

## Article

# Developing a Comprehensive Spectral-Biogeochemical Database of Midwestern Rivers for Water Quality Retrieval Using Remote Sensing Data: A Case Study of the Wabash River and Its Tributary, Indiana

Jing Tan <sup>1,\*</sup>, Keith A. Cherkauer <sup>1</sup> and Indrajeet Chaubey <sup>1,2</sup>

<sup>1</sup> Agricultural and Biological Engineering, Purdue University, 225 South University Street, West Lafayette, IN 47906, USA; cherkaue@purdue.edu (K.A.C.); ichaubey@purdue.edu (I.C.)

<sup>2</sup> Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

\* Correspondence: tan30@purdue.edu; Tel.: +1-765-491-0544

Academic Editors: Richard Gloaguen and Prasad S. Thenkabail

Received: 16 March 2016; Accepted: 14 June 2016; Published: 21 June 2016

**Abstract:** A comprehensive spectral-biogeochemical database was developed for the Wabash River and the Tippecanoe River in Indiana, United States. This database includes spectral measurements of river water, coincident *in situ* measurements of water quality parameters (chlorophyll (chl), non-algal particles (NAP), and colored dissolved organic matter (CDOM)), nutrients (total nitrogen (TN), total phosphorus (TP), and dissolved organic carbon (DOC)), water-column inherent optical properties (IOPs), water depths, substrate types, and bottom reflectance spectra collected in summer 2014. With this dataset, the temporal variability of water quality observations was first analyzed and studied. Second, radiative transfer models were inverted to retrieve water quality parameters using a look-up table (LUT) based spectrum matching methodology. Results found that the temporal variability of water quality parameters and nutrients in the Wabash River was closely associated with hydrologic conditions. Meanwhile, there were no significant correlations found between these parameters and streamflow for the Tippecanoe River, due to the two upstream reservoirs, which increase the settling of sediment and uptake of nutrients. The poor relationship between CDOM and DOC indicates that most DOC in the rivers was from human sources such as wastewater. It was also found that the source of water (surface runoff or combined sewer overflow (CSO)), water temperature, and nutrients were important factors controlling instream concentrations of phytoplankton. The LUT retrieved NAP concentrations were in good agreement with field measurements with slope close to 1.0 and the average estimation error was 4.1% of independently obtained lab measurements. The error for chl estimation was larger (37.7%), which is attributed to the fact that the specific absorption spectrum of chl was not well represented in this study. The LUT retrievals for CDOM experienced large variability, probably due to the small data range collected in this study and the insensitivity of  $R_{rs}$  to CDOM change. It is concluded that the success of the LUT method requires accurate spectral measurements and enough *a priori* information of the environment to construct a representative database for water quality retrieval. Therefore, future work will focus on continuing data collection in other seasons of the year and better characterization of the study area.

**Keywords:** water quality; radiative transfer models; look-up table; hyperspectral remote sensing

## 1. Introduction

Remote sensing provides a practical means for synoptic and multi-temporal monitoring of water quality. The water leaving signals that are captured by remote sensing instruments contain essential information on the constituents in the water column, and if applicable, water column depths and bottom properties. The potential of remote sensing to retrieve water quality parameters, bathymetry, and substrate type/composition has been studied for over two decades and there are four main approaches used: empirical, semi-empirical, analytical and radiative transfer methods. Significant attention has been paid to the empirical approach, which focuses on developing best-fit correlational models between remote sensing data (digital numbers, radiance, or reflectance) and measured water quality parameters [1]. A summary of empirical models for water quality assessment can be found in [2]. Instead of screening all wavelengths and finding the band combinations showing the highest correlation, the semi-empirical approach incorporates the spectral characteristics of the interested parameters into the statistical relationship development. For example, many previous studies have shown that the reflectance trough at ~670 nm and the scattering peak at ~700 nm can be used to develop successful models for chlorophyll (chl) estimation and that the scattering peak at ~700 nm is strongly correlated with concentrations of total suspended sediments (TSS) [3,4]. The analytical approach is based on the physical relationship between the inherent optical properties (IOPs) of the water column and measured apparent optical properties (AOPs). The IOPs are the properties of the medium itself and are not affected by the ambient light field. The AOPs are radiometric quantities that display enough stability and can be used for approximately describing the optical properties of the water body, e.g., the remote sensing reflectance. Remote sensing data can be inverted by using the analytical modeling approach to retrieve water column properties and bottom depths [5,6]. In the radiative transfer modeling approach, the software package HydroLight [7] is often required due to the heavy computation required for simulating the complexities of underwater light transfer processes through solving the full set of radiative transfer equations. While analytical models are typically developed by simplifying the full radiative transfer equations based on a set of given assumptions, e.g., level water surface or no internal light sources, the radiative transfer models do not have such constraints. The radiative transfer models can be inverted to extract water column and bottom properties from remote sensing data by using a look-up table based spectrum matching (LUT) methodology [8,9]. Such models have been successfully applied to coral reef mapping in the work of Lesser and Mobley [10].

The empirical and semi-empirical models developed for water quality assessment are often highly dependent on the data and limited to the locations where data are collected. In contrast, the analytical and radiative transfer models provide physical insight into how environmental conditions (such as water column properties, bottom properties, and sky and water surface conditions) quantitatively affect the water leaving signals. Therefore these physics-based models have several advantages: (1) they are repeatable given appropriate inputs from the sites studied; (2) they are easily transferrable between data collected by a variety of sensors; and (3) sensitivity and uncertainty of the models can be objectively determined [11]. The existence of optically complex waters [12] and those that are so shallow that water-leaving reflectance includes interference from bottom conditions [13,14] also necessitates these methods. Tan *et al.* [15] investigated the capabilities of Hyperion imagery for mapping the water quality conditions in river plumes at Lake Michigan. By studying the physical relationship between IOPs and AOPs, the spatial heterogeneity of water quality was adequately captured, which would be challenging for traditional *in situ* sampling or empirical modeling given the limited sample size and complex optical features.

Despite all the advantages described above, the success of these physics-based approaches depends on two requirements: (1) remote sensing reflectance spectra must be accurately measured; and (2) model inputs including the depth, bottom reflectance, and water IOPs must be accurate for the sites of interest [9]. While much attention has been paid to collecting coincident measurements of IOPs and AOPs for oceanic systems (e.g., NASA SeaWiFS Bio-optical Archive and Storage System, <http://seabass.gsfc.nasa.gov/>), inland waters, especially river systems, have been poorly observed,

even though what happens in ocean and coastal waters is highly dependent on these systems [16]. It is therefore important and necessary to develop a similar database/archive for bio-optical data of inland waters and make it accessible to the whole scientific community [17]. Such a database will provide valuable data for improving satellite algorithm development and product validation. In addition, the observations of IOPs in the database will also provide a fundamental linkage between the optical properties and the biogeochemical state of inland waters. For example, the change in beam attenuation is closely associated with particle size variations and can be used to study particle composition [18]. Especially for rivers that experience nutrient and sediment loads from terrestrial sources, the measurements of IOPs, when combined with climate and hydrologic flow regime, enable a better understanding of the biogeochemical state of river systems.

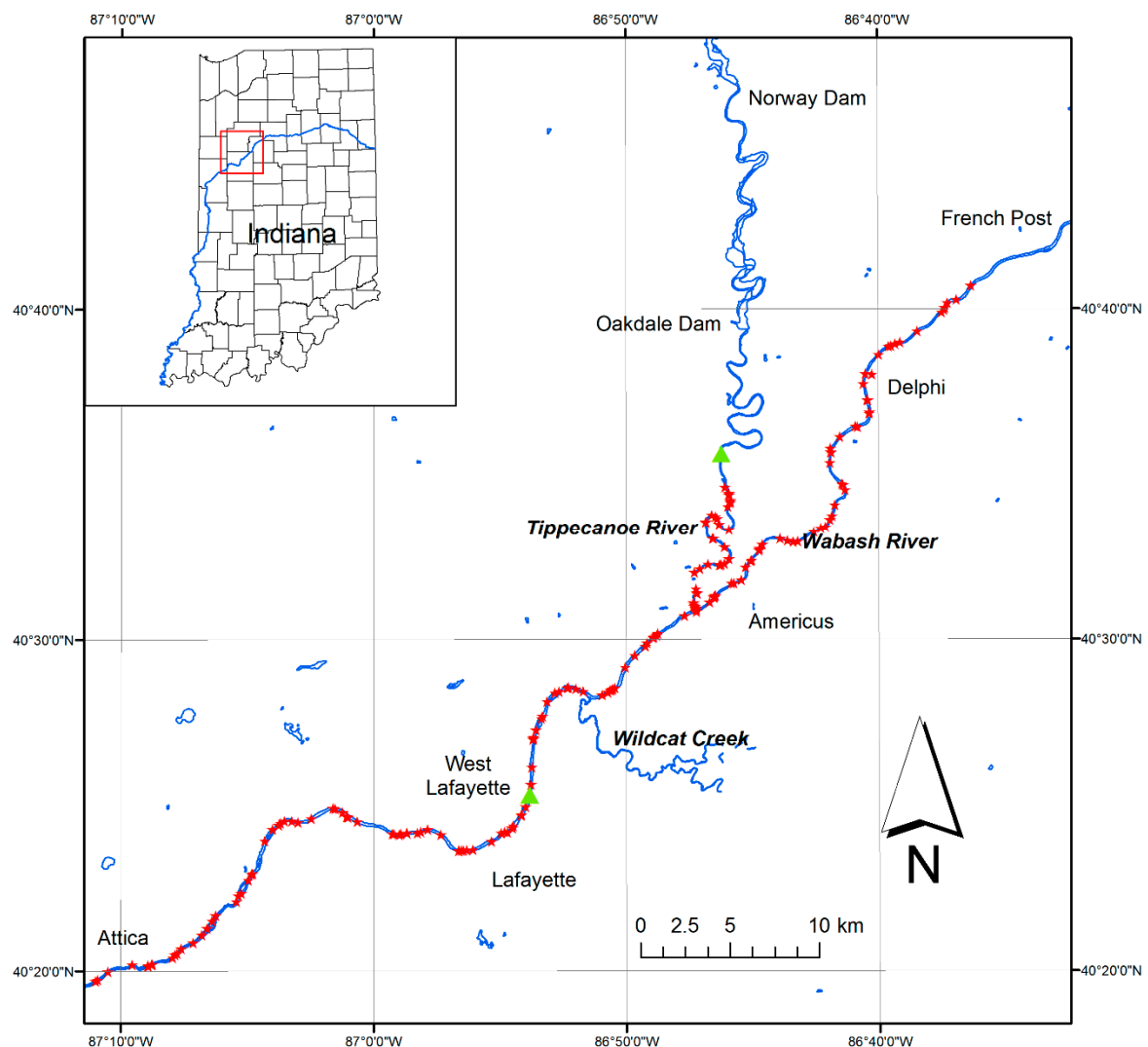
Recent years have seen growing interest in the development of hyperspectral imagers and in the application of hyperspectral data for water quality retrieval. Hyperspectral sensors typically collect data in narrow, contiguous spectral bands and are expected to yield advantages in estimation accuracy due to their ability to finely parse the visible spectrum. Lee and Carder [19] investigated how the number of spectral bands affected the retrieval of water column and bottom properties from remote sensing data and found that hyperspectral data performed much better for optically shallow waters. Although not designed for water targets, the satellite borne Hyperion imager is valid for adequately estimating water quality in coastal and estuary waters as well as the Great Lakes [15,20]. The Hyperspectral Imager for the Coastal Ocean (HICO) is the first hyperspectral sensor designed specifically for the coastal ocean and estuarial, riverine, or other shallow-water areas with optimized Signal-to-Noise Ratio (SNR) [21]. It has been successfully applied for the study of phytoplankton, colored dissolved organic matter (CDOM), turbidity, and bathymetry in coastal waters [22–25]. Other hyperspectral sensors such as the Hyperspectral Infrared Imager (HyspIRI) [26] show great potential for observing water quality of coastal and inland waters. However, the application of such satellite products for Inland River monitoring has been hampered since most rivers are not able to be appropriately resolved due to the coarse resolution. According to the work of Handcock *et al.* [27], the width of the river channel must be at least three pixels for reliable water measurements from remote sensing imagery. Although commercial satellites such as World View provide significantly higher spatial resolutions (0.5–2 m), the spectral configuration of these sensors are not completely suitable for remote sensing of inland waters [28]. Hyperspectral sensors mounted on airborne platforms provide a way to collect data of sufficiently high spatial resolution that rivers can be appropriately resolved and water quality parameters can be retrieved [3,29]. However, given the high cost in organization and realization, these airborne platforms are not affordable for agencies with small budgets and therefore regular monitoring of water quality using these platforms is not realistic. Alternatively, *in situ* sampling using a handheld spectrometer provides a cost-effective, convenient, and accurate approach for measuring spectral signatures of rivers and streams. Although limited in spatial coverage, it does help fill the gap of missing remote sensing data for rivers and streams. Furthermore, the more samples taken, the more useful it will prove in making recommendations for future work on remote sensing of water quality.

The overall goal of this study was to develop a comprehensive spectral-biogeochemical database for the Wabash River and its tributaries and evaluate the ability of the radiative transfer modeling approach by using the database for the retrievals of water quality parameters including concentrations of chl, non-algal particles (NAP), and CDOM. To fulfill the goal, our specific objectives were to: (1) collect extensive field data including *in situ* concentrations of water quality parameters and nutrients, measurements of IOPs, water depths, bottom albedos, and spectral signatures of river water; (2) analyze the temporal variability of water quality parameters, nutrients, and IOPs, as well as possible factors in affecting the temporal variability; and (3) apply the LUT method to the collected dataset and evaluate its capability for retrieving water quality parameters.

## 2. Study Area

The primary study area includes the reach of the Wabash River between French Post Park (about halfway between Delphi, Indiana, and Logansport, Indiana) and Attica, Indiana, and the reach of the Tippecanoe River between Indiana State Road 18 and its confluence with the Wabash River (Figure 1). Within the study area, the Wabash River has a length of about 90,000 m and ranges in width from 100 m to over 150 m. The Tippecanoe River reach flows approximately 10 km before entering the Wabash River. Two reservoirs, Lake Freeman and Lake Shafer, are located upstream from the confluence 48,000 m and 29,000 m, respectively. The Wabash River, which has an average annual flow of approximately  $1000 \text{ m}^3 \cdot \text{s}^{-1}$ , originates from west-central Ohio and is the largest drainage in Indiana. It drains an area of over  $8.5 \times 10^{10} \text{ m}^2$  that covers two-thirds of Indiana's 92 counties and had a population of approximately 4,366,000 in 2010. In the basin, land cover is dominated by agricultural row crops (62%) with approximately 20% forest and dispersed urbanization [30]. The Wabash River flows a distance of over 650,000 m from its headwaters to its confluence with the Ohio River and is the second largest tributary of the Ohio River. It is also the longest segment of free flowing river east of the Mississippi River. The Tippecanoe River (average flow of  $145 \text{ m}^3 \cdot \text{s}^{-1}$ ) is one of over 14 major tributaries contributing flow to the main Wabash River. Lakes and swamps are the major source of the Tippecanoe River and reduce the amount of sediments carried in the river. The river enters the Wabash River 19,000 m northeast of Lafayette, Indiana and is one of the nation's most biologically diverse rivers. The drainage basin of the Tippecanoe River is in the north central part of Indiana and drains approximately  $4.92 \times 10^9 \text{ km}^2$ . The land use in the basin is predominately agriculture, which represents approximately 87% of the land area. The Wabash River and its tributaries are a vital source for water supply and recreation in Indiana. Throughout the year water depth of the Wabash River (USGS 03335500) ranges from 0.6 m to 6.0 m while the Tippecanoe River (USGS 03333050) is shallower with typical depths of 0.6 m to 2.5 m. The riverbed of the Tippecanoe River is often visible through the water during summer when flows are extremely low.

Water quality impairment occurs on various segments of the Wabash River and the Tippecanoe River. Issues include those related to *Escherichia coli* (*E. coli*), nutrients, pH, dissolved oxygen, and impaired biotic communities, according to the Indiana and Illinois 2010, 2012, and 2014 Clean Water Act (CWA) Section 303(d) listings. Major pollution sources in the watershed include nonpoint sources from agricultural and urban run-off, and point sources from treated and untreated (from combined sewer overflows) municipal wastewater. Both river play important roles in transporting pollutants downstream. According to the Ohio River Valley Water Sanitation Commission (ORSANCO, <http://www.orsanco.org/wabash-river-project>), the Wabash River is one of the largest contributors of nutrient loadings to the Mississippi River and the Gulf of Mexico. Approximately  $1.0 \times 10^7 \text{ kg}$  of total phosphorus and  $1.39 \times 10^8 \text{ kg}$  of total nitrogen are estimated to be contributed by the Wabash River watershed to the Gulf of Mexico each year [31]. Water quality conditions in the two river reaches within our study area are quite different. Based on our previous sampling experience, water in the Tippecanoe River carries significantly lower sediment loads than the Wabash River, likely due to the presence of the two upstream reservoirs, Lake Freeman and Lake Shafer, which greatly reduce sediments in the river. The water quality of the Wabash River is complex and dominated by both phytoplankton, usually measured in terms of chl concentrations, and NAP otherwise known as inorganic sediments. The water quality of the Wabash River is closely associated with flow and seasonal dynamics. During spring, sediment and nutrient loads in the Wabash River are typically the highest as a consequence of intense agricultural activities and high agricultural runoff, with lowest values occurring during summer. However, a significant increase of sediments and nutrients is often found in the rivers after summer storm events. These nutrients delivered from terrestrial environment, in return, cause algal blooms in the river and turn the water to be visibly green.



**Figure 1.** Main study area includes two reaches of the Wabash River, including the confluence with the Tippecanoe River. Field spectrometer measurements and water samples (marked as red stars) were collected through the summer of 2014. Triangles indicate United States Geological Survey (USGS) real-time streamflow monitoring stations.

### 3. Materials and Methods

#### 3.1. Field Data

Regular sampling was conducted using a boat platform on a total of 28 dates during May, June, and July 2014, resulting in a total of 213 samples from the Wabash River and the Tippecanoe River. Surface water samples of each site were collected and stored in brown polyethylene bottles until returned to the laboratory for further analysis. Above-water measurements were taken with a GER 1500 field spectrometer (Spectral Vista Corporation, <http://www.spectravista.com/>) and a Spectralon panel at each station by following NASA's standard operating protocols of satellite ocean color remote sensing [32]. Spectral range of the spectrometer is 350–1050 nm with 1.5 nm sampling interval. To avoid significant changes in illumination conditions, measurements between water target, sky, and the Spectralon panel were done within a very short time period. Sky conditions were also recorded for each station when spectral measurements occurred. For each site, water depths were measured using an ultrasonic device and recorded. An YSI sonde (YSI, <https://www.ysi.com>) was used to take instantaneous measurements of water temperature, conductivity, salinity, and dissolved oxygen.

Water measurements were accompanied also by underwater video, which was used to determine the substrate type. During low flow conditions when the streambed emerged, the albedo of various river bottom types was collected and related back to the classification of streambed materials from the video. Locations of all sampled sites were recorded using a handheld GPS device. In addition, daily discharge data from the Wabash River and the Tippecanoe River were obtained from USGS stations USGS 03335500 and USGS 03333050, respectively.

All field spectrometer measurements were processed to remove sky and sun glint by using a constant water surface reflection coefficient [33]. Therefore, remote sensing reflectance,  $R_{rs}$ , was calculated using the following equation:

$$R_{rs} = \frac{L_u - \rho L_s}{E_d} \quad (1)$$

where  $L_u$  is the total upwelling radiance,  $L_s$  is the sky radiance,  $\rho$  is the water surface reflection coefficient which is 0.028, and  $E_d$  is the measured downwelling solar irradiance.

### 3.2. Laboratory Measurements

All water samples were stored in the dark and on ice until returned to the laboratory for the determination of the concentrations of water quality parameters and nutrients (total nitrogen (TN), total phosphorous (TP), and dissolved organic carbon (DOC)), as well as spectral absorption properties of chl, NAP and CDOM.

According to the standard methods of the American Public Health Association (APHA) [34], a sub-sample was filtered onto Whatman GF/F filters, then extracted in 90% acetone solution and analyzed spectrophotometrically to determine the concentration of chl (denoted as (chl), where "(X)" indicates a concentration of X). The concentrations of TSS ((TSS)) were measured gravimetrically on pre-weighted Whatman GF/F filters after rinsing with pure water. It should be noted that TSS includes both organic and inorganic sediments, *i.e.*, chl and NAP. The organic part of TSS can be converted from (chl) using a ratio of 0.02 which is typical for mesotrophic and eutrophic systems [35]. With this ratio, the amount of organic sediments was calculated for each sample collected in summer 2014 and it was found that the organic sediments took up only a small part of the total mass (<10%). Therefore, the concentrations of NAP ((NAP)) were assumed to approximate (TSS) in this study. The concentration of DOC ((DOC)) was estimated by chemical analysis of a filtered 250 mL sample using the EPA 415.1 method. The concentrations of TP and TN ((TP) and (TN), respectively) were analyzed using an autoanalyzer after subjecting unfiltered and filtered water samples to alkaline persulfate digestion.

CDOM absorption ( $a_{\text{cdom}}(440)$ ) was measured using a laboratory spectrophotometer after filtration through 0.45  $\mu\text{m}$  membrane filters. Total absorption of particulate matter ( $a_p(\lambda)$ ) was acquired using the quantitative filter technique by measuring the particles retained on to Whatman GF/F filters spectrophotometrically. The filters were then bleached with hot methanol so that pigments were extracted. The absorption spectra of NAP ( $a_{\text{nap}}(\lambda)$ ) were determined through measurements of particles remaining on these bleached filters. The difference between  $a_p(\lambda)$  and  $a_{\text{nap}}(\lambda)$  gave an estimate of the absorption of phytoplankton ( $a_{\text{ph}}(\lambda)$ ). All spectral absorption measurements were made at 1 nm increments between 350 nm and 900 nm. The detailed lab procedure can be found in NASA protocols [32].

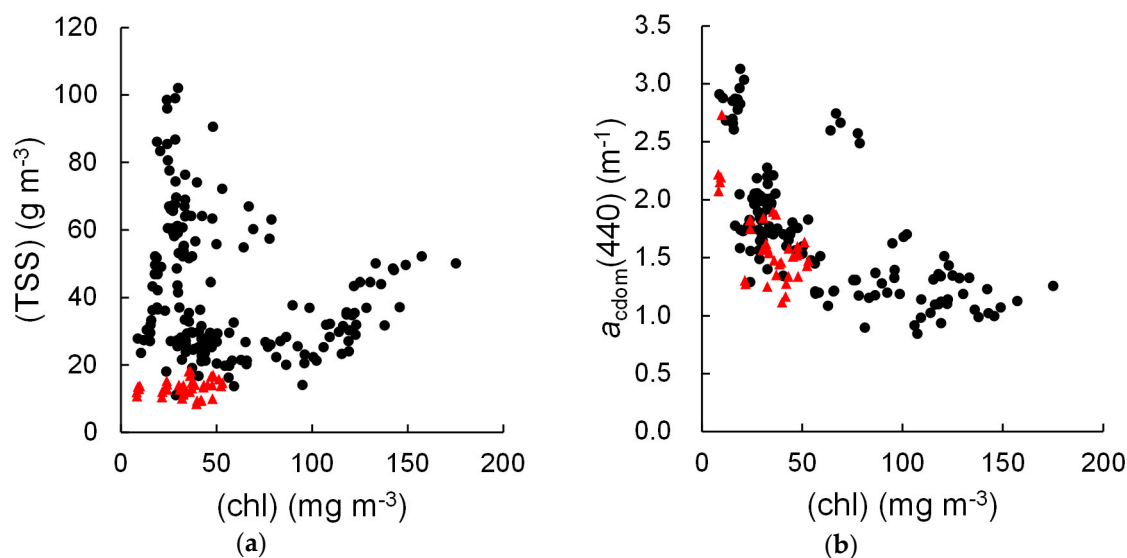
Null point corrections were performed to the lab measured absorption to remove residual offsets due to filter manufacturing and scattering artifacts caused by particle loading. For  $a_{\text{cdom}}(\lambda)$  and  $a_p(\lambda)$  correction, the average from 750 to 760 nm was forced to be null. The absorption of non-algal particles  $a_{\text{nap}}(\lambda)$  was corrected using the average absorption measured between 890 nm and 900 nm and a pathlength amplification factor of 2 [36]. Specific absorption coefficients (absorption per unit of mass concentration) of CDOM, chl and NAP were then estimated after corrections. The averaged values of the specific absorption coefficients were further fitted to exponential functions and used to represent the specific inherent optical properties (SIOPs) of the study area.

The temporal variability of water quality parameters, nutrients, and IOPs in the Wabash River and the Tippecanoe River were analyzed. Pearson's correlation ( $r$ ) analyses and significance tests were performed to explore possible factors influencing the temporal variability of these parameters. In particular, the daily distribution of (chl) sampled within the Wabash River was also evaluated using box-plots and the Mann–Whitney–Wilcoxon test was performed to determine if significant changes occurred.

### 3.3. Biophysical Characteristics

#### 3.3.1. Water Quality Observations

The Mann–Whitney–Wilcoxon test shows that the Wabash River and the Tippecanoe River experience different hydrologic regimes ( $p < 0.05$ ). Since the Wabash River and the Tippecanoe River also experience different optical properties, as shown in Section 3.3.4, they were analyzed separately. A summary of water quality observations from the Wabash River and the Tippecanoe River during the summer of 2014 is presented in Table 1. For the Wabash River, (chl) experienced large variability ranging from  $8.9 \text{ mg} \cdot \text{m}^{-3}$  to  $175.3 \text{ mg} \cdot \text{m}^{-3}$ , which spans three orders of magnitude. Concentrations of TSS also experienced large variability ranging from  $11.0 \text{ g} \cdot \text{m}^{-3}$  to  $102.0 \text{ g} \cdot \text{m}^{-3}$ . In contrast, (chl) and (TSS) in the Tippecanoe River were both lower and less variable, which is most likely due to the two upstream reservoirs serving as a settling basin. The CDOM level ( $a_{\text{cdom}}(440)$ ) in both rivers was similar and ranged from low to moderate ( $0.8 \text{ m}^{-1}$ – $3.1 \text{ m}^{-1}$ , and  $1.1 \text{ m}^{-1}$ – $2.7 \text{ m}^{-1}$ , respectively). This is consistent with (DOC) in the two rivers (Table 1). Values of (TN) were similar between the Wabash River and the Tippecanoe River, but (TP) in the Wabash River was much higher with the highest value exceeding the Indiana nutrient benchmark of  $0.3 \text{ g} \cdot \text{m}^{-3}$ . Both (TSS) and (CDOM) varied independently of phytoplankton (Figure 2). Therefore, the Wabash River and the Tippecanoe River are optically complex with non-algal particles and organics competing with phytoplankton and belong to the category of Case 2 waters [12]. The measured water depths of the Wabash River were generally higher than those of the Tippecanoe River.



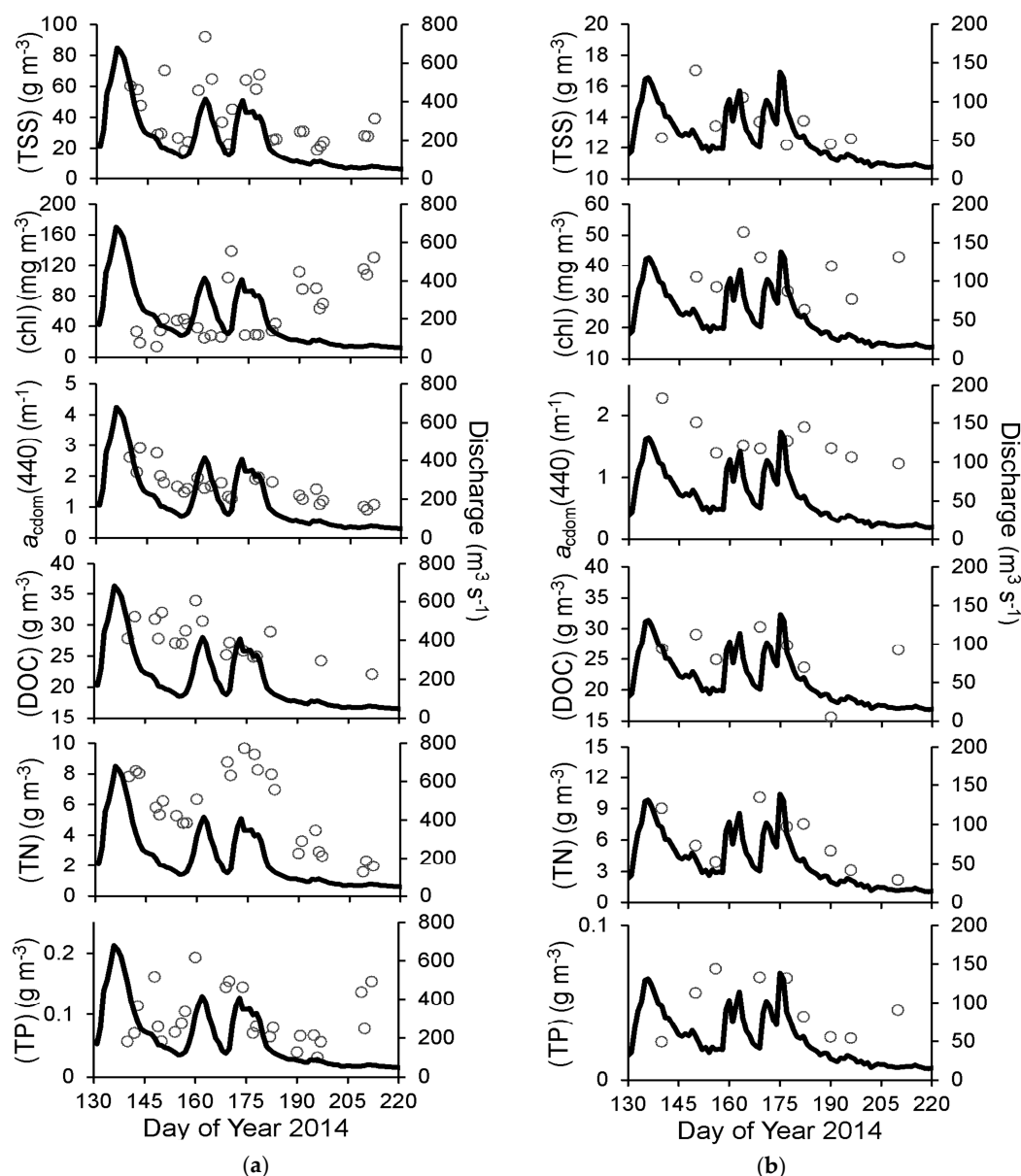
**Figure 2.** Scatterplots of measured (a) concentrations of total suspended sediments ((TSS)) and (b) concentrations of colored dissolved organic matter ( $a_{\text{cdom}}(440)$ ) versus chlorophyll concentrations ((chl)) for samples in the Wabash River (circles) and the Tippecanoe River (triangles) in summer 2014.

**Table 1.** Summary of water quality observations of the Wabash River and the Tippecanoe River in summer 2014.

|            |      | (chl)<br>mg·m <sup>-3</sup> | (TSS)<br>g·m <sup>-3</sup> | <i>a</i> <sub>cdom</sub> (440)<br>m <sup>-1</sup> | (TP)<br>g·m <sup>-3</sup> | (TN)<br>g·m <sup>-3</sup> | (DOC)<br>g·m <sup>-3</sup> | Depth<br>m |
|------------|------|-----------------------------|----------------------------|---|---------------------------|---------------------------|----------------------------|------------|
| Wabash     | Mean | 56.0                        | 42.2                       | 1.7   | 0.10                      | 5.9                       | 27.1                       | 2.0        |
|            | Min  | 8.9                         | 11.0                       | 0.8   | 0.02                      | 1.3                       | 8.9                        | 0.4        |
|            | Max  | 175.3                       | 102.0                      | 3.1   | 0.47                      | 10.0                      | 41.5                       | 4.4        |
| Tippecanoe | Mean | 34.2                        | 13.5                       | 1.6   | 0.05                      | 6.3                       | 25.8                       | 1.3        |
|            | Min  | 8.3                         | 8.5                        | 1.1   | 0.01                      | 2.0                       | 8.2                        | 0.3        |
|            | Max  | 53.1                        | 20.4                       | 2.7   | 0.12                      | 10.6                      | 35.5                       | 2.8        |

Figure 3 shows time series of the measured water quality parameters and nutrients for summer 2014. It was found that the overall pattern of daily averaged TSS concentrations in the Wabash River followed that of streamflow (Figure 3a). High concentrations of TSS were typically found in the river when streamflow increased. The bloom of phytoplankton in the Wabash River usually occurred after streamflow peaked (Figure 3a). In addition, the changes in the level of TN, TP, and carbon in the Wabash River were also associated with those of streamflow. However, such observations were not obviously displayed for the Tippecanoe River. Pearson's correlation coefficients (*r*) were further calculated between each parameter and streamflow (Table 2). Results show that the concentrations of TSS, chl, CDOM, and TN in the Wabash River were significantly correlated with streamflow. The Wabash River watershed is dominated by agricultural land use, which means large amounts of sediments and nutrients were delivered from terrestrial sources to the river during storm events. The increase of (TP) in late July (Julian Day 209, 210, and 212) when streamflow was low coincided with the combined sewer overflow (CSO) events which delivered significant amount of TP into the river [37]. In addition, the relatively low (TP) on 20 and 22 May (Julian Day 140 and 142) was caused by the relatively low amount of TP delivered into the river and the high streamflow up to 680 m<sup>3</sup>·s<sup>-1</sup> on 16 May (Julian Day 136). When these points were removed from the analysis, the correlation between (TP) and streamflow became significant (*p* = 0.025) and increased to 0.50. None of the observed water quality parameters and nutrients showed significant correlation with streamflow for the Tippecanoe River, except for the level of CDOM (Table 2). Such results indicate stream-based instead of runoff-based sources of sediments and nutrients, even though the major land use type in the Tippecanoe River watershed is also agriculture. This is mostly attributed to the two reservoirs located upstream of the Tippecanoe River study reach, which increase the residence time of water increasing the settling of sediment and uptake of nutrients.

There are two major sources of CDOM: (1) allochthonous—derived from the decomposition of woody plants in terrestrial environments; and (2) autochthonous—derived from the decomposition of algae and aquatic vegetation within the rivers. Since CDOM concentrations were significantly correlated with streamflow in both rivers (Table 2), it is highly likely that autochthonous is the dominant source in these rivers. It has to be noted that CDOM is only a portion of DOC that absorbs light. Therefore, it is not surprising that DOC showed no significant correlation with streamflow in both rivers, which means that most of DOC in the rivers is uncolored and from human sources such as wastewater discharge or from CSO that are prevalent in our study area. For many remote sensing of water quality studies (e.g., [38]), it is assumed that remote estimates of CDOM can be used to predict (DOC). However, a weak relationship was observed between CDOM and (DOC) based on our field measurements (Figure 4). This is common for water bodies affected by human activities ([39]). Therefore, the use of CDOM for estimating (DOC) in inland water bodies should be cautioned and field validation is needed unless more is known about the CDOM-DOC relationship.



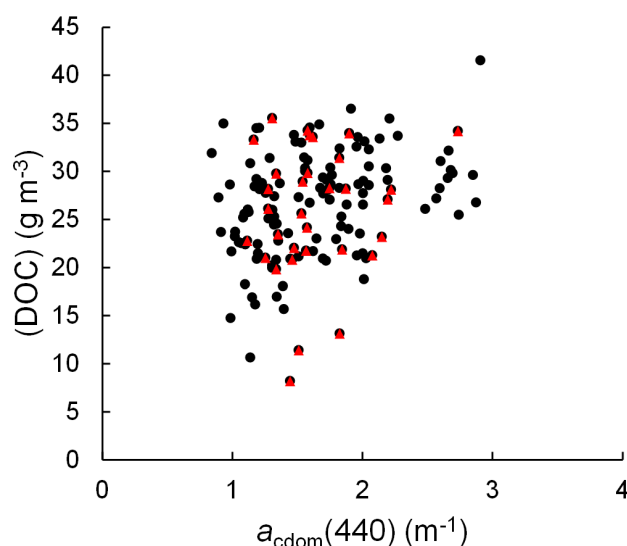
**Figure 3.** Time series of measured concentrations of water quality parameters and nutrients (circles) versus streamflow (solid line) of: (a) the Wabash River; and (b) the Tippecanoe River.

**Table 2.** Pearson's correlations between concentrations and streamflow of the Wabash River and the Tippecanoe River. The values in bold text represent those correlations that were significant ( $p < 0.05$ ).

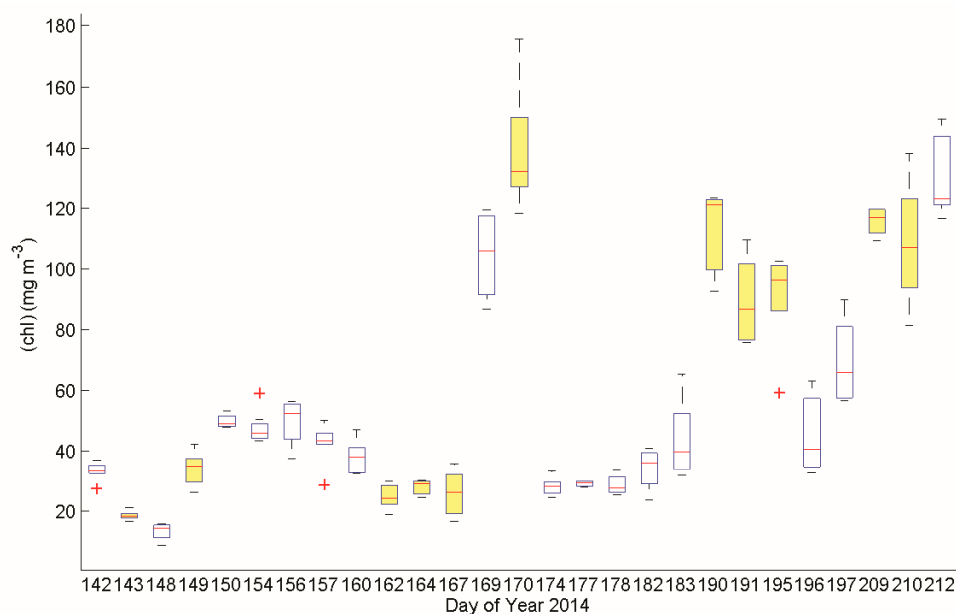
|            | (chl)        | (TSS)       | $a_{\text{cdom}}(440)$ | (TP) | (TN)        | (DOC) |
|------------|--------------|-------------|------------------------|------|-------------|-------|
| Wabash     | <b>−0.68</b> | <b>0.82</b> | <b>0.68</b>            | 0.11 | <b>0.74</b> | 0.31  |
| Tippecanoe | 0.02         | 0.36        | <b>0.72</b>            | 0.04 | 0.66        | 0.33  |

As we closely examine the measured (chl) in the Wabash River (Figure 5), it is found that variability of (chl) increased between sampled sites when (chl) increased. Significant increases of (chl) were found around early June (Day 149–160), middle June (Day 169–170), middle July (Day 190–195), and late July (Day 209–210). Specifically, the phytoplankton blooms on Day 149–160, 169–170, and 190–195 followed increases of (TN) and (TP). Following the algae blooms there were decreases of (TN) and (TP) due to biological uptake and transformations. The increase of (chl) on Day 209–210 is believed to be a result

of increased (TP) due to CSO input (Figure 3a). The magnitude of the (chl) increase was relatively lower in early June as compared to the other bloom events, which might be caused by the relatively lower amount of TP delivered into the river as well as the relatively lower water temperature (ranged approximately 21 °C–23 °C). The decrease of (chl) on Day 196–197 when (TP) and streamflow were relatively stable is believed to be a result of decreased water temperature, which dropped from 26 °C to 23 °C as shown by our field data. Therefore, it is concluded that the source of water (surface runoff or CSO) to a river, water temperature, and nutrients are important factors controlling instream concentrations of phytoplankton.



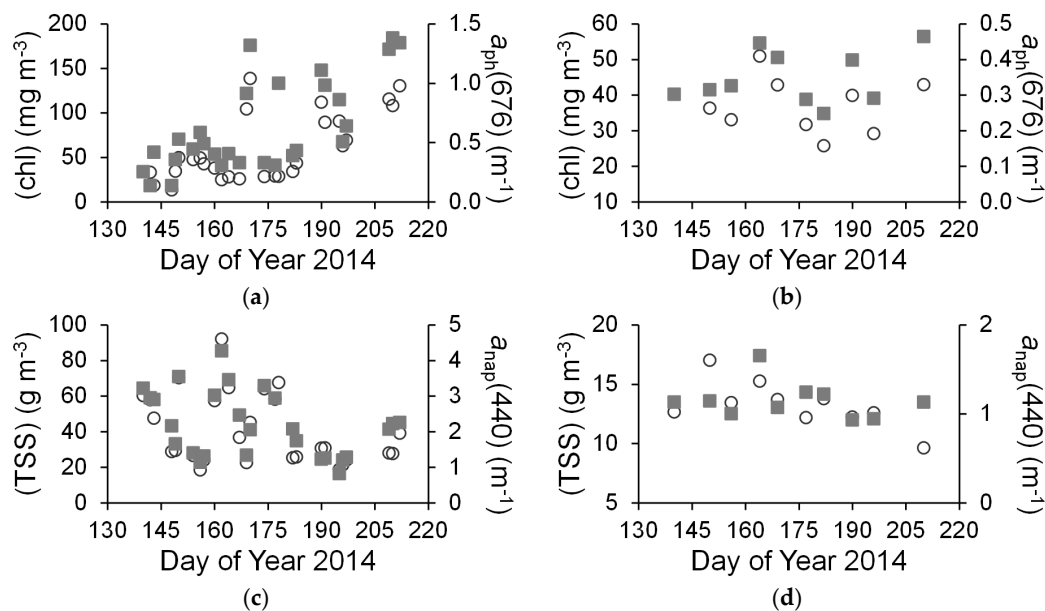
**Figure 4.** Scatterplot showing measured DOC concentrations ((DOC)) *versus* concentrations of colored dissolved organic matter ( $a_{\text{cdom}}(440)$ ) for samples collected for the Wabash River (circles) and the Tippecanoe River (triangles) in summer 2014.



**Figure 5.** Boxplots of measured chlorophyll concentrations ((chl)) for samples collected in the Wabash River during the summer of 2014. Boxes filled with yellow color indicate that there is a statistically significant different ( $p < 0.05$ ) between observations on the highlighted day and the previous day.

### 3.3.2. Inherent Optical Properties

Absorption by phytoplankton at 676 nm ( $a_{ph}(676)$ ) as estimated from the collected water samples closely paralleled changes in extracted (chl) in the Wabash River with  $r$  equal to 0.90 (Figure 6a). The temporal variability in chl absorption was dominated by algal blooms, which were caused by increased nutrients delivered from terrestrial sources during runoff events. Significant increases in chl absorption were found on days when (chl) increased. Significant correlation between  $a_{ph}(676)$  and (chl) were also found in the Tippecanoe River ( $r = 0.93$ ,  $p < 0.05$ ), although absorption by chl was lower and less variable (Figure 6b).



**Figure 6.** Scatterplots of measured inherent optical properties (filled squares) versus water quality parameters (circles) of the Wabash River (left panel) and the Tippecanoe River (right panel): (a,b)—absorption of phytoplankton at 676 nm ( $a_{ph}(676)$ ) vs. chlorophyll concentrations ((chl)); (c,d)—absorption of non-algal particles at 440 nm ( $a_{nap}(440)$ ) vs. concentrations of total suspended sediments ((TSS)).

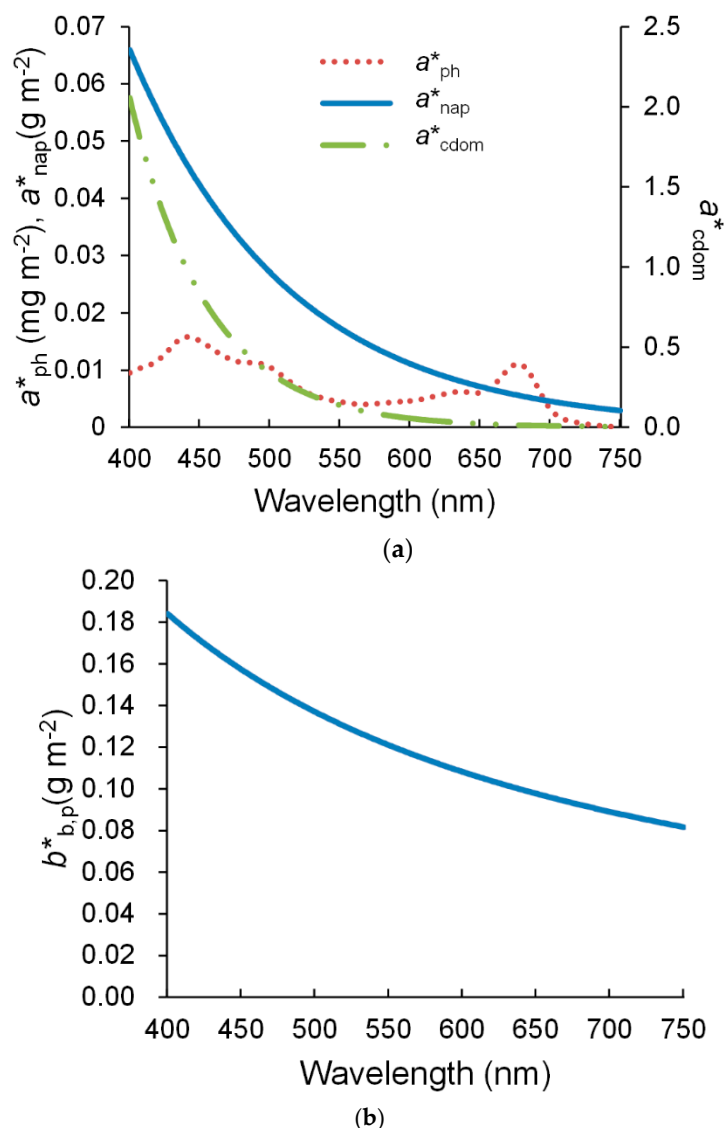
The observed changes in absorption by NAP at 440 nm ( $a_{nap}(440)$ ) corresponded strongly to the variation in (TSS) in the Wabash River (Figure 6c). The absorption coefficient of NAP at 440 nm, increased from about  $1 \text{ m}^{-1}$  to  $> 3 \text{ m}^{-1}$  during summer runoff events. In July (after Day 180) when no rainfall was observed,  $a_{nap}(440)$  was low and much less variable since the residence time of water in the river channel was longer and most of the sediment had settled to the bottom of the channel. The Tippecanoe River experiences much lower absorption caused by non-algal particles, ranging from  $0.9 \text{ m}^{-1}$  to  $1.6 \text{ m}^{-1}$  (Figure 6d), primarily because the amount of sediment in the Tippecanoe River is much lower than that in the Wabash River. No significant correlation existed between  $a_{nap}(440)$  and (TSS) in the Tippecanoe River and  $r$  only equaled to 0.35, which indicates that non-algal particles only constituted a part of TSS.

The spectral absorption of CDOM,  $a_{cdom}(\lambda)$ , can be described using an exponential function,

$$a_{cdom}(\lambda) = a_{cdom}(440) \times e^{-S_{cdom}(\lambda-440)} \quad (2)$$

with the exponential slope  $S_{cdom}$  estimated by non-linear regression. The derived values of  $S_{cdom}$  had a narrow range ( $0.0157\text{--}0.0207 \text{ nm}^{-1}$ ), which is in good agreement with those reported for inland and coastal waters [40–42]. The specific absorption of CDOM,  $a^*_{cdom}(\lambda)$ , was acquired by fitting the

ensemble mean of lab estimated specific absorption using Equation (2), with the corresponding  $S_{\text{cdom}}$  equal to  $0.018 \text{ nm}^{-1}$  (Figure 7a).



**Figure 7.** Specific inherent optical properties for the Wabash River and the Tippecanoe River: (a) absorptions; and (b) backscattering.

Similarly, an exponential function was fit to the spectral absorption of non-algal particulate matter,  $a_{\text{nap}}(\lambda)$ ,

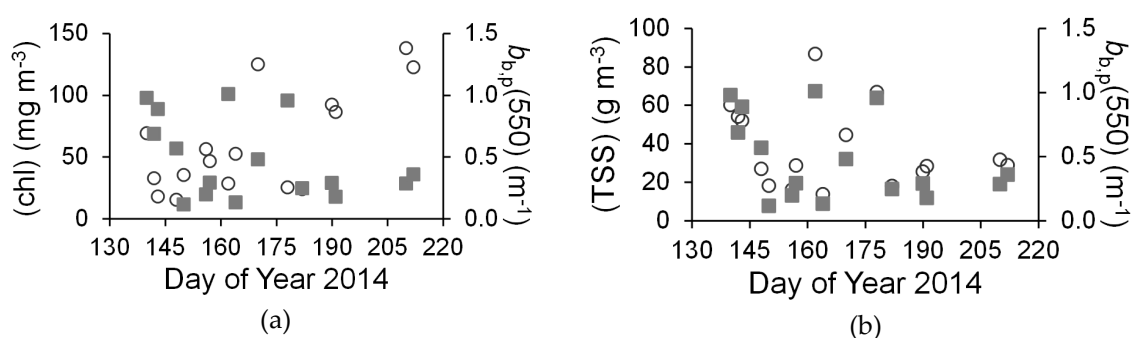
$$a_{\text{nap}}(\lambda) = (\text{TSS}) \times a_{\text{nap}}^*(440) \times e^{-S_{\text{nap}}(\lambda-440)} \quad (3)$$

where  $a_{\text{nap}}^*(440)$  is the specific absorption coefficient at 440 nm for NAP and (TSS) equals (NAP). The exponential slopes of NAP,  $S_{\text{nap}}$ , were estimated by non-linear regression and ranged from  $0.0076 \text{ nm}^{-1}$  to  $0.01 \text{ nm}^{-1}$ . These values are also similar to those reported for inland and coastal waters [40,41]. The  $a_{\text{nap}}^*(\lambda)$  was retrieved following the same method as for the  $S_{\text{nap}}$  and the corresponding measurement was  $0.089 \text{ nm}^{-1}$  (Figure 7a).

There were no systematic differences between the two rivers in the mean spectral shape of phytoplankton absorption,  $a_{\text{ph}}(\lambda)$ . Coefficients of variations of  $a_{\text{ph}}(\lambda)$  ranged from 14% to 51% for all wavelengths over the spectral range from 400 nm to 700 nm, with high variations observed at around 400–420 nm and 600–650 nm. The high variations could be ascribed to the different compositions of

chlorophyll b, chlorophyll c, and other accessory pigments in the Wabash River and the Tippecanoe River. The small bump around 480 nm and 645 nm is likely to be caused by the comparatively high concentrations of chlorophyll b and chlorophyll c for some sampled sites (Figure 7a).

The retrieved average backscattering coefficients are shown in Figure 7b, with  $b_{b,p}^*(550)$  equal to  $0.0012 \text{ m}^2 \cdot \text{g}^{-1}$  and  $\gamma$  equal to 1.3. There were no significant differences between the two rivers' backscattering properties. The retrieved  $b_{b,p}^*(550)$  values lied between  $0.006$  and  $0.02 \text{ m}^2 \cdot \text{g}^{-1}$  and the power exponent  $\gamma$  ranged from 0.5 to 2.0, typical for Case 2 waters as reported by previous literature [43]. The temporal variability of particulate backscattering at 550 nm ( $b_{b,p}(550)$ ) was closely associated with (TSS) and showed weak correlation with (chl) (Figure 8), implying the dominance of the backscattering by non-algal particles.



**Figure 8.** Variability of the retrieved backscattering coefficients of particles at 550 nm ( $b_{b,p}(550)$ , filled squares) with measured (a) concentrations of total suspended sediments ((TSS)) (circles) and (b) concentrations of chlorophyll ((chl)) (circles) in the Wabash River and the Tippecanoe River.

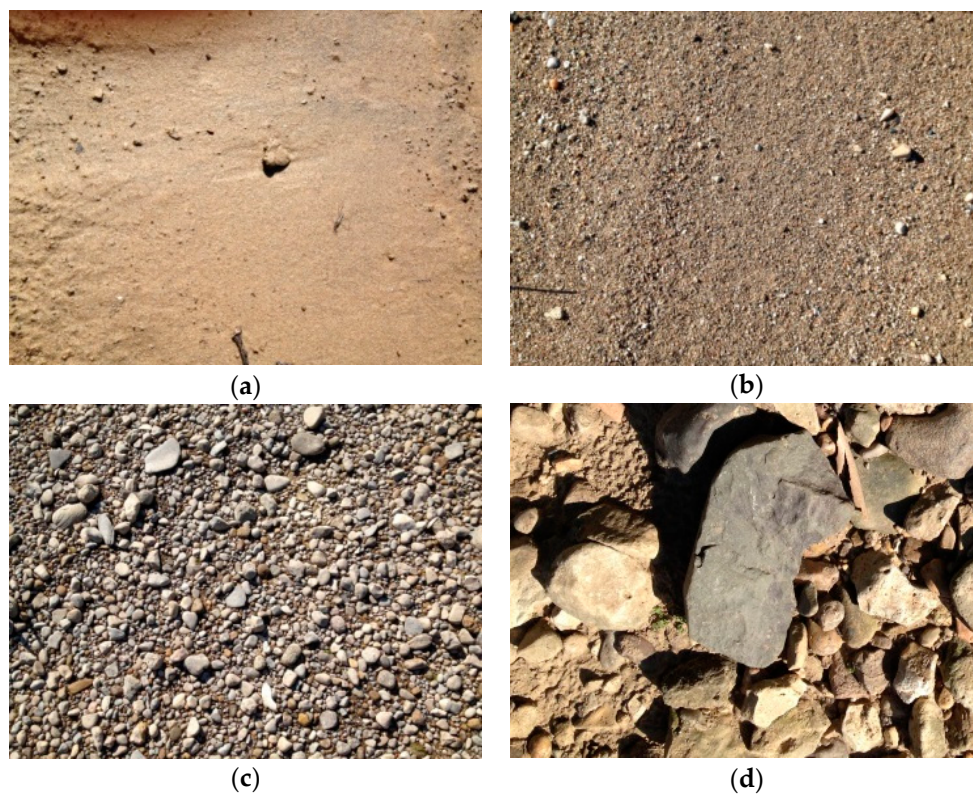
### 3.3.3. Bottom Properties

The measured bottom depths of sampled sites ranged from 0.3 m to 4.4 m, including both optically deep and shallow water (Table 1). Substrate type was categorized into six types based on the sediment size: boulder ( $>256 \text{ mm}$ ), cobble ( $255 \text{ mm}–64 \text{ mm}$ ), gravel ( $63 \text{ mm}–2 \text{ mm}$ ), sand ( $1 \text{ mm}–0.25 \text{ mm}$ ), fines ( $<0.24 \text{ mm}$ ), and hardpan (mixture of fines and clay), based on definitions from [44]. As observed, major substrate types for our study area consisted of fines, sand, gravel, and cobble with sand predominating (Figure 9). Cobbles and gravels were mostly found in the upstream portions of the Wabash River reach from French Post to Delphi and in the Tippecanoe River. The bottom of the Wabash River from Delphi to Attica was dominated by sand with fines occasionally found near the bank.

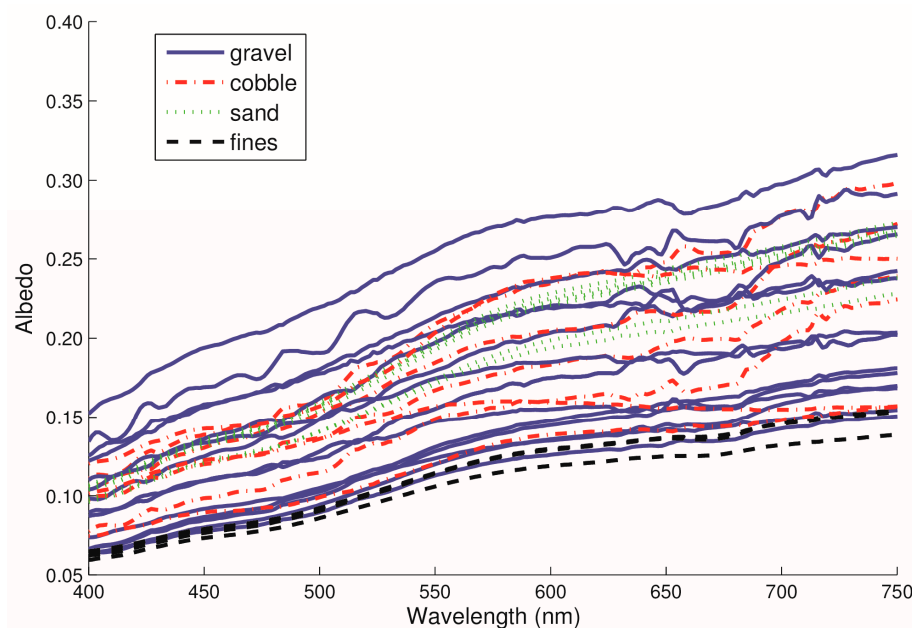
Figure 10 shows the measured albedo for different substrate types. The spectral shapes of the albedos are similar and it is hard to exactly discern each type since the ranges of measured albedos overlap with each other. Therefore, in this study, averaged albedo was used as the bottom reflectance spectrum.

### 3.3.4. Spectral Characteristics

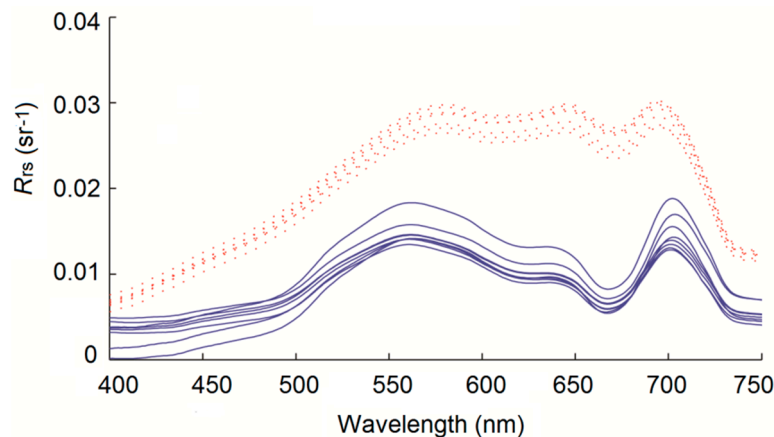
The measured spectra can be categorized into two types: (1) phytoplankton dominated and (2) sediment dominated, as shown by Figure 11. Spectra from phytoplankton dominated water experienced low reflectance in blue ( $400–500 \text{ nm}$ ) and red ( $600–700 \text{ nm}$ ) wavelengths due to the absorption by chl and other pigments. In particular, the local minimum at  $677 \text{ nm}$  and peak at  $704 \text{ nm}$  were caused by the decreasing absorption of chlorophyll and increasing absorption of water as well as the fluorescence of chl [45]. For the sediment dominated spectra, the reflectance values of these waters are relatively high in the green and red wavelengths, especially from  $560$  to  $700 \text{ nm}$ , and they lack the reflectance trough and peak in the red region caused by the absorption characteristics of chl.



**Figure 9.** Bottom types identified for the Wabash River and the Tippecanoe River: (a) fines; (b) sand; (c) gravel; and (d) cobble.



**Figure 10.** Albedo measured for different bottom types of the Wabash River and the Tippecanoe River.



**Figure 11.** Examples showing phytoplankton dominated (solid lines) and sediment dominated (dotted line) spectra.

### 3.4. Look-Up-Table Approach

The LUT methodology was used for the retrieval of water quality parameters [8]. The remote sensing reflectance,  $R_{rs}$ , can be computed exactly by solving the radiative transfer equation, as long as the environmental inputs including the water-column IOPs (the water absorption and scattering properties), the sky and water surface conditions, and water depths and bottom boundary conditions are known [46]. Therefore, the LUT methodology includes two major steps: (1) assemble a database of  $R_{rs}$  corresponding to different environmental inputs; and (2) compare the field measured  $R_{rs}$  to the spectrum in the database and find the closest match. The environmental inputs corresponding to the closest match are then considered to be the real conditions that generate the field measured  $R_{rs}$ .

No direct measurements of the backscattering coefficients of particles were available as part of this study. A subset of the samples collected in summer 2014 was selected to calibrate the backscattering properties and the remaining samples were used for model validation. Backscattering coefficients of suspend particles (including both chl and NAP) were lumped into one variable,  $b_{b,p}(\lambda)$ , which can be expressed using a power function,

$$b_{b,p}(\lambda) = (TSS) \times b_{b,p}^*(550) \times \left(\frac{550}{\lambda}\right)^\gamma \quad (4)$$

where  $b_{b,p}^*(550)$  is the specific backscattering coefficient at 550 nm,  $\lambda$  is the wavelength, and  $\gamma$  is the spectral shape parameter. The two unknowns  $b_{b,p}^*(550)$  and  $\gamma$  of the selected sites were determined by using the LUT methodology. The specific backscattering spectra were then estimated by normalizing the backscattering to the measured (TSS) and the average was used to represent the specific backscattering properties of the Wabash River and the Tippecanoe River.

Together with the lab measured absorption and bottom albedo collected in the field, the retrieved backscattering coefficients were used to construct the  $R_{rs}$  database using the HydroLight-EcoLight 5.2.2 radiative transfer model [7]. Since the Wabash River and the Tippecanoe River could be optically shallow during low flow conditions, water depth was also considered as a parameter. Therefore, to simulate  $R_{rs}$  spectrum, four main parameters are needed: (chl), (TSS), the absorption of CDOM at 440 nm,  $a_{cdom}(440)$ , and water depth. For the initial LUT, (chl) ranged from  $2 \text{ mg} \cdot \text{m}^{-3}$  to  $180 \text{ mg} \cdot \text{m}^{-3}$  at increments of  $2 \text{ mg} \cdot \text{m}^{-3}$ , (TSS) ranged from  $2 \text{ g} \cdot \text{m}^{-3}$  to  $180 \text{ g} \cdot \text{m}^{-3}$  at increments of  $2 \text{ g} \cdot \text{m}^{-3}$ , and  $a_{cdom}(440)$  ranged from 0 to  $5 \text{ m}^{-1}$  with increments of  $0.25 \text{ m}^{-1}$ . Water depths were set to start from 0.25 m with increments of 0.25 m according to [9]. During the iteration of water depths, no further simulations were executed if the  $R_{rs}$  spectrum showed no change, which means all light has been absorbed and/or scattered at this depth. This depth is referred as the maximum depth,  $D_{max}$ , and was recorded. If the retrieved depth is less than the maximum depth, it indicates optically shallow

water; otherwise, it suggests that the water depth of the specific site is equal to or greater than  $D_{\max}$ . All simulations were run using the Case 2 IOP model from 400 nm to 750 nm with 5 nm interval and assumed clear sky condition. Fluorescence of chl and CDOM were included. Only one solar zenith angle was used:  $\theta_s = 30^\circ$ . The wind speed was set to be  $5 \text{ m} \cdot \text{s}^{-1}$ . Details of the parameters used in HydroLight-EcoLight simulations are listed in Table 3.

**Table 3.** List of specifications of all the parameters for HydroLight-EcoLight simulations.

| Parameters                        | Specifications   |
|-----------------------------------|--|
| (chl)                             | $2 \text{ mg} \cdot \text{m}^{-3}$ to $180 \text{ mg} \cdot \text{m}^{-3}$ , $2 \text{ mg} \cdot \text{m}^{-3}$ increment  |
| (TSS)                             | $2 \text{ g} \cdot \text{m}^{-3}$ to $180 \text{ g} \cdot \text{m}^{-3}$ , $2 \text{ g} \cdot \text{m}^{-3}$ increment   |
| $a_{\text{cdom}}(440)$            | 0 m to 5 m, 0.25 m increment   |
| Bottom reflectance                | Value obtained in field (Section 3.3.3)  |
| Water depth                       | Start from 0.25 m, 0.25 increment  |
| Wavelength                        | 400 nm to 750 nm, 5 nm increment   |
| IOPs                              | Case 2 model<br>Absorption of pure water from Pope and Fry [47]<br>Values of $a_{\text{chl}}$ , $a_{\text{cdom}}$ , and $a_{\text{nap}}$ obtained in lab (Section 3.3.2),<br>Values of $b_{\text{b,p}}$ calibrated from collected data (Section 3.4) |
| Wind Speed                        | 5 m/s (default)  |
| Solar zenith angle                | $30^\circ$   |
| Real index of refraction of water | 1.34 (default)   |
| Sky model                         | A semi-empirical model (based on Radtran, default)<br>Sun-earth distance based on sampling dates<br>Ozone climatology based on lat/long of sampling locations  |
| Cloud cover                       | 0  |
| Fluorescence                      | Fluorescence of chlorophyll and CDOM included  |

The closest matching spectrum is defined as the one in the database showing minimum difference with field measured  $R_{\text{rs}}$  spectrum. A term, *err*, is used to describe the difference

$$err = \sum_{j=1}^J [\tilde{R}_{rs}(\lambda_j) - R_{rs}(\lambda_j)]^2 \quad (5)$$

where  $\tilde{R}_{rs}(\lambda_j)$  is the database spectrum at band  $j$ ,  $R_{rs}(\lambda_j)$  is the field measured spectrum at band  $j$ , and  $J$  is the total number of band.

To better quantitatively describe the retrieval results for NAP and chl, two metrics were adopted: (1) concentration difference; and (2) percent difference. The concentration difference is defined as

$$\text{concentration difference} = \frac{1}{N} \sum_{i=1}^N [C(\text{LUT}; i) - C(\text{field}; i)] \quad (6)$$

and percent difference between the LUT retrieved and field measured concentrations is computed as

$$\text{percent difference} = \frac{100}{N} \sum_{i=1}^N \frac{C(\text{LUT}; i) - C(\text{field}; i)}{C(\text{field}; i)} \quad (7)$$

In Equations (6) and (7),  $C(\text{LUT}; i)$  is the LUT retrieved concentration for site  $i$ ,  $C(\text{field}; i)$  is the field measured concentration at site  $i$ , and  $N$  is the total number of sampled sites.

Although theoretically a given  $R_{rs}$  spectrum corresponds to a particular set of environmental/water quality conditions, incorrect information may be retrieved when inverting  $R_{rs}$  due to errors in the field measurements. Therefore, to investigate the non-uniqueness problem of the LUT retrieval, a subset of 100  $R_{rs}$  spectra in the database based on *a priori* information of the study area was randomly selected. Random errors of  $\pm 1\%$ ,  $\pm 2.5\%$ ,  $\pm 4\%$ , and  $\pm 5\%$  were then added to the selected  $R_{rs}$  spectra and the retrieved results of the subset were evaluated.

### 3.5. Database Distribution

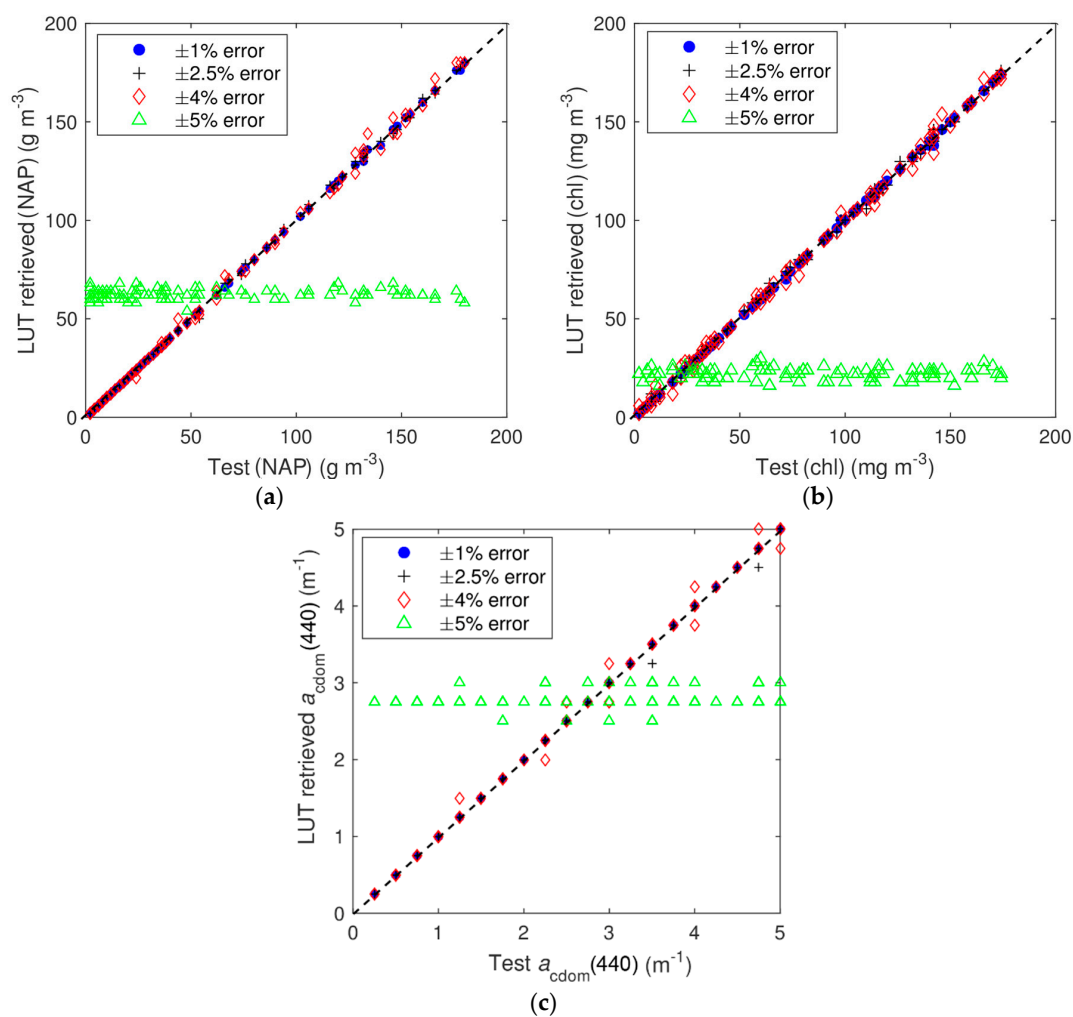
All the data collected in this study, including *in situ* water quality, nutrient level, IOPs, and spectral measurements, as well as all additional associated data (e.g., bottom albedos, water depth, date, time of day) and the LUT, were integrated into a database using Microsoft Access and distributed online through the Purdue University Research Repository (PURR—<http://purrr.purdue.edu/>). This dataset will be assigned a Document Object Identifier (DOI) from the Purdue Library and published. Purdue University will maintain the dataset for at least 10 years after the completion of this project. This published database will provide useful ground truth data for remote sensing of water quality in inland waters and valuable sources for further investigation of the relationship between optical and biogeochemical properties.

## 4. Results

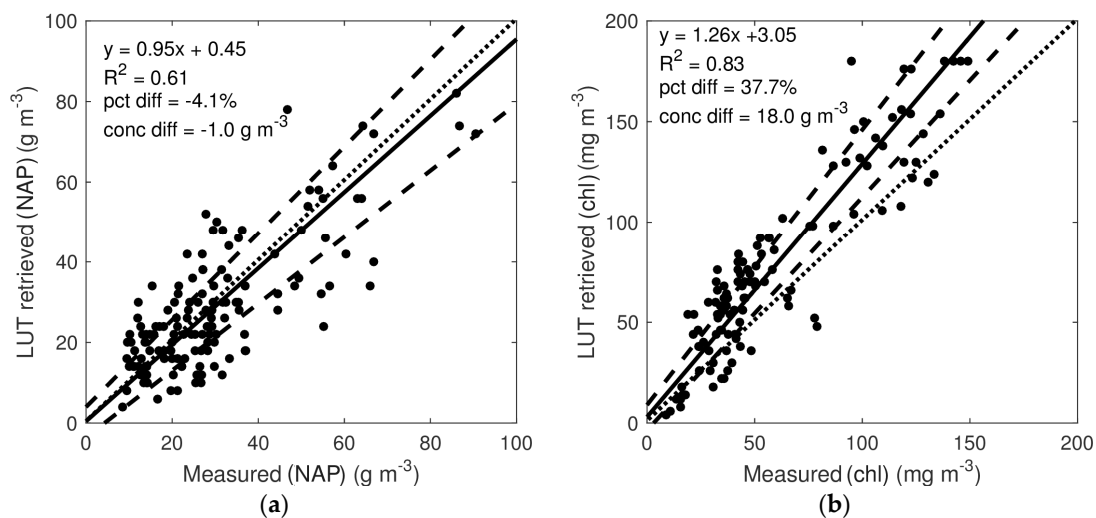
A total of 550,054 spectra were generated using HydroLight-EcoLight. Before we applied the entire database of  $R_{rs}$  spectra to the analysis of the field collected spectrometer data for the Wabash River and the Tippecanoe River, we first resample the field measured  $R_{rs}$  spectra (1.5 nm) with a cubic spline fit to correspond to the LUT wavelengths (5 nm). The database of  $R_{rs}$  spectra was created using a specialized version of HydroLight (*i.e.*, EcoLight) and it took about six days to complete all of the simulations. Although it is time consuming to build the  $R_{rs}$  database, it is one-time effort and water quality parameters can be retrieved much more quickly by searching through the database than by completing the EcoLight simulations. This searching was implemented in C++ and it took approximately 15 s to find the closest match for each field measured spectrum.

Figure 12 shows the results of using a subset of  $R_{rs}$  spectra in the database for testing the ability of the LUT method for water quality retrieval. The water quality conditions corresponding to the selected  $R_{rs}$  spectra are referred to as test measurements here. As seen from the figure, when the error level of  $R_{rs}$  is within  $\pm 4\%$ , the LUT retrieved (chl), (NAP), and  $a_{cdom}(\lambda)$  are all in good agreement with the values of the test measurements. However, when the error level increased to  $\pm 5\%$ , the estimated values were completely different from what was expected, in this case optically deep turbid water was retrieved as optically shallow clear water. Therefore, accurate field measurements of  $R_{rs}$  are very important for the success of the LUT method. To avoid such non-uniqueness problems in this study, we constrained the inversion by restricting water depths for sites that were optically shallow. This was implemented by searching a subset of the database limited to locations where the water depths were within 0.25 m of the field measured values. The results of constrained inversions are similar to those of the unconstrained inversions, which are discussed below.

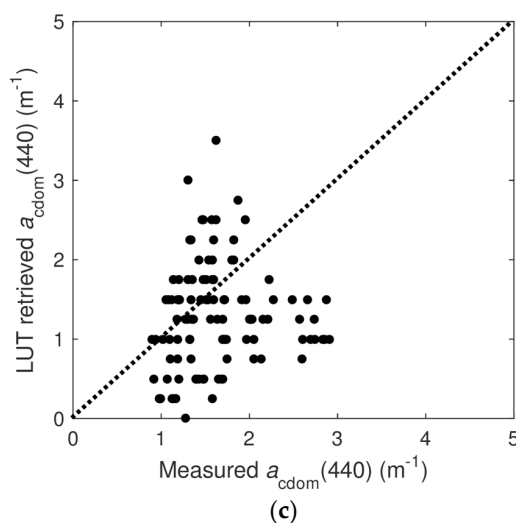
Values for (NAP), (chl), and  $a_{cdom}(440)$  were simultaneously retrieved by finding the closest matching LUT spectra to the field measurements (Figure 13). It is clear that the resulting points fall close to the 1:1 line for (NAP) estimates. The average percent difference calculated for (NAP) estimation is 4.1% and the concentration is  $-1.0 \text{ g} \cdot \text{m}^{-3}$ . It is thus concluded that the LUT retrieved (NAP) values are in close agreement with coincident *in situ* measurements. The LUT estimates of (chl) tend to be higher than field measured values with the regression slope of 1.26. As compared to the (NAP) estimates, the average error for (chl) estimation is larger, which is 37.7% or  $18.0 \text{ mg} \cdot \text{m}^{-3}$ . No statistically significant relationship was found between modeled and measured  $a_{cdom}(440)$  (Figure 13c). Although the points for CDOM comparison visibly cluster near the 1:1 line, the high variability restricts us from making any conclusion.



**Figure 12.** Comparison between test and look-up table estimated concentrations of: (a) non-algal particles ((NAP)); (b) chlorophyll ((chl)); and (c) colored organic matter ( $a_{\text{cdom}}(440)$ ), for the Wabash River and the Tippecanoe River. The dotted line represents 1:1 line.



**Figure 13.** Cont.

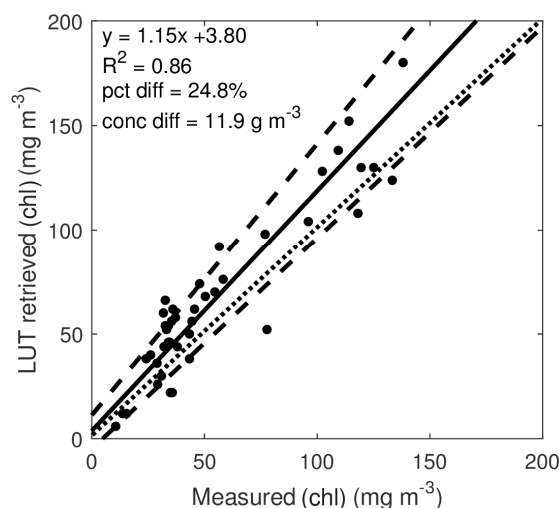


**Figure 13.** Comparison between measured and look-up table estimated (unconstrained inversion) concentrations of: (a) non-algal particles ((NAP)); (b) chlorophyll ((chl)); and (c) colored organic matter ( $a_{\text{cdom}}(440)$ ) for the Wabash River and the Tippecanoe River. The dotted line represents 1:1 line and the dashed lines represent 95% confidence interval.

## 5. Discussion

Our lab results show that the measured absorption coefficients of chl ( $a_{\text{ph}}^*(\lambda)$ ) exhibited high variations at around 400–420 nm and 600–650 nm. This is consistent with numerous laboratory and field studies of  $a_{\text{ph}}^*(\lambda)$  in case 2 water over the last two decades [48–50]. Such variability can be attributed to pigment composition [51] and packaging [52]. Given the dynamic physical and chemical conditions of inland rivers, the structure of phytoplankton community is highly variable. It is thus most likely that the robust relationship between (chl) and  $a_{\text{ph}}^*(\lambda)$  found in open oceans may not work well in inland river systems due to the contribution of accessory pigments to absorption and pigment packaging [50]. To confirm this, we reanalyzed our samples by selecting those where the measured  $a_{\text{ph}}^*(\lambda)$  experienced fewer features related to other pigments (Figure 14). By limiting the analysis to more uniform samples, we found that the regression slope became 1.15 and the average error dropped to 24.8%. Therefore, in order to improve the accuracy of (chl) retrieval, future work should include studying the pigment composition and phytoplankton cell size for better quantification of the relationship between (chl) and  $a_{\text{ph}}^*(\lambda)$  in our study area.

Based on our observations the change in  $a_{\text{cdom}}(440)$  was within  $3 \text{ m}^{-1}$  (Table 1), while the value can be as high as  $40 \text{ m}^{-1}$  for inland water [38]. In [14] we also found that  $R_{\text{rs}}$  is not sensitive to the observed changes in  $a_{\text{cdom}}(440)$ , therefore it is highly likely that the small observed changes cannot be adequately captured by the LUT methodology. It is also possible that the LUT retrievals for  $a_{\text{cdom}}(440)$  display large uncertainties at low CDOM levels [39], but produce an overall good 1:1 fit with a wider data range than the CDOM levels in the Wabash River and the Tippecanoe River sampled in summer 2014. This needs further investigation with more data collected for the rivers during other seasons of the year, for example, in spring when agricultural activities are intense and streamflow is high.



**Figure 14.** Comparison between measured and look-up table estimated (unconstrained inversion) values of chlorophyll concentrations ((chl)) for selected samples showing fewer features of accessory pigments. The dotted line represents 1:1 line and the dashed lines represent 95% confidence interval.

## 6. Conclusions

In this study, a comprehensive spectral-biogeochemical database of the Wabash River and the Tippecanoe River, Indiana, was developed. This database mainly includes remote sensing reflectance spectra of river water taken using a hand-held spectrometer, IOPs, concentrations of water quality parameters (chl, NAP, and CDOM) and nutrients (TP, TN, DOC), water depths, substrate types, and bottom reflectance spectra collected in summer 2014. Our results show that the temporal variability of water quality parameters and nutrients of the Wabash River in summer 2014 were significantly associated with hydrologic regime. Summer runoff events and CSOs that were prevalent in our study area played an important role in delivering nutrients and sediments to the Wabash River. In contrast, none of the water quality parameters and nutrients showed significant correlation with streamflow for the Tippecanoe River except for CDOM, due to the two upstream reservoirs which increase the residence time of water. It is highly likely that most of DOC in the rivers is uncolored and from human sources such as wastewater discharged from CSOs. Nutrients inputs, water temperature and the intensity of major runoff events are important factors controlling instream concentrations of phytoplankton. The LUT methodology was further applied to the dataset for inversion of field measured  $R_{rs}$  spectra. Significant linear relationships existed between the LUT retrieved and field measured values of (NAP) with a slope close to 1.0. The average percent difference of (NAP) estimates was 4.1% and the concentration difference was  $-1.0 \text{ g} \cdot \text{m}^{-3}$ . The average error between the LUT retrieved and field measured (chl) values was larger (37.7% or  $18.0 \text{ mg} \cdot \text{m}^{-3}$ ). However, after reselecting samples that were less likely to be influenced by other pigments, the average error decreased to 24.8% and the regression slope was close to 1.0. It is concluded that the specific absorption spectrum of chl was not well characterized, which affects the accuracy of (chl) retrieval. No significant relationship was found between the LUT retrieved and field measured CDOM values. The large variability of the LUT retrieved CDOM values could be due to the fact of small data ranges and the insensitivity of  $R_{rs}$  to the change in CDOM.

The initial evaluation of the database gives us reason to believe that the LUT method will prove to be a general and robust way of retrieving water quality parameters in our study area. The success of the LUT method depends on the accurate and appropriate measurements of IOPs and  $R_{rs}$ . Therefore, further improvements in the retrievals can be implemented by continuing data collection of the rivers in other seasons of the year, improving the characterization of water IOPs, and adding additional reflectance spectra and water IOPs as well as other environmental information (e.g., wind speed,

cloud cover *etc.*) to the existing database. Although the LUT method is best suited for specific local environments with enough *a priori* environmental information, as was done here, there are still limitations. For example, it is not universally applicable because it would be impractical to develop a database that is big enough to cover any range of water quality conditions. In addition, the retrieval of water quality parameters in such a database would be too computationally intensive. As we noticed, the inherent granularity (*i.e.*, discrete increments), in the database affects the estimates of errors in the retrievals as well as the time for database building and searching. Therefore, in order to extend the LUT method to a larger area, for example, the entire Wabash River in Indiana, it is necessary to find the optimal inherent granularity that produces water quality estimates within acceptable error limits while maintaining reasonable computing times. To handle shallow water conditions, it would be beneficial to determine if bottom contributions exist at given depths before building the database so that the computing time can be further reduced, as opposed to simply constraining the inversion.

**Acknowledgments:** This work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program—Grant “NNX14AL31H”. We thank Purdue Water Community (PWC) and Indiana Water Resource Center (IWRRC) for providing the financial support for the purchase of the Hydrolight software license. We greatly appreciate Reuben Goforth and his research group in the Department of Forestry and Natural Resources at Purdue University for scheduling our boat sampling trips into their fieldwork timeline. We are also thankful to Amanda Montgomery at the Ecohydrology Laboratory, Purdue University and Megan Heller at the Soil Microbiology Laboratory, Purdue University for their help in nutrient analyses.

**Author Contributions:** Jing Tan and Keith Cherkauer conceived and designed the fieldwork; Jing Tan performed the field work and experiments and analyzed the data; Indrajeet Chaubey contributed reagents/materials/analysis tools; and Jing Tan wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                  |
|------|----------------------------------|
| Chl  | chlorophyll                      |
| NAP  | non-algal particles              |
| TSS  | total suspended solids           |
| CDOM | colored dissolved organic matter |
| DOC  | dissolved organic carbon         |
| TN   | total nitrogen                   |
| TP   | total phosphorus                 |
| LUT  | look-up table                    |
| CSO  | combined sewer overflow          |

## References

1. Ritchie, J.C.; Schiebe, F.R. Water quality. In *Remote Sensing in Hydrology and Water Management*; Schultz, G.A., Engman, E.T., Eds.; Springer-Verlag: Berlin, Germany, 2000; pp. 287–303.
2. Ritchie, J.C.; Zimba, P.V.; Everitt, J.H. Remote sensing techniques to assess water quality. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 695–704. [[CrossRef](#)]
3. Olmanson, L.G.; Brezonik, P.L.; Bauer, M.E. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi river and its tributaries in Minnesota. *Remote Sens. Environ.* **2013**, *130*, 254–265. [[CrossRef](#)]
4. Tan, J.; Cherkauer, K.A.; Chaubey, I. Using hyperspectral data to quantify water-quality parameters in the Wabash river and its tributaries, Indiana. *Int. J. Remote Sens.* **2015**, *36*, 5466–5484. [[CrossRef](#)]
5. Lee, Z.P.; Carder, K.L.; Mobley, C.D.; Steward, R.G.; Patch, J.S. Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization. *Appl. Opt.* **1999**, *38*, 3831–3843. [[CrossRef](#)] [[PubMed](#)]
6. Lee, Z.; Carder, K.L.; Chen, R.F.; Peacock, T.G. Properties of the water column and bottom derived from airborne visible infrared imaging spectrometer (AVIRIS) data. *J. Geophys. Res. Oceans* **2001**, *106*, 11639–11651. [[CrossRef](#)]
7. Mobley, C.D.; Sundman, L.K. *Hydrolight 5.2 Users' Guide*; Sequoia Scientific: Bellevue, WA, USA, 2013.

8. Louchard, E.M.; Reid, R.P.; Stephens, F.C.; Davis, C.O.; Leathers, R.A.; Downes, T.V. Optical remote sensing of benthic habitats and bathymetry in coastal environments at Lee Stocking island, Bahamas: A comparative spectral classification approach. *Limnol. Oceanogr.* **2003**, *48*, 511–521. [[CrossRef](#)]
9. Mobley, C.D.; Sundman, L.K.; Davis, C.O.; Bowles, J.H.; Downes, T.V.; Leathers, R.A.; Montes, M.J.; Bissett, W.P.; Kohler, D.D.R.; Reid, R.P.; *et al.* Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. *Appl. Opt.* **2005**, *44*, 3576–3592. [[CrossRef](#)] [[PubMed](#)]
10. Lesser, M.P.; Mobley, C.D. Bathymetry, water optical properties, and benthic classification of coral reefs using hyperspectral remote sensing imagery. *Coral Reefs* **2007**, *26*, 819–829. [[CrossRef](#)]
11. Lee, Z.P.; Arnone, R.; Hu, C.M.; Werdell, P.J.; Lubac, B. Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Appl. Opt.* **2010**, *49*, 369–381. [[CrossRef](#)] [[PubMed](#)]
12. IOCCG. *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex Waters*; Reports of the International Ocean-Colour Coordinating Group No. 3; Sathyendranath, S., Ed.; IOCCG: Dartmouth, NS, Canada, 2000. Available online: [http://www.ioccg.org/reports\\_ioccg.html](http://www.ioccg.org/reports_ioccg.html) (accessed on 3 February 2016).
13. Maritorena, S.; Morel, A.; Gentili, B. Diffuse-reflectance of oceanic shallow waters-influence of water depth and bottom albedo. *Limnol. Oceanogr.* **1994**, *39*, 1689–1703. [[CrossRef](#)]
14. Lee, Z.P.; Carder, K.L.; Mobley, C.D.; Steward, R.G.; Patch, J.S. Hyperspectral remote sensing for shallow waters. I. A semianalytical model. *Appl. Opt.* **1998**, *37*, 6329–6338. [[CrossRef](#)] [[PubMed](#)]
15. Tan, J.; Cherkauer, K.A.; Chaubey, I.; Troy, C.D.; Essig, R. Water quality estimation of river plumes in southern Lake Michigan using Hyperion. *J. Great Lakes Res.* **2016**, *42*, 524–535. [[CrossRef](#)]
16. Palmer, S.C.J.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2015**, *157*, 1–8. [[CrossRef](#)]
17. Mouw, C.B.; Greb, S.; Aurin, D.; DiGiacomo, P.M.; Lee, Z.; Twardowski, M.; Binding, C.; Hu, C.M.; Ma, R.H.; Moore, T.; *et al.* Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **2015**, *160*, 15–30. [[CrossRef](#)]
18. Boss, E.S.; Collier, R.; Larson, G.; Fennel, K.; Pegau, W.S. Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR. *Hydrobiologia* **2007**, *574*, 149–159. [[CrossRef](#)]
19. Lee, Z.P.; Carder, K.L. Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data. *Appl. Opt.* **2002**, *41*, 2191–2201. [[CrossRef](#)] [[PubMed](#)]
20. Brando, V.E.; Dekker, A.G. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Trans. Geosci. Remote* **2003**, *41*, 1378–1387. [[CrossRef](#)]
21. Lucke, R.L.; Corson, M.; McGlothlin, N.R.; Butcher, S.D.; Wood, D.L.; Korwan, D.R.; Li, R.R.; Snyder, W.A.; Davis, C.O.; Chen, D.T. Hyperspectral imager for the coastal ocean: Instrument description and first images. *Appl. Opt.* **2011**, *50*, 1501–1516. [[CrossRef](#)] [[PubMed](#)]
22. Gitelson, A.A.; Gao, B.C.; Li, R.R.; Berdnikov, S.; Saprygin, V. Estimation of chlorophyll-a concentration in productive turbid waters using a hyperspectral imager for the coastal ocean-the Azov sea case study. *Environ. Res. Lett.* **2011**, *6*. [[CrossRef](#)]
23. Ryan, J.P.; Davis, C.O.; Tuffiaro, N.B.; Kudela, R.M.; Gao, B.C. Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in Monterey bay, CA, USA. *Remote Sens.* **2014**, *6*, 1007–1025. [[CrossRef](#)]
24. Garcia, R.A.; Fearn, P.R.C.S.; McKinna, L.I.W. Detecting trend and seasonal changes in bathymetry derived from HICO imagery: A case study of shark bay, Western Australia. *Remote Sens. Environ.* **2014**, *147*, 186–205. [[CrossRef](#)]
25. Keith, D.J.; Schaeffer, B.A.; Lunetta, R.S.; Gould, R.W.; Rocha, K.; Cobb, D.J. Remote sensing of selected water-quality indicators with the hyperspectral imager for the coastal ocean (HICO) sensor. *Int. J. Remote Sens.* **2014**, *35*, 2927–2962. [[CrossRef](#)]
26. Devred, E.; Turpie, K.R.; Moses, W.; Klemas, V.V.; Moisan, T.; Babin, M.; Toro-Farmer, G.; Forget, M.H.; Jo, Y.H. Future retrievals of water column bio-optical properties using the hyperspectral infrared imager (HyspIRI). *Remote Sens. (Basel)* **2013**, *5*, 6812–6837. [[CrossRef](#)]
27. Handcock, R.N.; Gillespie, A.R.; Cherkauer, K.A.; Kay, J.E.; Burges, S.J.; Kampf, S.K. Accuracy and uncertainty of thermal-infrared remote sensing of stream temperatures at multiple spatial scales. *Remote Sens. Environ.* **2006**, *100*, 427–440. [[CrossRef](#)]

28. Lee, Z.; Weidemann, A.; Arnone, R. Combined Effect of reduced band number and increased bandwidth on shallow water remote sensing: The case of worldview 2. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2577–2586. [CrossRef]
29. Shafique, N.A.; Fulk, F.; Autrey, B.C.; Flotemersch, J. Hyperspectral remote sensing of water quality parameters for large rivers in the Ohio River basin. In Proceedings of the First Interagency Conference on Research in the Watersheds, USDA Agricultural Research Service, Washington, DC, USA, 27–30 October 2003.
30. United States Army Corps of Engineers. Wabash River Watershed Section 729 Initial Watershed Assessment. Available online: <http://www.lrl.usace.army.mil/Portals/64/docs/CWProjects/WabashStudy.pdf> (accessed on 3 February 2016).
31. United States Environmental Protection Agency. Wabash River Watershed Water Quality Trading Feasibility Study Final Report. Available online: [http://www.ctic.org/media/pdf/Wabash%20WQT%20Feasibility%20Study\\_091411\\_final%20report.pdf](http://www.ctic.org/media/pdf/Wabash%20WQT%20Feasibility%20Study_091411_final%20report.pdf) (accessed on 3 February 2016).
32. Mueller, J.L.; Fargion, G.S.; Macclain, C.R. *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4*; NASA Technical Memo 211621. Goddard Space Flight Space Center: Greenbelt, MD, USA, 2003. Available online: <http://oceancolor.gsfc.nasa.gov/DOCS/> (accessed on 3 February 2016).
33. Mobley, C.D. Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Opt.* **1999**, *38*, 7442–7455. [CrossRef] [PubMed]
34. Eaton, A.D.; Clesceri, L.S.; Greenberg, A.E.; Franson, M.A.H. *Standard Methods for the Examination of Water and Wastewater*, 20th ed.; American Public Health Association: Washington, DC, USA, 1998; Volume 1.
35. Likens, G.E. *Plankton of Inland Waters*; Academic Press: Oxford, UK, 2010.
36. Roesler, C.S. Theoretical and experimental approaches to improve the accuracy of particulate absorption coefficients derived from the quantitative filter technique. *Limnol. Oceanogr.* **1998**, *43*, 1649–1660. [CrossRef]
37. Moore, M.; Lafayette Water Pollution Control Department, Lafayette, IN, USA. Personal Communication, 2015.
38. Kutser, T.; Pierson, D.C.; Kallio, K.Y.; Reinart, A.; Sobek, S. Mapping lake CDOM by satellite remote sensing. *Remote Sens. Environ.* **2005**, *94*, 535–540. [CrossRef]
39. Brezonik, P.L.; Olmanson, L.G.; Finlay, J.C.; Bauer, M.E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sens. Environ.* **2015**, *157*, 199–215. [CrossRef]
40. Roesler, C.S.; Perry, M.J.; Carder, K.L. Modeling *in situ* phytoplankton absorption from total absorption-spectra in productive inland marine waters. *Limnol. Oceanogr.* **1989**, *34*, 1510–1523. [CrossRef]
41. Babin, M.; Morel, A.; Fournier-Sicre, V.; Fell, F.; Stramski, D. Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. *Limnol. Oceanogr.* **2003**, *48*, 843–859. [CrossRef]
42. Binding, C.E.; Jerome, J.H.; Bukata, R.P.; Booty, W.G. Spectral absorption properties of dissolved and particulate matter in Lake Erie. *Remote Sens. Environ.* **2008**, *112*, 1702–1711. [CrossRef]
43. IOCCG. *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*; Reports of the International Ocean-Colour Coordinating Group No. 5; Lee, Z.-P., Ed.; IOCCG: Dartmouth, NS, Canada, 2006. Available online: [http://www.ioccg.org/reports\\_ioccg.html](http://www.ioccg.org/reports_ioccg.html) (accessed on 3 February 2016).
44. Wentworth, C.K. A scale of grade and class terms for clastic sediments. *J. Geol.* **1922**, *30*, 377–392. [CrossRef]
45. Vasilkov, A.P.; Kopelevich, O.V. On the reasons of the appearance of the maximum near 700 nm in the upwelling radiance spectrum in the sea. *Okeanologiya* **1982**, *22*, 945–950.
46. Mobley, C.D. *Light and Water: Radiative Transfer in Natural Waters*; Academic Press: London, UK, 1994.
47. Pope, R.M.; Fry, E.S. Absorption spectrum (380–770 nm) of pure water. II. Integrating cavity measurements. *Appl. Opt.* **1997**, *36*, 8710–8723. [CrossRef] [PubMed]
48. Babin, M.; Theriault, J.-C.; Legendre, L.; Condal, A. Variations in the specific absorption coefficient for natural phytoplankton assemblages: Impact on estimates of primary production. *Limnol. Oceanogr.* **1993**, *38*, 154–177. [CrossRef]
49. Babin, M.; Stramski, D.; Ferrari, G.M.; Claustre, H.; Bricaud, A.; Obolensky, G.; Hoepffner, N. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res.* **2003**, *108*. [CrossRef]
50. Perkins, M.; Effler, S.W.; Strait, C.M. Phytoplankton absorption and the chlorophyll a-specific absorption coefficient in dynamic Onondaga Lake. *Inland Waters* **2014**, *4*, 133–146. [CrossRef]

51. Hoepffner, N.; Sathyendranath, S. Effect of pigment composition on absorption properties of phytoplankton. *Mar. Ecol. Prog. Ser.* **1991**, *73*, 11–23. [[CrossRef](#)]
52. Bricaud, A.; Morel, A.; Babin, M.; Allali, K.; Claustre, H. Variations of light absorption by suspended particles with the chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *J. Geophys. Res.* **1998**, *103*, 31033–31044. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).