algorithms. To provide measures of the overall bias and uncertainty associated with the GLF, we calculated both the ratio of satellite to *in situ* chlorophyll and the absolute percent difference for each matched pair of observations (Bailey and Werdell, 2006). The absolute percent difference (PD) is calculated as:

$$PD_i = 100 * \frac{|X_i - Y_i|}{Y_i} \tag{3}$$

where the $X_i$s are the satellite values and the $Y_i$s are the *in situ* values. We refer to the median value of the absolute percent differences as $MPD$.

We also calculate the median of the ratios, an indicator of the overall bias of the estimates, and the semi-interquartile range (SIQR) of the ratios, a measure of the spread of the data. The SIQR is defined as:

$$SIQR = \frac{Q_3 - Q_1}{2} \tag{4}$$

In which $Q_1$ is the 25th percentile of the ratios and $Q_3$ is the 75th percentile. Thus 50% of the observed ratios fall between $Q_1$ and $Q_3$.

In addition, we calculated the log-transformed model error or residual ($\delta_i$), the mean square error (MSE), and the root mean square error (RMSE) following Campbell and O'Reilly (2006). These are defined as:

$$\delta_i = log_{10}(C^i_{mod}) - log_{10}(C^i_{obs}) = log_{10}\left(\frac{C^i_{mod}}{C^i_{obs}}\right) \tag{5}$$

in which $C^i_{mod}$ and $C^i_{obs}$ represent the *ith* modeled and observed values; bias is the average value of the residuals, and the mean square error is defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\delta_i)^2, \tag{6}$$

and the root mean square error (RMSE) is simply $\sqrt{MSE}$.

Finally, we used Major Axis (MA) Model II regression as implemented in the R package lmodel2 (Legendre, 2014) to assess the linear association between the modeled and the observed chlorophyll values. Because the distributions of both the observed and modeled chlorophyll values are approximately log-normal (see Lesht et al. (2013) Fig. 2), we based our regression calculations on the log-transformed values.

## Results

Figure 1 shows the relationship between the *in situ* chlorophyll measurements and the maximum band ratio ($Max\{Rrs_{443}, Rrs_{488}\}/Rrs_{547}$) obtained from the corresponding satellite images. The points measured between 2002-2011 that were used to determine the original GLF coefficients are shown as the gray crosses. The new (2012-2015) points are color coded by lake (the official lake colors are: Erie = red, Huron = green, Michigan = blue, Ontario = pink, Superior = yellow), season (darker shade indicates summer sampling), and shape coded by year. The curves illustrate the original GLF model (black solid line) and the model fits to the reduced (by adding the homogeneity and HISATZEN constraints) 2002-2011 data (black dashed line) and to the full (2002-2015) data (blue dashed line).

*Effects of adding the homogeneity and HISATZEN constraints*

Adding the homogeneity and HISATZEN constraints reduced the total number of accepted matches for the period 2002-2011 from 774 used in the original GLF analysis to 708. Three points were eliminated by using the pixel-based HISATZEN test. Of the 63 matches that were rejected by the homogeneity test, 28 were from Lake Erie, 20 from Lake Superior, and 15 from Lake Ontario. Of the 20 rejected matches from Lake Superior, 13 were samples from one station (SU06, 45.55861N, 86.37694W) which, though in deep water (~165 m), is relatively close to the northeastern shore of the lake and near a river mouth. In the other lakes, the rejected matches also tended to be from stations close to shore where spatial gradients in chlorophyll might be expected, but no other single station accounted for so many of the matches that failed the test.

With a regression slope of 0.978, the original GLF model predicts slightly lower values of chlorophyll than were observed when it is applied to the reduced 2002-2011 data (Table 1), but the RMSE is lower and the $r^2$ value of the regression is higher, indicating that some of the error in the original fit resulted from including non-representative points. When the GLF model is fit to the reduced 2002-2011 data, the change in the model coefficients is small (Table 2).

*Original GLF applied to 2012-2015 data*

Comparison between the original GLF model-predicted and observed 2012-2015 chlorophyll values are shown in Fig. 2. The annotations in panel (a) note the results (intercept = 0.036, slope = 1.063, and $r$ = 0.911) of the Model II regression along with the root-mean square error (0.181) and bias (0.040) The residuals of the fitted models $log_{10}(C_{mod}/C_{obs})$) are plotted as a function of the model-predicted values in panel (c). The mean and standard deviation ($\sigma$) of the residuals and fractions of residuals within $1\sigma$ of the means also are noted.

*Original GLF applied to 2002-2015 data*

The relationship between the modeled (original GLF) and the observed chlorophyll values for the full (2002-2015) data set is illustrated in Fig. 3. Model II regression between the 956 log-transformed modeled and *in situ* values yields an intercept of -0.000, slope of 0.999, and $r^2$ of 0.820 (panel a). The quantile-quantile plot (panel b) and plot of the relative frequency distributions (panel d) show that the distributions of modeled and observed values are very similar throughout most of the range, with a modest deviation at the lowest end of the distributions. The distribution of the log of the ratio of the modeled to

observed values (panel c) shows that the errors are symmetric and distributed log-normally with a mean value very close to zero (0.04) and standard deviation of 0.176 (log units).

## Discussion

The original GLF was developed to establish a simple method for estimating surface chlorophyll concentrations in the offshore waters of the Great Lakes from readily available satellite observations. Because the GLF was intended to provide a single relationship that could be applied to the lakes universally, it is to be understood that this generality might involve some sacrifice of local accuracy (*i.e*, the algorithm is not tuned by lake or by subregion with a lake). In particular, we would not expect the algorithm to apply to regions of the lakes not represented in the sample data (*e.g.* shallow waters, nearshore areas, and embayments). We also want to emphasize that because the data used to derive the GLF model coefficients were obtained exclusively from observations made as part of the annual GLNPO monitoring program, strictly speaking, the GLF is an algorithm that can be used to estimate the surface chlorophyll values as measured by the GLNPO program. As such, the degree to which the GLF estimates the "true" surface chlorophyll concentrations depends on the accuracy and representativeness of the underlying *in situ* measurements. Of course, this limitation is true of any method fit to a particular set of field observations.

Both the range and the scatter of the new (2012-2015) chlorophyll values are similar to those typical of the values collected between 2002-2011. The new observations fall within the cloud of points formed by the original data (Fig. 1), reinforcing our conclusions (Lesht et al., 2013) that the GLNPO chlorophyll data can be represented by a simple function of the MODIS maximum band ratio and that function is temporally stable.

Examination of the fit and residuals when the original GLF is used to model the 2012-2015 data (Fig. 2) shows the model performs well through most of the range of modeled chlorophyll values. Despite the cluster of positive residuals at higher values of modeled chlorophyll, 72% of the residuals fall within $1\sigma$ of the mean. The largest residuals (at higher chlorophyll values ($> \sim$8 mg $m^{-3}$)) consist of observations from the western and central basins of Lake Erie where it is most likely that non-chlorophyll constituents such as suspended mineral sediment and colored dissolved organic material interfere with the band-ratio retrievals (Binding et al., 2008, 2010, 2012; Shuchman et al., 2013). Because the uncertainty of the GLF retrievals is relatively high at these values we recommend caution when the GLF is applied to satellite observations with very low (< -0.20) values of the maximum band ratio. Observations of this magnitude, however, represent a small fraction (<1%) of the observed values.

The intercept of the GLF when applied to the 2012-2015 data is greater than zero and the slope is greater than one. Our original work showed that the slope and intercept of linear relationship between the modeled and observed values varied when the GLF was applied to individual years, and similarly varied when the GLF was applied to five-year subsets of the original data (Lesht et al. (2013), Tables 5 and 6). This variation reflects the unavoidable uncertainties associated with the sampling, the matching of the field data to the satellite imagery, the error in the reflectance values, and the degree to which interfering substances influence the observations (Lesht et al., 2012). To determine whether the fit obtained for 2012-2015 (Fig. 2) was anomalous, we compared the sample intercept and slope to empirical distributions obtained by using bootstrap techniques (Wu, 1986) in which we calculated the Model II fit between the GLF and observations for 10,000

randomly selected four-year subsets of data taken from the full (2002-2015) set. Figure 4 shows that the intercept and slope values obtained for 2012-2015, though greater than zero and one, are within the 95% bootstrap confidence intervals (-0.0364, 0.0420 for the intercept and 0.9164, 1.1261 for the slope).

The coefficient values determined from the maximum likelihood fit to the reduced sets of 2002-2011 and 2002-2015 values are very similar to the original GLF coefficients (Table 2). As we noted in our original derivation of the GLF, because the coefficients are determined with the constraint that the intercept and slope of the linear relationship between the log-transformed modeled and observed values be zero and one respectively, familiar least-squared methods cannot be used to assess the coefficient uncertainties. Rather, similarity between the different model coefficients can be determined by comparison with the empirically derived coefficient distributions obtained by using sub-sampling methods (Lesht et al. (2013), Fig. 9). Reference to this figure shows that the updated coefficient values are very close to the median values of their respective empirical distributions. The substantial reduction in mean absolute error (from 0.142 to 0.130) when the GLF is fit to the data set that includes the homogeneity constraint probably reflects elimination of questionable matching points.

The original GLF model performs very well when compared to the entire 2002-2015 data set (Fig. 3). The Model II intercept and slope of the linear relationship between $log_{10}(C_{mod})$ and $log_{10}(C_{obs})$ are -0.000 and 0.999 respectively, with low bias and RMSE comparable to that obtained in the original fit based on the pre-2012 data. Furthermore, the residuals are normally distributed (panel (c)) and the frequency distribution of the modeled chlorophyll values closely matches that of the observed values (panel (d)). Taken together,

these results suggest that the original GLF can continue to be used without substantial modification.

## Conclusions

Observations from the four most recent years (2012-2015) of GLNPO monitoring confirm that there is a clear relationship between the field-measured chlorophyll concentrations and the satellite-observed reflectance band ratio. Estimates of chlorophyll concentration obtained from the GLF model closely agree with the new observations, though the model accuracy diminishes at higher ($> \sim 8$ mg m$^{-3}$) estimated chlorophyll concentrations, approximately corresponding to $log_{10}(BandRatio)$ values below -0.2. Adding a homogeneity constraint to the procedure used for selecting matched field and satellite observations resulted in a substantial reduction of the mean absolute error between the estimated and observed chlorophyll values, suggesting that some of the error in the original fit resulted from instances of spatial inhomogeneity. When the original GLF was tested with the full data set (2002-2015) the slope and intercept of the linear relationship between $log_{10}(C_{mod})$ and $log_{10}(C_{obs})$ was -0.000 and 0.999 respectively indicating excellent accuracy overall. An alternative to the original model that included a 4th-order term improved the model performance only slightly.

Algorithms that convert in-space observations of radiance to estimates of surface water chlorophyll concentration in the Great Lakes make it possible to take advantage of the frequent, synoptic, and high resolution measurements provided by satellite remote sensing in studies of bio-geophysical processes in the lakes. Because they are intended to be used by broad communities not necessarily expert in remote sensing or modeling, simply pre-

senting a new algorithm with an initial assessment of its validity may not be sufficient for its ultimate adoption. We argue that continued analysis of the algorithms' performance is necessary both to establish confidence in their application and to assess their limitations. We encourage both algorithm developers and other users to test and assess published algorithms in their work and to share their results for the benefit of the entire community.

## Acknowledgements

This work was supported by the USEPA Great Lakes National Program Office as part of EPA Contract No. EP-C-10-060, Technical, Analytical, and Regulatory Mission Support with CSC under the direction of Louis Blume, Project Manager. Although the research described in this article has been funded by the U.S. Environmental Protection Agency, it has not been subjected to Agency review. Any opinions expressed in this publication are those of the author(s) and do not, necessarily, reflect the official positions and policies of the U.S. EPA. Any mention of products or trade names does not constitute recommendation for use by the U.S. EPA. We gratefully acknowledge the contributions of Dr. Thomas Johengen of the University of Michigan and Dr. James Watkins of Cornell University who led the field data collection efforts and who generously provided their insight. We also thank three anonymous reviewers for their thoughtful and helpful suggestions.

## References

Augusiak, J., den Brink, P.J.V., Grimm, V., 2014. Merging validation and evaluation of ecologial models to 'evaludation': A review of terminology and a practical approach. Ecological Modelling 280, 117–128. doi:10.1016/j.ecolmodel.2013.11.009.

299 Bailey, S., Werdell, P., 2006. A multi-sensor approach for the on-orbit validation of ocean
300 color satellite data products. Remote Sensing of Environment 102, 12–23.

301 Bailey, S.W., Franz, B.A., Werdell, P.J., 2010. Estimation of near-infrared water-leaving
302 reflectance for satellite ocean color data processing. Optics Express 18, 7521–7527.

303 Baith, K., Lindsay, R., Fu, G., McClain, C., 2001. SeaDAS, a data analysis system for
304 ocean-color satellite sensors. EOS Transactions of the American Geophysical Union
305 82, 202.

306 Binding, C.E., Greenberg, T.A., Bukata, R.P., 2012. An analysis of MODIS-derived algal
307 and mineral turbidity in lake erie. J. Great Lakes Research 38, 107–116.

308 Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2008. Spectral absorption proper-
309 ties of dissolved and particulate matter in Lake Erie. Remote Sensing of Environment
310 112, 1702–1711. doi:10.1016/j.rse.2007.08.017.

311 Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2010. Suspended particulate mat-
312 ter in Lake Erie derived from MODIS aquatic colour imagery. International Journal
313 of Remote Sensing 31, 5239–5255. doi:10.1080/01431160903302973.

314 Campbell, J.W., O'Reilly, J.E., 2006. Metrics for quantifying the uncertainty in a chloro-
315 phyll algorithm: explicit equations and examples using the OC4.v4 algorithm and
316 NOMAD data. Report from the Ocean Color Bio-optical Algorithm Mini-workshop.
317 URL: http://oceancolor.gsfc.nasa.gov/MEETING/OCBAM/docs.

318 Kahru, M., Kudela, R.M., Anderson, C.R., Manzano-Sarabia, M., Mitchell, B.G., 2014.

Evaluation of satellite retrievals of ocean chlorophyll-a in the California Current. Remote Sensing 6, 8524–8540. doi:10.3390/rs6098524.

Legendre, P., 2014. lmodel2: Model II Regression. URL: http://CRAN.R-project.org/package=lmodel2. R package version 1.7-2.

Lesht, B.M., Barbiero, R.P., Warren, G.J., 2012. Satellite ocean color algorithms: a review of applications to the Great Lakes. J. Great Lakes Research 38, 49–60. doi:10.1016/j.jglr.2011.10.005.

Lesht, B.M., Barbiero, R.P., Warren, G.J., 2013. A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the Great Lakes. J. Great Lakes Research 39, 138–152.

Murray, K., Müller, S., Turlach, B.A., 2013. Revisiting fitting monotone polynomials to data. Computational Statistics 28, 1989–2005. doi:10.1007/s00180-012-0390-5.

Nash, J.C., Varadhan, R., 2011. Unifying Optimization Algorithms to Aid Software System Users: optimx for R. Journal of Statistical Software 43, 1–14. URL: http://www.jstatsoft.org/v43/i09/.

O'Reilly, J.E., Maritorena, S., Siegel, D.A., O'Brien, M.C., Toole, D., Mitchell, B.G., Kahru, M., Chavez, F.P., Strutton, P., Cota, G.F., Hooker, S.B., McClain, C.R., Carder, K.L., Müeller-Karger, F., Harding, L., Magnuson, A., Phinney, D., Moore, G.F., Aiken, J., Arrigo, K.R., Letelier, R., Culver, M., 2000. Ocean Color Chlorophyll a Algorithms for SeaWiFS, OC2 and OC4: Version 4. SeaWiFS postlaunch

technical report series. Vol 11, Part 3, Ch. 2. NASA-TM-2000-206892. NASA Goddard Space Flight Center. Greenbelt, Maryland.

Press, W., Teukolsky, S.A., 1992. Fitting straight line data with errors in both coordinates. Comput. Phys. 6, 274–276.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Shuchman, R.A., Leshkevich, G.A., Sayers, M.J., Johengen, T.H., Brooks, C.N., Pozdnyakov, D., 2013. An algorithm to retrieve chlorophyll, dissolved organic carbon, and suspended minerals from Great Lakes satellite data. J. Great Lakes Research 39 (Suppl. 1), 14–33.

U.S.E.P.A., 2010. Sampling and Analytical Procedures for GLNPO's Open Lake Water Quality Survey of the Great Lakes. URL: http://www.epa.gov/glnpo/monitoring/sop/index.html.

Werdell, P.J., Bailey, S., Fargio, G., Pietras, C., Knobelspiesse, K., Feldman, G., McClain, C., 2003. Unique data repository facilitates ocean color sensor validation. EOS Transactions of the American Geophysical Union 84, 377.

Wu, C., 1986. Jacknife, bootstrap and other resampling methods in regression analysis (with discussions). Annals of Statistics 14, 1261–1350. doi:10.1214/aos/1176350142.

Table 1: Chlorophyll validation statistics (after Bailey and Werdell (2006)) for different sets of coefficients and data

| Data | Coefficients | Ratio (SQIR)[a] | % Difference[b] | Intercept | Slope | $r^2$ | RMSE | N |
|---|---|---|---|---|---|---|---|---|
| 2002-2011 | GLF[1] | 0.967 (±0.255) | 24.85 | 0.000 | 1.000 | 0.800 | 0.183 | 774 |
| 2002-2011 | GLF[2] | 0.973 (±0.255) | 24.56 | 0.000 | 1.000 | 0.800 | 0.181 | 774 |
| 2002-2011[3] | GLF[1] | 0.946 (±0.224) | 22.76 | -0.012 | 0.978 | 0.821 | 0.168 | 708 |
| 2012-2015[3] | GLF[1] | 1.083 (±0.266) | 24.90 | 0.036 | 1.063 | 0.830 | 0.180 | 248 |
| 2002-2015[3] | GLF[1] | 0.973 (±0.240) | 23.26 | 0.000 | 0.999 | 0.820 | 0.174 | 956 |
| 2002-2015[3] | GLFv2[2] | 0.976 (±0.242) | 23.54 | 0.000 | 1.000 | 0.822 | 0.173 | 956 |
| 2002-2015[3] | GLFv2[2,4] | 0.977 (±0.240) | 23.09 | 0.000 | 1.000 | 0.824 | 0.172 | 956 |

Regression statistics (Model II) are based on log transformed values.

[a] Median model to *in situ* ratio and semi-interquartile range

[b] Median absolute percent difference

[1] Coefficients (see Line 1 of Table 2 below) from Lesht et al. (2013)

[2] Coefficients (see Line 2 of Table 2 below) determined using Eq. 2

[3] Including homogeneity and pixel-based HISATZEN tests

[4] 4th order model

Table 2: Best-fit coefficients determined for the Great Lakes Fit (GLF) model for different models, datasets, and fitting methods. $c_i$ are the model coefficient values, $N$ is the number of matched points and $MAE$ is the mean absolute error between the observed and modeled chlorophyll values.

| Years | Fit Method | $N$ | $c_0$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $MAE$ |
|---|---|---|---|---|---|---|---|---|
| 2002-2011 | NASA Iterative[1] | 774 | 0.3429 | -3.3925 | 3.3412 | 0.7857 | | 0.142 |
| 2002-2011 | Max Likelihood[2] | 774 | 0.3430 | -3.3942 | 3.3433 | 0.7858 | | 0.138 |
| 2002-2011[3] | Max Likelihood[2] | 708 | 0.3660 | -3.4553 | 3.1773 | 1.1559 | - | 0.130 |
| 2002-2015[3] | Max Likelihood[2] | 956 | 0.3578 | -3.2742 | 2.4548 | 0.7291 | - | 0.131 |
| 2002-2015[3,4] | Max Likelihood[2] | 956 | 0.3592 | -3.5244 | 2.5975 | 10.8482 | -22.8493 | 0.129 |

[1] Lesht et al. (2013)

[2] Equation 2

[3] Including homogeneity and pixel-based HISATZEN tests

[4] 4th order model

## List of Figures

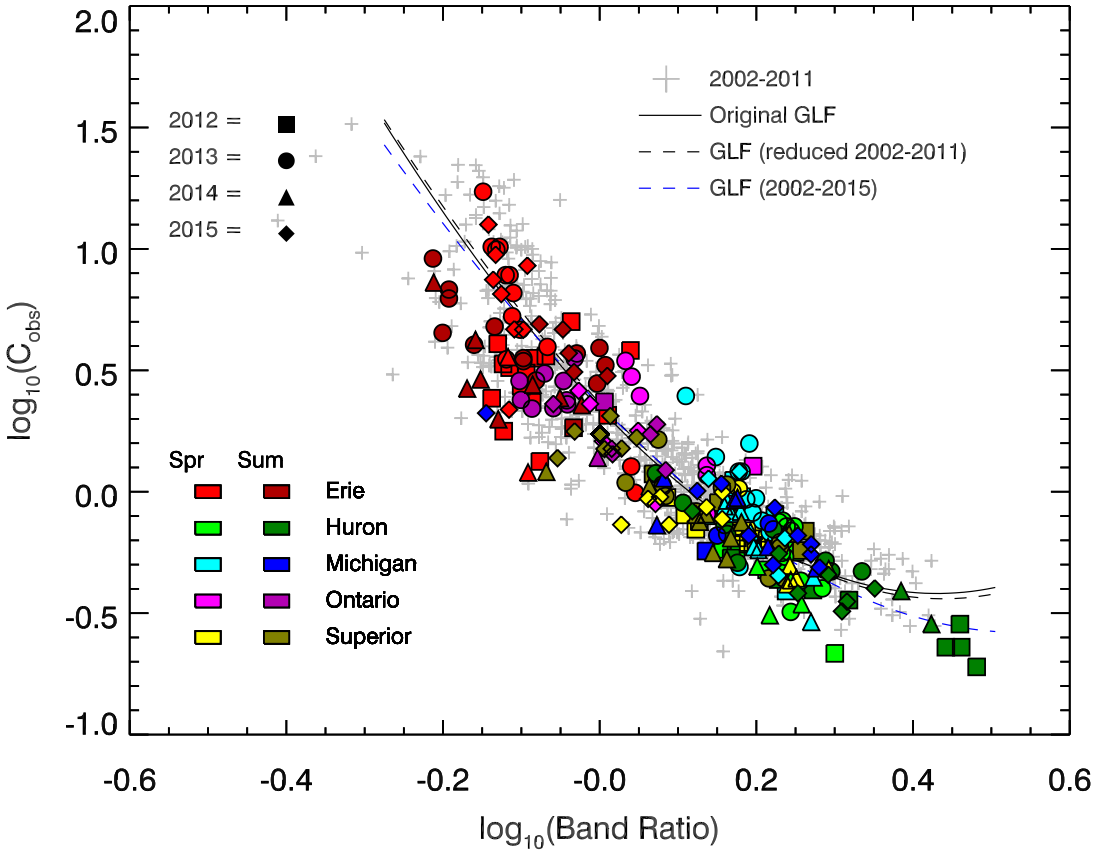


Figure 1: Log transformed observed chlorophyll ($\log_{10}(C_{obs})$) vs. $\log_{10}$(Band Ratio) for MODIS. The black solid line shows the original GLF model ((Lesht et al., 2013), Table 2). The black dashed line is the model derived from the same time period (2002-2011) after elimination of points that fail the HISATZEN and homogeneity tests. The blue dashed line shows the third-order model fit to the entire (2002-2015) dataset.

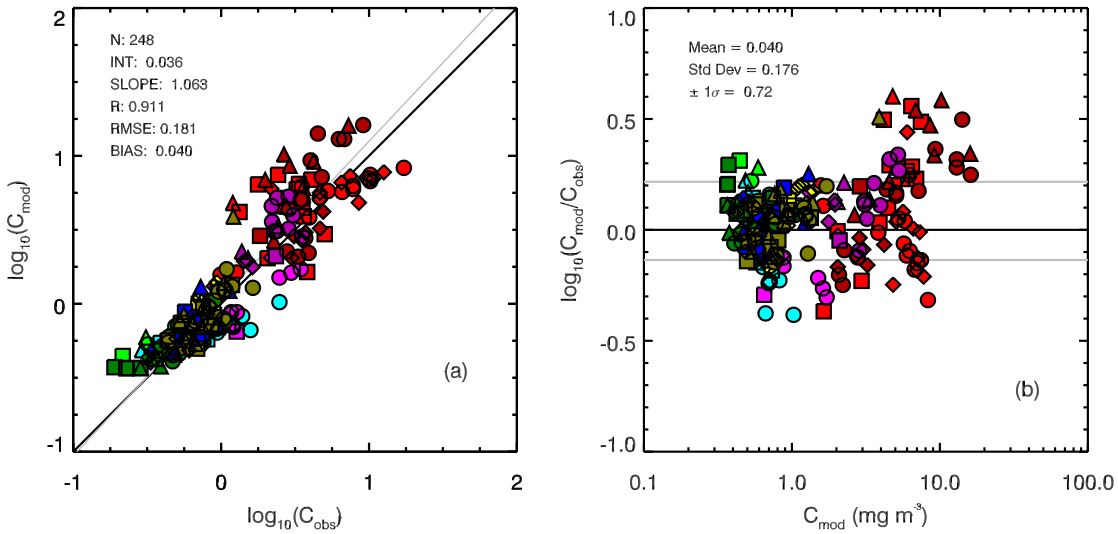


Figure 2: Scatter plots between $log_{10}(C_{mod})$ and $log_{10}(C_{obs})$ (a) and between residuals ($log_{10}(C_{mod}/C_{obs})$ and $C_{mod}$ (b) for the original GLF model fit to 2012-2015 data. Gray line in panel (a) shows the Model II regression; the black line is 1:1. Gray lines in panels (b) are $\pm 1\sigma$ of the residual mean. Lake color and year shape coding is the same as in Fig. 1.

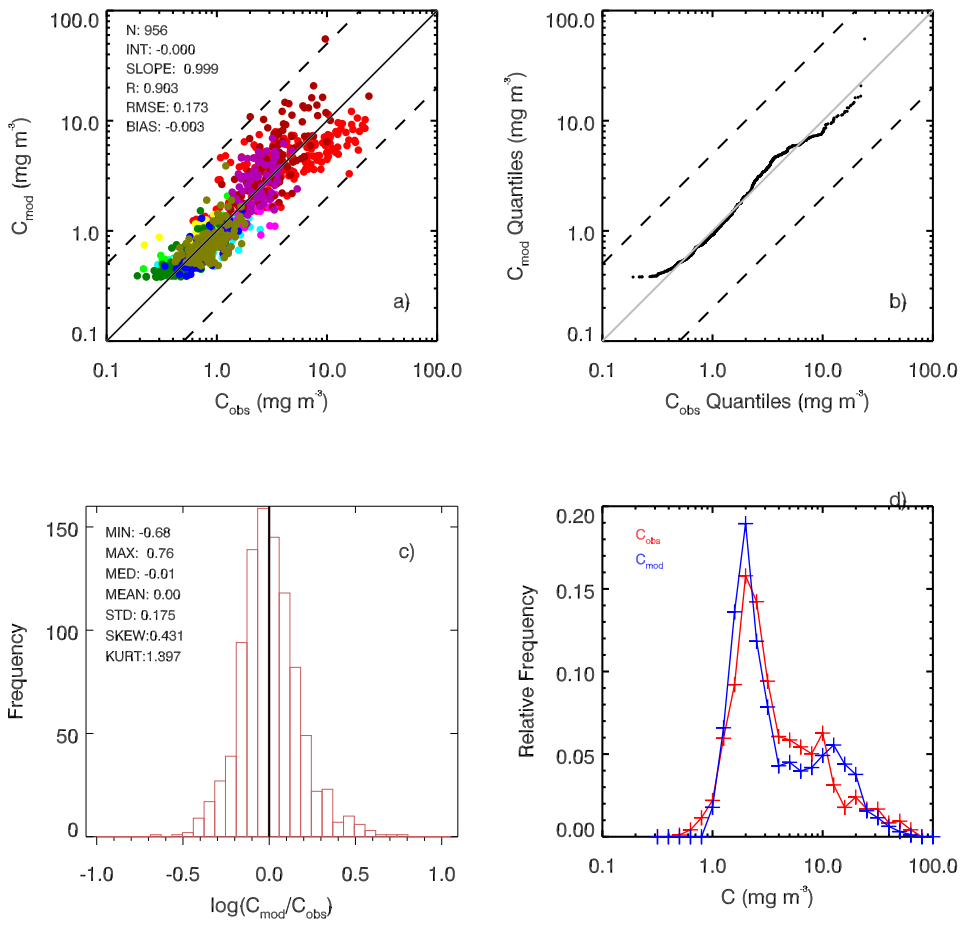


Figure 3: Comparison of GLF modeled chlorophyll ($C_{mod}$) with *in situ* GLNPO observations ($C_{obs}$) 2002-2015: a) Scatterplot of $C_{mod}$ and $C_{obs}$; b) (Quantile-quantile plot of $C_{mod}$ versus $C_{obs}$; c) Frequency distribution of $log(C_{mod}/C_{obs})$; d) Relative frequency of $C_{mod}$ (blue curve) and $C_{obs}$ (red curve). Lake color code (panel a) is the same as in Fig. 1.

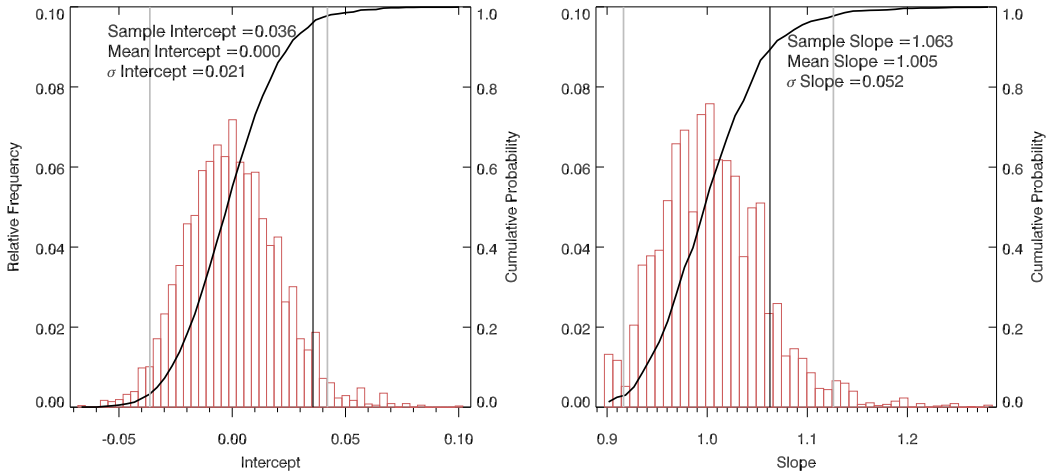


Figure 4: Bootstrap frequency distributions of the intercepts (left panel) and slopes (right panel) of the Model II regressions of $log_{10}(C_{mod})$ against $log_{10}(C_{obs})$ using the original GLF coefficients and 10,000 simulated four-year subsets of the 2002-2105 data set. The slope and intercept obtained for the 2012-2015 subset are indicated by the dark vertical lines and cumulative probability distributions by the overlaid curve. The 95% bootstrap confidence intervals are indicated by the light gray lines.