

To appear in the *International Journal of Remote Sensing*  
Vol. 00, No. 00, Month 2014, 1–26

# Optimizing a remotely-sensed proxy for plankton biomass in Lake Kivu

A. KNOX\*†, E. BERTUZZO†, L. MARI†‡, D. ODERMATT§¶, E. VERRECCHIA|, and  
A. RINALDO†□

†Faculté de l’Environnement Naturel, Architectural et Construit, École Polytechnique  
Fédérale de Lausanne, 1015 Lausanne, Switzerland

‡Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133  
Milano, Italy

§Brockmann Consult, 21502 Geesthacht, Germany

¶Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters  
Research and Management, 6047 Kastanienbaum, Switzerland

|Institut des Dynamiques de la Surface Terrestre, Université de Lausanne,  
1015 Lausanne, Switzerland

□Dipartimento di Ingegneria Civile, Edile ed Ambientale, Università di Padova,  
35131 Padova, Italy

(Received June 17, 2014)

Many regions of the world, including inland lakes, present with sub-optimal conditions for the remotely-sensed retrieval of optical signals, thus challenging the limits of available satellite data processing tools, such as atmospheric correction models (ACM) and water-constituent retrieval (WCR) algorithms. Working in such regions, however, can improve our understanding of remote-sensing tools and their applicability in new contexts in addition to potentially offering useful information about aquatic ecology. Here, we assess and compare 32 combinations of two ACMs, two WCRs and three binary categories of data quality standards to optimize a remotely-sensed proxy of plankton biomass in Lake Kivu. Each parameter set is compared against the available ground-truth match-ups using Spearman’s right-tailed  $\rho$ . Focusing on the best sets from each ACM-WCR combination, their performances are discussed with regard to data distributions, sample sizes, spatial completeness and seasonality. The results of this study may be of interest both for ecological studies on Lake Kivu and for epidemiological studies of disease, such as for cholera, whose dynamics in other regions of the world are often associated with plankton biomass.

## 1. Introduction

### 1.1 Context

Proxies of plankton biomass, such as *chlorophyll-a* (*chl-a*) concentrations or plankton absorption coefficients, can be important indicators of aquatic ecological processes. They can also be useful in the context of disease dynamics, as is the case for cholera, a devastating diarrheal disease. This is because the causative agent

---

This work was supported by the ERC Advanced Grant RINEC-227612, the SNF Project CR2312-138104-1 and the ENAC/EPFL discretionary fund, as well as the EU-FP-7 Marie Curie PEOPLE/IAPP project WaterS.

\*Corresponding author. Email: allyn.knox@epfl.ch

of cholera, *Vibrio cholerae*, lives as a member of the bacterioplankton community. Indeed, cholera outbreaks have been linked to plankton blooms in the Bay of Bengal where the pathogen is endemic. This cholera-plankton link is often found using remotely-sensed *chl-a* as a proxy for plankton biomass (Lobitz et al. 2000; De Magny et al. 2008; Ford et al. 2009; Jutla, Akanda, and Islam 2010; Jutla et al. 2013). The correlation between plankton blooms and cholera outbreaks may, however, depend on the local assemblages of plankton (Tamplin et al. 1990; Asplund et al. 2011).

This raised the question as to whether such a relationship exists for endemic cholera in the African Great Lakes (AGL) region, where the AGL may serve as cholera reservoirs (Olago, Marchall, and Wandiga 2007; Bompangue et al. 2008, 2009). As of this writing, the only publication to investigate the plankton-cholera relationship in the AGL region found the correlation indistinguishable from rainfall-mediated seasonality in the number of cholera cases, at least for Lake Tanganyika (Bompangue et al. 2011). Plankton also responds to seasonal shifts in the rainfall regime, with lower biomass during the rainy season (generally Oct.-May), than in the dry season (June-Sept. and around January); compositional shifts occur in January, May and September, including the presence of two different blooming species May-Sept. and Sept-Dec. (Sarmiento, Isumbisho, and Descy 2006). To further our understanding of plankton dynamics in Lake Kivu (Figure 1), we assess and attempt to optimize a remotely-sensed proxy for plankton biomass. This could possibly have implications for the mathematical modeling of cholera dynamics around Lake Kivu, described in Finger et al. (2014).

### Figure 1

For this study, *in situ chl-a* measurements at 7 locations (see Figure 1) were made available by Sarmiento, Descy, and Darchambeau (2012), to be described further in section 2.3. Although these *in situ* measurements are of high quality, they are of insufficient spatial and temporal coverage to be used alone in a study of cholera epidemiology, thus necessitating a remotely-sensed database.

## 1.2 Tools

Basically, the retrieval of water constituents by remote sensing consists of:

### 1. Signal processing

- the application of an ACM that removes atmospheric effects from the calibrated and geolocated top-of-atmosphere radiance, yielding the water-leaving radiance ( $L_w$ ) at each wavelength, which is then normalized to  $nL_w$  (Chavula et al. 2009);
- the use of a WCR algorithm to retrieve biological values like *chl-a* from the  $nL_w$  (Gurlin, Gitelson, and Moses 2011; Ocean Color Biology Group 2012);

### 2. Post-processing

- the application of data quality standards;
- an optional binning step, not used here, which consists of temporal and/or spatial interpolations.

The specific methods used to generate the remotely-sensed *chl-a* estimates for the Bompangue et al. (2011) cholera study on Lake Tanganyika are given in Horion et al. (2010). Briefly, they tested eight combinations of four ACM with two WCR, namely the over-ocean SeaDAS standard ACM and three ACM designed for coastal

regions, each with a different relative humidity, combined with one of two *chl-a* algorithms, the Clark algorithm or the MODIS OC3 algorithm. According to Horion et al. (2010), the most promising combination for Lake Tanganyika was the coastal ACM with 90% relative humidity coupled to the OC3 algorithm ( $R^2 = .64$  and  $N = 12$ ). In preliminary studies, we confirmed that this conclusion was also valid for Lake Kivu when the same ACM-WCR combinations were tested (Knox 2012). Thus, the coastal ACM with 90% relative humidity and the OC3 bio-optical algorithm are used here. Also of interest in this study are a short-wave infrared (SWIR) ACM and a WCR for retrieving the plankton absorption coefficient at a given wavelength, the quasi-analytic algorithm (QAA) (Lee et al. 1996; Lee, Carder, and Arnone 2002; Lyon et al. 2004). To our knowledge, neither has been applied for the retrieval of plankton information in the AGL region.

Both ACM tested in this study are specifically designed for coastal regions, but not specifically for inland lakes. Both account for multiple scattering aerosol calculations. The coastal ACM with 90% humidity consists of a fixed aerosol type and represents a significant improvement over its predecessors when applied to optically complex waters because it does not rely on the black pixel assumption (Ahmad et al. 2010; Bailey, Franz, and Werdell 2010). The SWIR model uses a 2-band correction in the short-wave infrared spectrum (1240 and 2130 nm bands) to calibrate radiances in the visible spectrum. However, although promising for turbid waters, it does not perform well in clear waters (Wang and Shi 2007; Wang, Son, and Shi 2009; Wang, Shi, and Tang 2011). Other drawbacks of the SWIR approach can include a large number of invalid output pixels and a high error rate due to the low signal-to-noise ratio in the SWIR correction bands (Werdell, Franz, and Bailey 2010).

One drawback of the OC3 algorithm, as well as similar band-arithmetic algorithms relying on the blue-green part of the visible spectrum, is its underlying assumption of a strong correlation between *chl-a* and colored dissolved organic matter, which is valid in the open ocean but not in coastal and inland regions. The quasi-analytic algorithm (QAA), on the other hand, is appropriate for these complex waters because it uses a spectral inversion algorithm based on radiative transfer theory to obtain estimates of phytoplankton absorption (Lee, Carder, and Arnone 2002; Lee et al. 2009; Shang et al. 2011; Odermatt et al. 2012). Although *chl-a* is more commonly used, the plankton absorption coefficient may provide a more reliable measure of primary production (Lee et al. 1996; Lee, Carder, and Arnone 2002; Lyon et al. 2004). Together, its potential robustness in retrieving relative plankton biomass values in inland or other complex waters and its utility as a measure of primary production make it an especially interesting algorithm to test in an inland lake.

### 1.3 Challenges

The optimization of a remotely-sensed plankton database for Lake Kivu requires an assessment of the applicability of remote sensing techniques to the specific conditions of an inland lake. Indeed, as ACMs and WCRs are typically designed for oceans, their applicability to Lake Kivu is not guaranteed.

Extensive cloud cover and haze diffusion over Lake Kivu, especially during the rainy season, cause turbid and heterogeneous atmospheric conditions. Consequently, the ratio of water-leaving to atmospherically scattered signals (including adjacency effects) decreases to critical levels, potentially decreasing the quantity and quality of plankton biomass estimates and thereby reducing their utility. Unfor-

Unfortunately, the automatic identification or correction of such conditions is currently not solved; therefore greater random variability in the dataset is inevitable (Ocean Color Biology Group 2012). Finally, the aerosols over southern Africa are unlike other aerosol types in that, relative to other regions, there is both a strong negative correlation between the single scattering albedo and the wavelength and also a considerably higher aerosol absorption (Russell et al. 2010). The latter is typically assumed to be negligible for the maritime aerosols for which most ACM were developed.

The surface area of Lake Kivu is also smaller than that of two AGL neighbors, Lakes Tanganyika and Malawi, whose plankton dynamics have been previously studied using satellite-based remote sensing techniques (Chavula et al. 2009; Horion et al. 2010). Unlike these neighbors, Lake Kivu also features a prominent island, which consequently increases the ratio of shoreline to surface area and thus, the impact of land proximity on retrieved data. Together, cloud cover, atypical aerosol absorption characteristics and land proximity reduce the quantity and quality of retrieved data, therefore necessitating thorough data assessment and optimization.

## 2. Methods

### 2.1 Data Treatment

All available MODIS-Aqua L1A images over Lake Kivu from June 2002 to June 2009 were obtained and processed to geolocated L1B data. After downloading the necessary data from the Ocean Color Biology Group, the ACM and bio-optical algorithms were applied using SeaDAS versions 6.2 to 6.4 (Ocean Color Biology Group 2012). The retrieved products and the associated reflectance (albedo) and flags were exported as ASCII files. Post-processing was performed in MATLAB.

Prior to post-processing, the QAA plankton absorption estimate had some negative values, which are physically impossible. However, because plankton absorption is calculated algebraically by subtracting absorption by water and by detritus from total absorption at a given wavelength (Lee et al. 2009; Naik et al. 2009), negative values may occur under conditions for which the algorithm is inapplicable. Negative values were therefore excluded from analysis.

### 2.2 Experimental Setup

Each processing and post-processing step was treated as a binary parameter influencing the quality of the retrieved data (Table 1). The two possible values for each parameter were coded as (+) and (-). With five such parameters, there are 32 possible parameter sets, all of which were tested.

**Table 1**

As illustrated in Table 1, parameters are as follows:

$X_1$ : The two ACM tested in this study are the coastal ACM with 90% relative humidity and the SWIR ACM.

$X_2$ : The first bio-optical algorithm tested is the MODIS standard empirical blue-green band-ratio algorithm known as OC3, using spectral bands at 443-555 nm. The second is QAA, which decomposes reflectance into apparent and inherent optical properties based on a sequence of algebraic expressions. Here we use the plankton absorption coefficient at a single wavelength (490 nm).

$X_3$ : Images yielding fewer than 5% or 15% of possible data-yielding pixels were excluded from analysis. Note that, here, in the post-processing phase, the total number of data-yielding pixels was set individually for each ACM-WCR combination to avoid *a priori* biasing in the parameter set selection process. The rationale is that low pixel yield is indicative of extensive cloud cover, where cloud-covered pixels yield no data and the feasible retrieval of nearby pixels is compromised by increased atmospheric turbidity. Removing the most affected images (i.e. those with lowest pixel yields) can help improve data reliability (Horion et al. 2010).

$X_4$ : Heterogeneity in the atmosphere and water can diminish the reliability of retrieved data in a given pixel (Hu, Lee, and Franz 2012; Ocean Color Biology Group 2012). Since the standard set of confidence flags results in the exclusion of all available data, only selected flags were applied, based on preliminary analyses (Knox 2012); see Table 2 for more details on the less permissive (-) and more permissive (+) flag combinations applied.

$X_5$ : The maximum albedo threshold was tested at double (.054) and triple (.081) the default value designed for clear over-ocean atmospheres (Ocean Color Biology Group 2012). This is because all pixels have values exceeding that of the default threshold, due to the high atmospheric turbidity over Lake Kivu.

**Table 2**

### 2.3 Ground-Truthing

Surface *in situ* measurements of plankton biomass (*chl-a*) were made available from 7 pelagic sites on Lake Kivu (Figures 1 and 2). These 255 *in situ* measurements, both published and unpublished, span the years 2002-2009. They were collected and analyzed according to methods described in Sarmiento, Isumbisho, and Descy (2006). The Ishungu and Kibuye sites provide, respectively, 62% and 22% of the *in situ chl-a* measurements, with the Bukavu Bay site yielding 8% and the remaining 8% coming from the other 4 sites. These *in situ* values are log-normally distributed, as is expected for *chl-a*. A timeline and frequency histogram of the *in situ* data for all sites can be found in Figure 2.

**Figure 2**

Ground-truth match-ups were assigned by matching the geographic coordinates of the *in situ* sites and the corresponding remotely-sensed geo-located 1 km<sup>2</sup> pixels. A same-day satellite image was matched to the *in situ* sampling date whenever possible; however, in the absence of a valid pixel value, the match-up was made using a temporal interpolation from two images, one dating to just before and one to just after the *in situ* sampling date, provided that they were within 3 days of each other and only 1-2 days of the sampling date. The geographic coordinates were the same. For both the same-day and the temporally-interpolated match-ups, the satellite overpass was within 3 hours of the time of day of the *in situ* sampling, thus minimizing the impact of diurnal fluctuations in plankton biomass.

### 2.4 Data Analysis

Although *chl-a* is usually log-normally distributed, this rule does not always hold true for smaller data subsets, which is the case for the ground-truth match-ups

in our study (data not shown). Because normality of log-transformed data cannot be assumed, Pearson's  $r$  was deemed inappropriate for this study, and the untransformed, true-value data were therefore analyzed using a right-tailed pairwise Spearman's  $\rho$ . Despite our best efforts to minimize the impact of data distortions on the results, the representativeness of small samples is always a concern in data-limited settings.

Several additional analyses were performed. Representative frequency histograms are shown of the processed raw values (including negative values, hence the use of raw data values rather than log-transformed values) as well as the log-transformed post-processed values. Sample sizes and  $\rho$  were plotted in a regression; spatial completeness statistics, timelines and selected images are shown, and parameter set-specific variations in the availability of match-ups are displayed with a highlight on seasonality.

Note that although data analysis was performed using the untransformed true values, frequency histograms are presented using log-transformed data whenever a log-normal distribution is expected. Additionally, graphical representations of the ground-truthing results and lake-wide images of the plankton biomass estimations use center-reduced data, or z-scores. While still imperfect, this technique facilitates the visual comparison of the different parameter sets by homogenizing the scale for the plankton absorption coefficients and *chl-a* estimates. Furthermore, as our ultimate purpose is the correlation of plankton biomass dynamics and their possible epidemiological implications, we are more interested in the relative plankton biomass, not in the absolute values. The z-scores of each data set are equal to the individual values minus the mean of the full data set (including but not limited to the ground-truth match-ups), and divided by the standard deviation. A z-score of 0 is equal to the mean, and a z-score of 1 is equal to one standard deviation above the mean. A high-performing data set would be expected to vary, relative to itself, to the same extent as the *in situ* data do relative to themselves, thus yielding similar z-scores.

### 3. Results & Discussion

#### 3.1 Parameter Set Performances

Looking at Table 3, the best parameter sets from each ACM-WCR group (in bold) all share the more restrictive confidence flag setting ( $X_4$ ). The minimum pixel yield ( $X_3$ ) is more exclusive (-) with the Coastal ACMs but less exclusive (+) with SWIR-OC3 and of no consequence with SWIR-QAA. The albedo threshold, at the tested values, was of no importance. Overall, the average  $\rho$  of 0.49 leaves something to be desired. However, it is not immediately clear from the tabularized results which ACM-WCR combination is best.

**Table 3**

Sample sizes varied between 6 and 39 (mean  $N = 19$ ) across parameter sets, with the SWIR ACM generally having smaller sample sizes (mean  $N = 13$  vs 26) and poorer coverage (mean coverage = .12 vs .23) but higher  $\rho$  (mean  $\rho = .57$  vs .40) than the Coastal ACM. Indeed, a negative trend between sample size and  $\rho$  emerges, which is discussed further in Section 3.2.2.

As for the comparison between QAA and OC3, the differences in coverage and in  $\rho$  are minimal, but QAA is more robust in attaining statistically significant p-

values, possibly due to larger sample sizes ( $N = 23$  vs 16). This bias toward QAA may have resulted from the exclusion of negative values if their exclusion acted as an additional targeted filter not applied to the OC3 parameter sets. In fact, despite its below-average  $\rho$  the Coastal-QAA group appears to be the most robust to noisy data, with significant p-values across all post-processing options (parameter sets 17-24).

Finally, the best-performing Coastal-OC3 parameter set is not characteristic of its group in terms of sample size (lower) or  $\rho$  (higher); however, as the most restrictive in terms of data quality (confidence flags and minimum pixel threshold), it may simply be a sign that the Coastal-OC3 combination is susceptible to noisy data and that removing the noise improves performance. This parameter set is also unique in its high rate of same-day match-ups (89% of match-ups), and while the potential impact of this difference remains elusive, it could be non-trivial.

The selected best sets from each ACM-WCR combination all, in their standardized forms (z-scores), follow roughly the same pattern as the *in situ* data. This can be seen in Figure 3 in the timelines and regressions of the z-scores.

### Figure 3

Across all sets, greater deviations from the regression lines are found in the high *chl-a* waters, which, incidentally, are only captured at the Ishungu site. However, while heterogeneities in site location cannot be excluded as a possible explanation for the greater unexplained variability, the most likely explanation is that spatially limited and often heterogeneous bloom conditions cannot be captured in the same way when using different methods (*in situ* biological sampling and remote sensing) at different spatial scales (point-source vs  $1 \text{ km}^2$  average).

If the distributions were all log-normal and the data noise-free, then the z-scores of the remotely-sensed and *in situ* data would match perfectly, i.e.  $y = x$ . By that standard, the Coastal-QAA set performs best, despite having one outlier (visible in Figure 3(e)), where it reproduces but overshoots a peak in the corresponding *in situ* z-score.

## 3.2 Interpretation Biases

In the context of data limitation and quality concerns,  $\rho$  and ground-truth regressions, alone, are insufficient to fully compare the different parameter sets. Hence, the need for further analyses of possibly hidden biases.

### 3.2.1 Data Distribution

Figure 4 shows representative frequency histograms before and after post-processing, demonstrating not only the importance of post-processing but also the fidelity of the final data distributions to the expected log-normal distributions. While all distributions prior to post-processing show evidence of significant noise, there is an important distinction between the post-processed SWIR distributions, in panels *b* and *d*, and the post-processed Coastal distributions, in panels *f* and *h*.

### Figure 4

Namely, the retention of a long right tail for both post-processed SWIR distributions shows evidence of a susceptibility to noise in the form of high values. This may reveal a susceptibility of SWIR to adjacency effects from bright objects (e.g. land) which saturate the signal and distort the retrieved values, or it may

simply reflect the relatively high proportion of invalid output pixels known to be problematic with SWIR (Werdell, Franz, and Bailey 2010). The post-processed distributions produced using the Coastal ACM, on the other hand, are log-normal, as would be expected of plankton biomass estimates. These results suggest greater applicability of the Coastal ACM in the study location, despite the high  $\rho$  of a few SWIR parameter sets.

### 3.2.2 Sample Size

Returning to the question of sample size, we also find great differences between the SWIR and Coastal ACMs, not only in the average sample size, discussed earlier in section 3.1, but also in the strength of the relationship between sample size and  $\rho$ , shown in Figure 5.

### Figure 5

For the SWIR ACM, the slope ( $b$ ) between  $N$  and  $\rho$  is different from zero ( $b = -.0219$ , 90%,  $CI = -.0404$  to  $-.0034$ ,  $p = 0.1$ ). Despite the visible trend, however, this is not the case for the Coastal ACM, for which the impact of sample size on  $\rho$  is less clear ( $b = -.0058$ , 90%,  $CI = -.0122$  to  $.0064$ ,  $p = 0.1$ .)

Focusing on the seven parameter sets above their regression lines (grey line for Coastal-WCR sets and black line for SWIR-WCR sets), nearly all apply the more rigorous confidence flag settings, with the exception of two from the Coastal-QAA group. In fact, when controlling for the possibly-significant impact of sample size, four of the seven above-average sets are from the Coastal-QAA group. incidentally, these four also have the largest sample sizes  $N$  of the seven sets.

Figure 5 thus offers further evidence of the increased applicability of this ACM-WCR combination in the study region. Evidence presented elsewhere (Table 3 and Figures 3 and 4) already casted doubts about the applicability of SWIR ACM in the study location, at least compared to the Coastal ACM; however, the single high-performing Coastal-OC3 set still cannot be discarded, as the available evidence simply suggests that it is more sensitive than Coastal-QAA to the relaxation of data inclusion criteria (Horion et al. 2010; Knox 2012).

### 3.2.3 Spatial Completeness

Just over two-thirds of the images provided data, with an average pixel yield that is 18% of all possible post-processed data-yielding pixels (1845 pixels). Indeed, lack of data and of spatial coverage overall is a limitation of this study. This can be seen clearly in Figure 6, where the grey background represents possible data-yielding pixels for which data could not be retrieved on that day.

### Figure 6

The first four columns show retrievals from the four ground-truthing match-ups common to all parameter sets (see Table 4), whereas the last column shows one of the best retrieval days. Note that three of the displayed images (both SWIR sets in the 4th column and Coastal-QAA in the 3rd column) failed to provide the needed ground-truth match-up data, which were obtained using temporal interpolation.

In the fifth column of Figure 6, the importance of land proximity also becomes clear, especially for Bukavu and Kabuno Bays. Together with the limited number of *in situ* data points from these two sites, the lack of pixel retrievals helps explain the lack of ground-truth match-ups for either site. Distinct differences between the SWIR and Coastal ACMs are evident in the coloring of Figure 6. Importantly, while not perfect, the Coastal ACM is better able to retrieve coastal pixels and



to minimize the impact of land proximity on the resulting water-leaving radiance than is the SWIR ACM, which suffers from improbably high z-scores for all pixels located closest to the coast. Indeed, the SWIR ACM's apparent susceptibility to adjacency effects due to land proximity is a likely explanation for the long right tail seen in the frequency histograms (see Figure 4, panels b and d). One strength of the Coastal ACM is precisely that it was designed specifically to account for land proximity. In pelagic waters, however, it is difficult to assess the Coastal ACM's apparently greater variability, since Figure 6 shows internally standardized z-scores rather than absolute values. One plausible explanation is that the Coastal ACM's greater robustness to adjacency effects and fewer extreme values amplify the visibility (in z-scores) of biologically relevant differences in true values.

While the lake images in Figure 6 provide visual snapshots, the seasonal component of spatial completeness is shown quantitatively in Figure 7. With values normalized to the maximum number of retrieved pixels for any given day and any parameter set, this figure complements Table 3, showing that the total number of retrieved pixels using the SWIR ACM is indeed half as much as with the coastal ACM. This may be due to limitations in the applicability of the SWIR correction bands over inland waters, including its inability to adequately retrieve the coastal pixels.

### Figure 7

Figure 7 clearly shows a temporal pattern in spatial completeness of the images, especially using the SWIR ACM. While the SWIR sets yield less than half of the average pixels-per-image retrieved by the Coastal sets (see Table 3), the difference per-day is less drastic, at least outside the months of February and JJA season. Lastly, although the two Coastal sets have comparable average completeness per-image (see Table 3), the Coastal-QAA set has the greatest completeness per day across all months.

Here, it becomes clear that no ACM-WCR combination retrieves well in the middle of the rainy season (DJF), despite drier weather around January. Only the Coastal-WCRs retrieve well during the long dry season (JJA), with the Coastal-QAA doing particularly well. Fortunately, the DJF season is unlikely to be the most important for plankton dynamics because blooms are less likely to occur (Sarmiento, Isumbisho, and Descy 2006). Furthermore, from an epidemiological perspective, the cholera case load and human-to-human transmission are increased at this time of year, consequently decreasing the relative impact of environmental factors like plankton dynamics is minimized.

#### 3.2.4 Seasonality and more

Figure 7 already provides a first appreciation of seasonality in spatial completeness, but few final comments can be made about the role of seasonality - and other differences - in the ground-truth match-ups. Such information, found in Table 4, complements and strengthens results discussed earlier.

### Table 4

First, as can be seen in Table 4, there are many differences among sets in the exact dates of their ground-truth match-ups, including the question of whether the match-up was obtained using a same-day pixel or a temporally-interpolated pixel. Additional differences are introduced by the season and location of the match-ups. All of this increases uncertainties of parameter set comparisons.

Namely, although very minor seasonality exists in the number and location of *in situ* measurements, the seasonality in ground-truth match-ups is consistent with the seasonality in remote-sensing data retrieval, shown in Figure 7. Namely, the Coastal ACM, which retrieves best in JJA, has an increased number of match-ups from this season while the SWIR ACM has none, and match-ups are nearly absent from the DJF season. Further, seasonality in the ground-truth match-ups also interacts with site location and no ACM-WCR combination retrieves well in the DJF season. Indeed, the low, but seasonally variable, numbers of ground-truth match-ups can be almost entirely attributed to the seasonality in retrievals for any given ACM-WCR combination.

#### 4. Conclusion

This is the first study to employ the SWIR atmospheric correction algorithm and the QAA water-constituents retrieval algorithm in an African Great Lake. That said, although SWIR produced the highest  $\rho$  values, its spatial completeness does not afford adequate large-scale spatiotemporal information about plankton dynamics in Lake Kivu. Indeed, its utility may be limited to specific times and places where its retrieval capacity is maximized, such as during the March-April-May or September-October-November seasons. Ultimately, the coastal atmospheric correction model with 90% relative humidity appears to be more appropriate for Lake Kivu.

It is more difficult to decide between Coastal-QAA and Coastal-OC3. This study highlights the robustness and potential utility of the QAA algorithm for inland lake settings. Indeed, the QAA algorithm is an interesting alternative to *chl-a* and should be considered for studies of limnologic primary productivity or plankton biomass as it may, in some cases, be more appropriate, especially for tropical inland lakes. Moreover, since QAA decomposes reflectance into apparent and inherent optical properties, it distinguishes between plankton, detritus and other water constituents, and it does so at each wavelength. QAA thus offers much more than what we could investigate here, having only incorporated plankton absorption at one wavelength. Its use in applications of inland water components retrieval deserves further exploration. However, when appropriate data quality standards can be applied, the OC3 algorithm may actually be higher-performing. Therefore, any study planning to use these data should consider both the best-performing Coastal-QAA and the best-performing Coastal-OC3 parameter sets and weigh the pros and cons prior to selection.

In spite of the difficult remote sensing conditions on Lake Kivu, notably the missing data and the adjacency effects from land and cloud cover, our study may serve as a reference for others attempting similar study in difficult conditions. Furthermore, our results may still prove useful in future studies due to the absence of other continuous data sources. Our study may not be directly comparable to those using  $R^2$  as the measure of ground-truthing performance, but for reference, Chavula et al. (2009), working on Lake Malawi, obtained  $R^2 = .6$  ( $N = 5$ ) using a band ratio of normalized reflectances, while Horion et al. (2010), working on Lake Tanganyika, obtained  $R^2 = .64$  ( $N = 12$ ) using the Coastal-OC3 combination and, notably, a minimum pixel threshold of 20% which would have been prohibitive in our study. As might be imagined, these other AGL studies also encountered significant challenges in retrieving sufficient high-quality data.

When attempting to optimize a remotely-sensed proxy in conditions of limited data quantity and quality, a balance must be found, simultaneously excluding the

maximum amount of low-quality data while still allowing for sufficient data quantity to perform ground-truth analyses. Estimating plankton biomass concentrations in Lake Kivu can potentially have important implications for spatially-explicit cholera models(Gatto et al. 2012; Mari et al. 2012; Rinaldo et al. 2012; Bertuzzo et al. 2014) that rely on the description of pathogen dynamics in environmental waters to characterize the force of the infection of the disease(Codeco 2001). A remotely sensed clue regarding pathogen concentrations, along with other climatic and environmental drivers(Finger et al. 2014), would crucially improve our ability to predict and, to some extent, control cholera dynamics in the endemic Lake Kivu region.

The authors of this paper declare no conflict of interest.

#### **Acknowledgements**

We thank colleagues for their support, especially B. Schaeffli, D. Tuia and G. Moser, as well as the experts on the SeaDAS forum. We also thank J-P Descy, F. Darchambeau, M. Isumbisho, and H. Sarmento for graciously providing the *in situ* data, which they produced thanks to financial support from several Belgian institutions: Commission Universitaire pour le Développement, Fonds Léopold III pour l'Exploration et Conservation de la Nature, Fondation pour Favoriser les Recherches Scientifiques en Afrique, Coopération Technique Belge, and the International Foundation for Science and the International Society of Limnology.

## References

- Ahmad, Z., B.A. Franz, C.R. McClain, E.J. Kwiatkowska, J. Werdell, E.P. Schettle, and B.N. Holben. 2010. "New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans." *Applied Optics* 49 (29): 5545–5560.
- Asplund, M.E., A-S. Rehnstam-Holm, V. Atnur, P. Raghunath, V. Saravanan, K. Härnström, B. Collin, I. Karunasagar, and A. Godhe. 2011. "Water column dynamics of *Vibrio* in relation to phytoplankton community composition and environmental conditions in a tropical coastal area." *Environmental Microbiology* (13): 27382751.
- Bailey, S.W., B.A. Franz, and P.J. Werdell. 2010. "Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing." *Optics Express* 18 (7): 7521–7527.
- Bertuzzo, E., F. Finger, L. Mari, M. Gatto, and A. Rinaldo. 2014. "On the probability of extinction of the Haiti cholera epidemic." *Stochastic Environmental Research and Risk Assessment* (online).
- Bompangue, D., P. Giraudoux, P. Handschumacher, M. Piarroux, B. Sudre, M. Ekwanzala, I. Kebela, and R. Piarroux. 2008. "Lakes as sources of cholera outbreaks, Democratic Republic of Congo." *Emerging Infectious Diseases* 14 (5): 798–800.
- Bompangue, D., P. Giraudoux, M. Piarroux, G. Mutombo, R. Shamavu, B. Sudre, A. Mutombo, V. Mondonge, and R. Piarroux. 2009. "Cholera epidemics, war and disasters around Goma and Lake Kivu: an eight-year study." *PLOS Neglected Tropical Diseases* 3 (5): 3436.
- Bompangue, D., P. Giraudoux, P-D. Plisnier, A.M Tinda, M. Piarroux, B. Sudre, S. Horion, J.M. Tamfum, B.K. Llunga, and R. Piarroux. 2011. "Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978-2008." *Emerging Infectious Diseases* 17 (11): 2026–2036.
- Chavula, G., P. Brezonik, P. Thenkabail, T. Johnson, and M. Bauer. 2009. "Estimating chlorophyll concentration in Lake Malawi from MODIS satellite imagery." *Physics and Chemistry of the Earth* 34: 756–760.
- Codeco, C. 2001. "Endemic And Epidemic Dynamics Of Cholera: The Role Of The Aquatic Reservoir." *Bmc Infectious Diseases* 1.
- De Magny, G.C., R. Murtugudde, M. Sapiano, A. Nizam, C.W. Brown, A.J. Busalacchi, M. Yunus, et al. 2008. "Environmental signatures associated with cholera epidemics." *Proceedings of the National Academy of Sciences* 105 (46): 17676–17681.
- Finger, F., A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2014. "Cholera in the lake Kivu region (DRC): integrating remote sensing and spatially-explicit epidemiological modeling." *Water Resources Research* Forthcoming.
- Ford, T.E., R.R. Colwell, J.B. Rose, S.S. Morse, D.J. Rogers, and T.L. Yates. 2009. "Using satellite images of environmental changes to predict infectious disease outbreaks." *Emerging Infectious Diseases* 15 (9).
- Gatto, M., L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. "Generalized reproduction numbers and the prediction of patterns in waterborne disease." *Proceedings of the National Academy of Sciences* 109 (48): 19703–19708.
- Gurlin, D., A.A. Gitelson, and W.J. Moses. 2011. "Remote estimation of chl-a concentration in turbid productive waters return to a simple two-band NIR-red model?." *Remote Sensing of Environment* 115: 3479–3490.
- Horion, S., N. Bergamino, Y. Cornet, J-P. Descy, and P-D. Plisnier. 2010. "Optimized extraction of daily bio-optical time series derived from MODIS/Aqua imagery for Lake Tanganyika, Africa." *Remote Sensing of Environment* 114: 781–791.
- Hu, C., Z. Lee, and B. Franz. 2012. "Chlorophyll a algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference." *Journal of Geophysical Research* 117: C010111.
- Jutla, A., A.S. Akanda, A. Huq, A. Syed, G. Faruque, R. Colwell, and S. Islam. 2013. "A water marker monitored by satellites to predict seasonal endemic cholera." *Remote*

- Sensing Letters* 4: 822–831.
- Jutla, A.S., A.S. Akanda, and S. Islam. 2010. “Tracking Cholera In Coastal Regions Using Satellite Observations.” *Journal of the American Water Resources Association* 46 (4): 651–662.
- Knox, A. 2012. “Endemic cholera in inlandAfrica: a study on Lake Kivu, DRC.” Master’s thesis. University of Lausanne.
- Lee, Z.P., K. L. Carder, and R.A. Arnone. 2002. “Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters.” *Applied Optics* 41: 5755–5772.
- Lee, Z.P., K. L. Carder, J. Marra, R. G. Steward, and M.J. Perry. 1996. “Estimating primary production at depth from remote sensing.” *Applied Optics* 35: 463–474.
- Lee, Z.P., B. Lubac, J. Werdell, and R. Arnone. 2009. *An update of the quasi-analytical algorithm (QAA.v5)*. Tech. rep.. Mississippi State University; Ocean Biology Processing Group, GSFC, NASA; Stennis Space Center.
- Lobitz, B., L. Beck, A. Huq, B. Wood, G. Fuchs, A.S.G. Faruque, and R.R. Colwell. 2000. “Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement.” *Proceedings of the National Academy of Sciences* 97 (4): 1438–1443.
- Lyon, P.E., F.E. Hoge, C.W. Wright, R.N. Swift, and J.K. Yungel. 2004. “Chlorophyll biomass in the global oceans: satellite retrieval using inherent optical properties.” *Applied Optics* 43 (31): 5886–5892.
- Mari, L., E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. “Modelling Cholera Epidemics: The Role Of Waterways, Human Mobility And Sanitation.” *Journal Of The Royal Society Interface* 9: 376–388.
- Naik, Puneeta, E.J. D’Sa, J.I. Goés, and H. do R. Gomes. 2009. “Particulate absorption properties from MODIS ocean color and four in-situ transects in the southeastern Bering Sea shelf during July, 2008.” *Proceedings of SPIE* 7473: 747302.
- Ocean Color Biology Group. 2012. “OceanColor Homepage and Forum.” <http://oceancolor.gsfc.nasa.gov/>.
- Odermatt, D., A. Gitelson, V.E. Brando, and M. Schaepman. 2012. “Review of constituent retrieval in optically deep and complex waters from satellite imagery.” *Remote Sensing of Environment* 118: 116–126.
- Olago, D., M. Marchall, and S.O. Wandiga. 2007. “Climatic, socio-economic, and health factors affecting human vulnerability to cholera in the Lake Victoria Basin, East Africa.” *Ambio* 36 (4): 350–358.
- Rinaldo, A., E. Bertuzzo, L. Mari, L. Righetto, M. Blokesch, M. Gatto, R. Casagrandi, M. Murray, S. Vesenbeckh, and I. Rodriguez-Iturbe. 2012. “Reassessment Of The 2010–2011 Haiti Cholera Outbreak And Rainfall-Driven Multi-Season Projections.” *Proceedings Of The National Academy Of Sciences* 109: 6602–6607.
- Russell, P.B., R.W. Bergstrom, Y. Shinozuka, A.D. Clarke, P.F. DeCarlo, J.L. Jimenez, J.M. Livingston, J. Redemann, O. Dubovik, and A. Strawa. 2010. “Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition.” *Atmospheric Chemistry and Physics* 10: 1155–1169.
- Sarmiento, H., J-P. Descy, and F. Darchambeau. 2012. “Phytoplankton.” In *Lake Kivu limnology and biogeochemistry of a tropical great lake*, edited by J-P. Descy, F. Darchambeau, and M. Schmid. Aquatic Ecology Series. chap. 5, 188. Springer.
- Sarmiento, H., M. Isumbisho, and J-P. Descy. 2006. “Phytoplankton ecology of Lake Kivu (Eastern Africa).” *Journal of Plankton Research* 28 (9): 815–829.
- Shang, S., Q. Dong, Z. Lee, Y. Li, Y. Xie, and M. Behrenfeld. 2011. “MODIS observed phytoplankton dynamics in the Taiwan Strait: an absorption-based analysis.” *Biogeosciences* 8: 841–850.
- Tamplin, M.L., A.L. Gauzens, A. Huq, D.A. Sack, and R.R. Colwell. 1990. “Attachment of *Vibrio cholerae* serogroup 01 to zooplankton and phytoplankton of Bangladesh waters.” *Applied and Environmental Microbiology* 56 (6): 1977–1980.
- Wang, M., and W. Shi. 2007. “The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing.” *Optics Express* 15 (24): 15722.

- Wang, M., W. Shi, and J. Tang. 2011. "Water property monitoring and assessment for China's inland Lake Taihu from MODIS-Aqua measurements." *Remote Sensing of Environment* 115: 841–854.
- Wang, M., SH. Son, and W. Shi. 2009. "Evaluation of MODIS SWIR and NIR-SWIR atmospheric correction algorithms using SeaBASS data." *Remote Sensing of Environment* 113: 635–644.
- Werdell, P.J., B.A. Franz, and S.W. Bailey. 2010. "Evaluation of shortwave infrared atmospheric correction for ocean color remote sensing of Chesapeake Bay." *Remote Sensing of Environment* 114: 2238–2247.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
Parameters	ACM	WCR	Minimum pixel yield	Flag setting	Maximum albedo
Coded (+)	Coastal	OC3	5%	Less exclusive	0.054
Coded (-)	SWIR	QAA	15%	More exclusive	0.081

Table 1. **Parameter Values.** For each parameter set and each parameter  $X_i$  in that set ( $i$  being an integer from 1 to 5), the two true values of the parameter are coded to (+) or (-) and entered into the experimental matrix.

Flag	Description	Coded Value	
		(+)	(-)
1. ATMFAIL	Atmospheric correction failure	✓	✓
3. PRODWARN*	One or more product warnings		✓
4. HIGLINT	High sun glint	✓	✓
6. HISATZEN	High sensor view zenith angle		✓
10. CLDICE	Probable cloud or ice contamination		✓
11. COCCOLITH	Coccolithofores detected		✓
12. TURBIDW*	Turbid water detected		✓
13. HISOLZEN	High solar zenith	✓	✓
15. LOWLW	Very low nLw (cloud shadow)		✓
16. CHLFAIL	Derived product algorithm failure	✓	✓
17. NAVWARN	Navigation quality is reduced		✓
18. ABSAER*	Possible absorbing aerosol		✓
20. MAXAERITER	Aerosol iterations exceeded max		✓
22. CHLWARN	Derived product quality is reduced	✓	✓
23. ATMWARN	Atmospheric correction is suspect	✓	✓
*SeaDAS flags not applied by default in standard ocean color processing.			

Table 2. **Confidence Flags ( $X_4$ )**. The flags incorporated into the more permissive (+) and more restrictive (-) parameter values are noted here. Whenever a flag is applied, it excludes data points obtained under the circumstances defined by the flag. All flags except three (asterisked) are SeaDAS default flags for ocean color processing (Ocean Color Biology Group 2012). Four additional default flags were omitted from analysis: flags 5 (saturated radiance) and 9 (straylight) due to over-masking, and flags 2 (land mask) and 26 (bad navigation) due to apparent flagging errors in the context of an inland lake. The two tested flag combinations were chosen based on preliminary experiments (Knox 2012).



Group	Set	Parameters					Main Results					
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	p	p-value	N	I <sup>a</sup>	Day <sup>b</sup>	Image <sup>c</sup>
	<b>1-2</b>	-	-	-	-	+/-	<b>0.74</b>	<b>&lt;.01</b>	<b>11</b>	<b>0.45</b>	<b>0.64</b>	<b>0.09</b>
SWIR-	3-4	-	-	-	+	+/-	0.40	<.05	19	0.47	0.74	0.16
QAA	5-6	-	-	+	-	+/-	0.74	<.01	11	0.45	0.64	0.09
	7-8	-	-	+	+	+/-	0.38	<.05	20	0.45	0.75	0.16
	9-10	-	+	-	-	+/-	0.60	n.s	6	0.67	0.67	0.08
SWIR-	11-12	-	+	-	+	+/-	0.49	n.s	10	0.60	0.80	0.15
OC3	<b>13-14</b>	-	+	+	-	+/-	<b>0.75</b>	<b>&lt;.01</b>	<b>11</b>	<b>0.45</b>	<b>0.64</b>	<b>0.09</b>
	15-16	-	+	+	+	+/-	0.48	<.05	16	0.44	0.75	0.15
	<b>17-18</b>	+	-	-	-	+/-	<b>0.47</b>	<b>&lt;.05</b>	<b>24</b>	<b>0.58</b>	<b>0.67</b>	<b>0.20</b>
Coastal-	19-20	+	-	-	+	+/-	0.37	<.05	37	0.57	0.57	0.24
QAA	21-22	+	-	+	-	+/-	0.45	<.05	25	0.56	0.68	0.19
	23-24	+	-	+	+	+/-	0.36	<.05	39	0.54	0.56	0.23
	<b>25-26</b>	+	+	-	-	+/-	<b>0.62</b>	<b>&lt;.05</b>	<b>9</b>	<b>0.67</b>	<b>0.89</b>	<b>0.21</b>
Coastal-	27-28	+	+	-	+	+/-	0.31	n.s	14	0.57	0.64	0.27
OC3	29-30	+	+	+	-	+/-	0.35	n.s	23	0.61	0.65	0.23
	31-32	+	+	+	+	+/-	0.28	n.s	36	0.58	0.56	0.29
<b>Averaged Results</b>							<b>0.49</b>	<b>11/16 sig.</b>	<b>19</b>	<b>0.54</b>	<b>0.68</b>	<b>0.18</b>
SWIR Only							0.57	6/8 sig.	13	0.50	0.70	0.12
Coastal Only							0.40	5/8 sig.	26	0.58	0.65	0.23
QAA Only							0.49	8/8 sig.	23	0.51	0.65	0.17
OC3 Only							0.49	3/8 sig.	16	0.57	0.70	0.18

Table 3. **Parameter Set Performances.** The four parameter sets chosen to represent each ACM-WCR group are in bold. <sup>a</sup>The proportion of match-ups from Ishungu; <sup>b</sup>the proportion generated as same-day match-ups without temporal interpolation, and <sup>c</sup>the proportion of data-yielding pixels, obtained by dividing the number of data-yielding pixels by the number of possible data-yielding pixels.

Date	Site	Matched <i>in situ</i> <sup>b</sup>	SWIR- QAA	SWIR- OC3	Coastal- QAA	Coastal- OC3
2002-08-20	Ishungu	0.9201			0.0550	
2002-10-15	Ishungu	1.1911	0.0573	0.9491	0.0304	0.3997
2003-01-27	Western	2.2641	0.6045	21.5564	0.0502	
2003-04-01	Ishungu	4.5364	0.6858	14.9965	0.2418	3.4912
2003-08-05	Ishungu	3.8565			0.0660	1.0016
2003-09-02	Ishungu	3.5345			0.0564	
2004-07-07	Ishungu	3.1844			0.0495	0.8035
2004-11-23	Ishungu	3.7171	1.5686	49.6493	0.0591	
2005-06-21	Kibuye	3.2513			0.0414	
2005-07-05	Ishungu	1.5689			0.0452	
2005-07-05	Kibuye	2.3571			0.0386	
2005-08-05	Ishungu	1.3236			0.0474	
2005-10-25	Kibuye	2.1719	0.0703	1.8019	0.0482	
2005-11-08	Kibuye	1.6296	0.1327	5.3299	0.0655	1.4060
2006-02-03	Kibuye	2.4725			0.0648	1.2598
2006-04-03	Ishungu	2.3848	1.2412	36.0386	0.0945	
2006-04-27	Kibuye	1.7183	0.1187	3.1222	0.0876	
2006-05-09	Ishungu	3.6007	1.1775	31.5146	0.1371	4.1648
2006-06-13	Kibuye	0.4962			0.0530	
2006-06-20	Ishungu	1.9938			0.0678	
2006-07-04	Kibuye	1.0696			0.0499	0.9314
2006-11-21	Kibuye	1.7366	0.2543	5.3961		
2007-06-19	Ishungu	1.3200			0.0433	
2008-06-03	Kibuye	0.7697	0.2574	4.1221	0.0405	
2008-07-03	Ishungu	1.3428			0.0608	0.9738

Months	All <i>in situ</i> ( <i>I</i> <sup>a</sup> )	Ground-truth match-ups ( <i>I</i> <sup>a</sup> )			
All Year	255 (.62)	11 (.45)	11 (.45)	24 (.58)	9 (.67)
DJF	56 (.59)	1 (0)	1 (0)	2 (0)	1 (0)
MAM	73 (.67)	4 (.75)	4 (.75)	4 (.75)	2 (1.00)
JJA	72 (.64)	1 (0)	1 (0)	13 (.62)	4 (.75)
SON	54 (.57)	5 (.40)	5 (.40)	5 (.60)	2 (.50)

Table 4. **Selected Ground-Truth Match-Ups.** Details of the specific ground-truth match-ups are shown with the true values (i.e. the absolute plankton absorption or *Chl-A* values) from the best representative of each ACM-WCR group. Italicized values were obtained using the temporal interpolation described in section 2.3. Distilling the raw information about match-ups into seasonal components, the number and location of match-ups is also shown, where months are grouped so as to maximize seasonality. <sup>a</sup>The proportion of match-ups from Ishungu by season; <sup>b</sup>*In situ* data were provided through the work of Sarmiento, Isumbisho, and Descy (2006) and Sarmiento, Descy, and Darchambeau (2012).

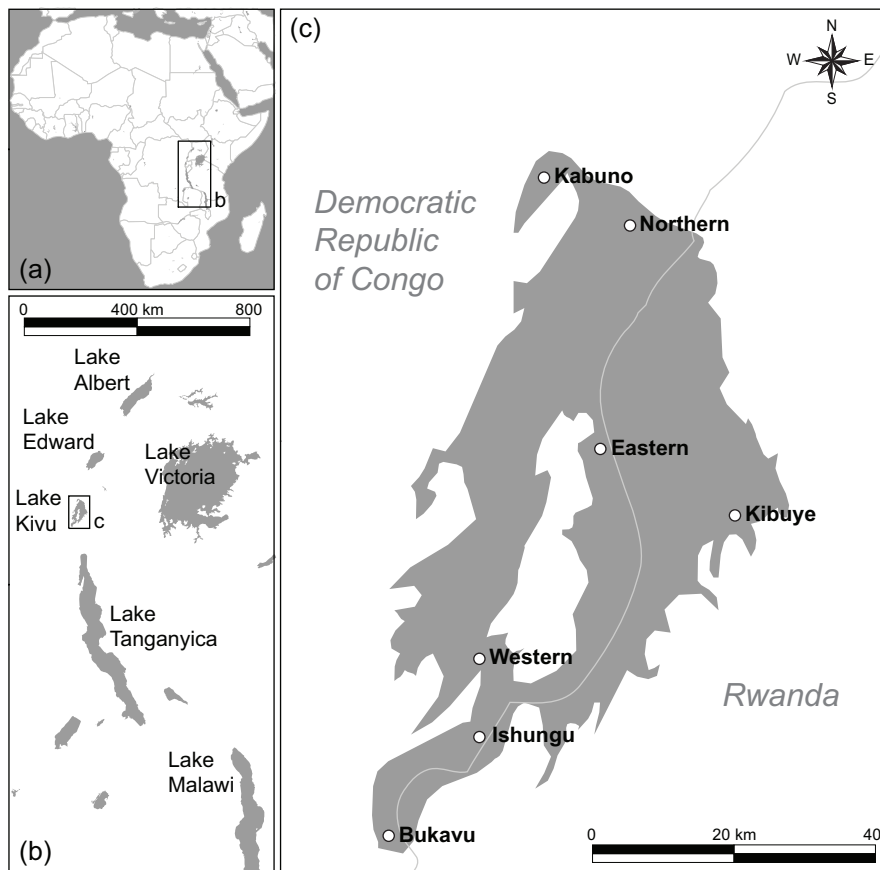


Figure 1. **Study Location.** (a) The African Great Lakes are found in East Africa, many along the Albertine Rift (western branch) of the East African Rift System. (b) Lake Kivu is one of the smaller AGL, located to the north of Lake Tanganyika and west of Lake Victoria. (c) Lake Kivu is situated on the border of the Democratic Republic of Congo to the West and Rwanda to the East. Data from seven pelagic *in situ* Chl-A sampling points were available for this study.

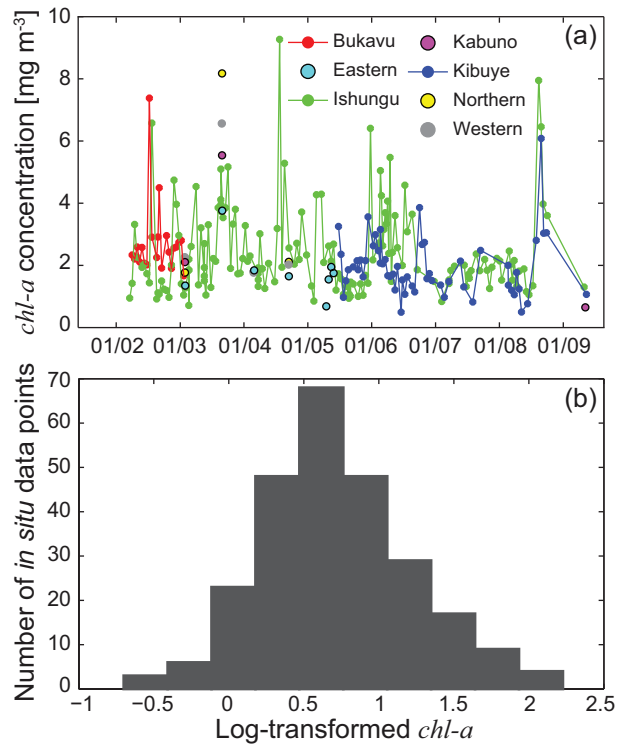


Figure 2. *In Situ Data.* (a) Timeline of true *chl-a* values obtained *in situ*, where the three sites with the most continuous data are drawn with a line to connect the dots while the others are plotted as unconnected dots. (b) Frequency histogram the normally-distributed log *in situ* data.

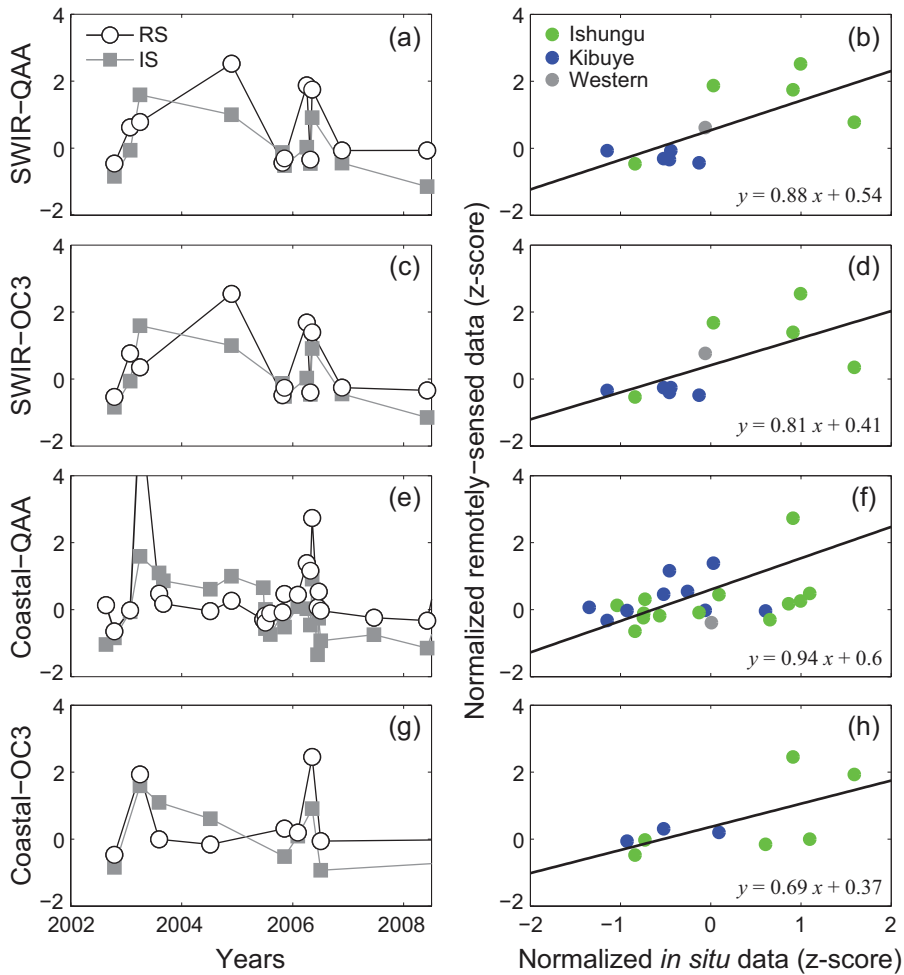


Figure 3. **Ground-Truthing Results.** Results of the center-reduced (z-scores) matched data for each of the four chosen ACM-WCR combinations. (a), (c), (e), and (g): Timelines of the z-scores of remotely-sensed data in open circles and z-scores of *in situ* data in closed squares. Note that the line connecting sequential ground-truth match-ups is included as a visual aid only. (b), (d), (f), and (h): Regressions of the z-scores of match-ups are colored according to their site location.

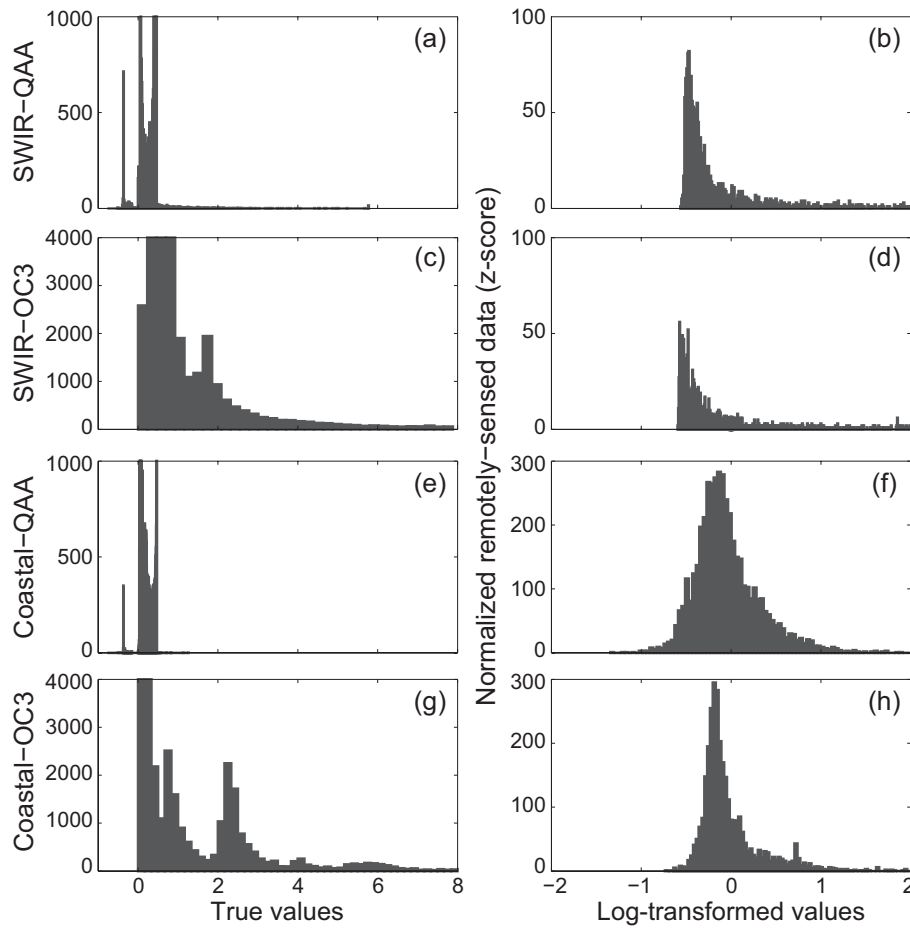


Figure 4. **Impact of Post-Processing.** Frequency histograms of the four chosen ACM-WCR combinations were, for computational reasons, generated using representative sub-samples of every fifth pixel latitude, fourth longitude and fifth in time. (a), (c), (e), and (g) show the distribution of true values before post-processing; these values would ideally be left-skewed with a positive maximum. Unsurprisingly, with the possible exception of SWIR-OC3, none meet that criteria, signifying the necessity of post-processing. (b), (d), (f), and (h) show the log-transformed values after post-processing; plotted on a log-scale, these values would ideally appear normally distributed, as is the case for the two Coastal sets in (f) and (h).

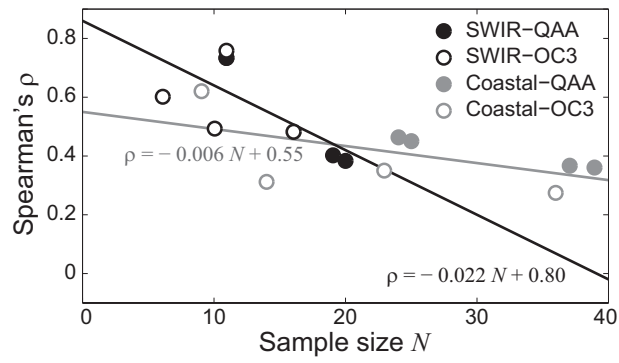


Figure 5. **Impact of Sample Size.** All parameter sets are plotted according to the sample size  $N$  and  $\rho$ , with SWIR sets in black and Coastal sets in grey, where closed circles are used for QAA and open circles for OC3. Since ACM is clearly an important determinant in the impact of  $N$  on  $\rho$ , two regression lines have been drawn, one each for SWIR (black) and Coastal (grey) sets.

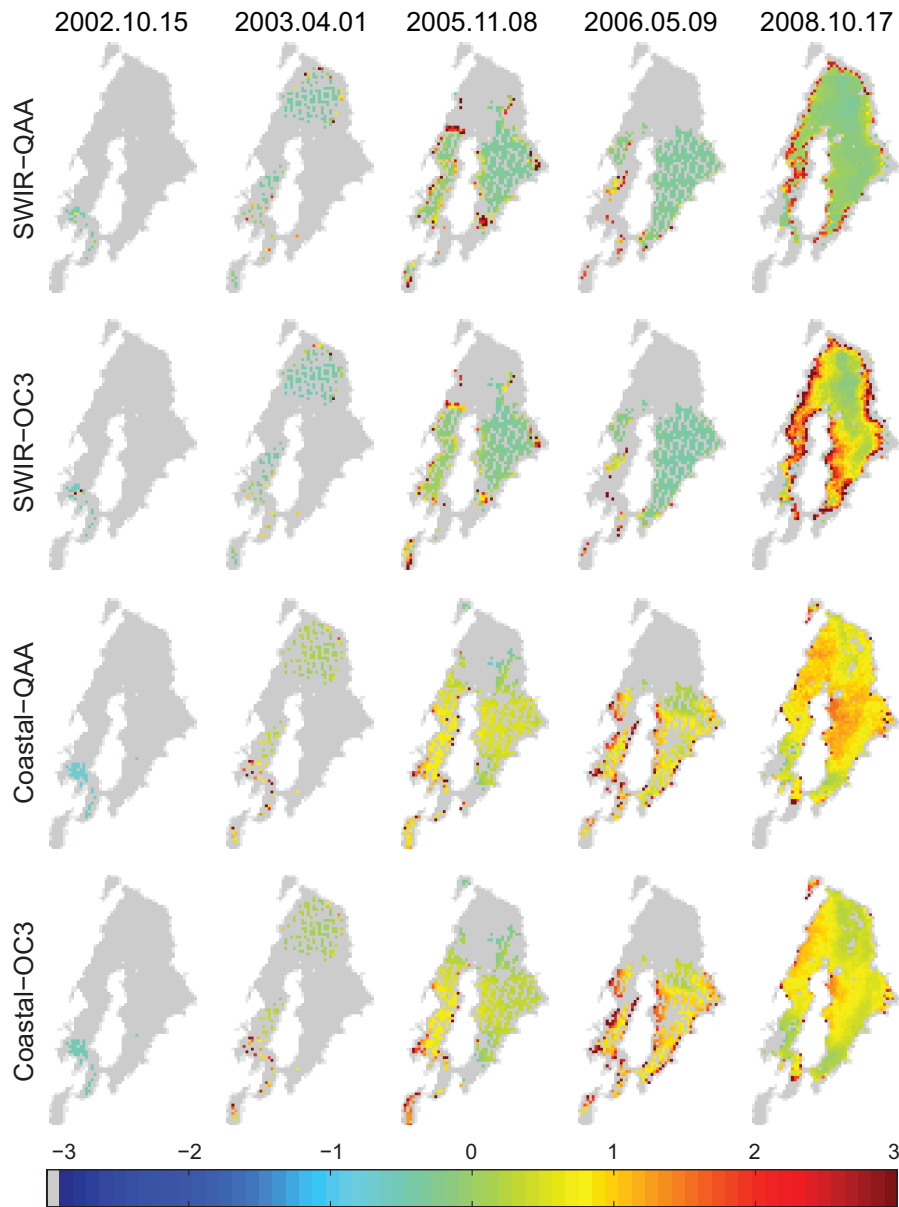


Figure 6. **Retrievals for Lake Kivu.** Standardized lake images (z-scores) where the colorbar show plankton proxies relative to data distribution of the selected ACM-WCR combinations. The first four columns show images typical of those use for ground-truthing while the fifth column shows one of the best imaging days, for which the *in situ* data was not available.



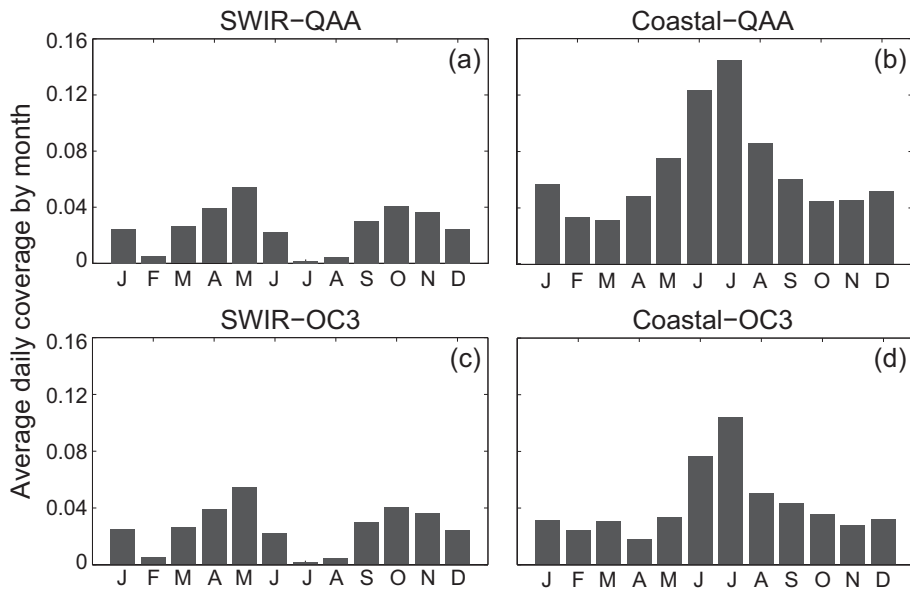


Figure 7. **Seasonality of Spatial Completeness.** Daily spatial completeness averaged by month, normalized to the product of the maximum unique data-yielding pixels and the number of days of a given month across the assessed timespan (July 6, 2002 to July 5, 2009). Unlike in Table 1, which shows average completeness per data-yielding image, the normalization used here enables a consideration for the total actual monthly pixel yield (i.e. the number of pixels per image and images per month) divided by the theoretical maximum monthly pixel yield (i.e. the maximum possible coverage multiplied by the number of days per month). (a) SWIR-QAA, (b) Coastal-QAA, (c) SWIR-OC3 and (d) Coastal-OC3.