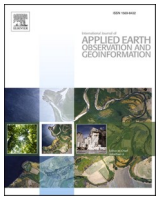
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103053

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Application of airborne hyperspectral imagery to retrieve spatiotemporal CDOM distribution using machine learning in a reservoir

Jinuk Kim [a], Wonjin Jang [a], Jin Hwi Kim [a], Jiwan Lee [a], Kyung Hwa Cho [b,c], Yong-Gu Lee [d], Kangmin Chon [d,e], Sanghyun Park [f], JongCheol Pyo [g], Yongeun Park [h,*], Seongjoon Kim [h]

[a] Department of Civil, Environmental and Plant Engineering, Graduate School, Konkuk University, Seoul 05029, Republic of Korea
[b] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea
[c] Graduate School of Carbon Neutrality, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea
[d] Department of Environmental Engineering, Kangwon National University, Gangwon-do 24341, Republic of Korea
[e] Department of Integrated Energy and Infra System, Kangwon National University, Gangwon-do t24341, Republic of Korea
[f] Water Quality Assessment Research Division, National Institute of Environmental Research, Environmental Research Complex, Incheon 22689, Republic of Korea
[g] Department of Environmental Engineering, Pusan National University, Busan 46241, Republic of Korea
[h] Department of Civil and Environmental Engineering, Konkuk University, Seoul 05029, Republic of Korea

ARTICLE INFO

Keywords:
CDOM
Hyperspectral imagery
Reflectance band selection
Absorption coefficient
Machine learning
Spatiotemporal distribution

ABSTRACT

Colored dissolved organic matter (CDOM) in inland waters is used as a proxy to estimate dissolved organic carbon (DOC) and may be a key indicator of water quality and nutrient enrichment. CDOM is optically active fraction of DOC so that remote sensing techniques can remotely monitor CDOM with wide spatial coverage. However, to effectively retrieve CDOM using optical algorithms, it may be critical to select the absorption coefficient at an appropriate wavelength as an output variable and to optimize input reflectance wavelengths. In this study, we constructed a CDOM retrieval model using airborne hyperspectral reflectance data and a machine learning model such as random forest. We evaluated the best combination of input wavelength bands and the CDOM absorption coefficient at various wavelengths. Seven sampling events for airborne hyperspectral imagery and CDOM absorption coefficient data from 350 nm to 440 nm over two years (2016–2017) were used, and the collected data helped train and validate the random forest model in a freshwater reservoir. An absorption coefficient of 355 nm was selected to best represent the CDOM concentration. The random forest exhibited the best performance for CDOM estimation with an R^2 of 0.85, Nash-Sutcliffe efficiency of 0.77, and percent bias of 3.88, by using a combination of three reflectance bands: 475, 497, and 660 nm. The results show that our model can be utilized to construct a CDOM retrieving algorithm and evaluate its spatiotemporal variation across a reservoir.

1. Introduction

Dissolved organic matter (DOM) is a heterogeneous mixture of molecules containing carbon, nutrients, and sulfur. In particular, colored dissolved organic matter (CDOM) absorbs shortwave radiation from ultraviolet to visible light and has a large influence on physical, chemical, and biological properties in water systems (Coble, 2007). CDOM is widely used to understand the dynamics of dissolved organic carbon (DOC) in inland and marine waters (Kutser et al., 2005; Yu et al, 2010) and may be an essential variable to assess biogeochemical process and nutrient enrichment (Tzortziou et al, 2015). Furthermore, there is a good correlation between DOC and total organic carbon (TOC) in lakes and rivers because carbon is mostly dissolved. In other words, TOC can be predicted from CDOM, which can be calculated through a proxy using remote sensing spectral data such as absorption coefficient and reflectance (Chang and Vannah, 2012; Colombo et al, 2008; Gholizadeh et al, 2016).

Many studies have used absorption coefficient at specific wavelengths as a surrogate for CDOM concentration (Griffin et al, 2018; Massicotte et al, 2017; Stedmon et al, 2000; Zhang et al, 2021a). CDOM absorption coefficients (a_{CDOM}) such as $a_{CDOM}(254)$, $a_{CDOM}(280)$, and $a_{CDOM}(320)$ were used to estimate CDOM distribution as well as to understand CDOM effects on attenuation of ultraviolet radiation (Laurion et al, 2000; Lavonen et al, 2015; McKnight et al, 2001; Morris et al, 1995; Zhang et al, 2021a). Furthermore, $a_{CDOM}(350)$, $a_{CDOM}(355)$, $a_{CDOM}(375)$, $a_{CDOM}(400)$, $a_{CDOM}(412)$, $a_{CDOM}(420)$, *and* $a_{CDOM}(440)$ have

* Corresponding author.
E-mail address: yepark@konkuk.ac.kr (Y. Park).

https://doi.org/10.1016/j.jag.2022.103053
Received 27 May 2022; Received in revised form 22 September 2022; Accepted 7 October 2022
Available online 17 October 2022

been used to remotely retrieve CDOM concentration in waters (Brezonik et al, 2015; Griffin et al, 2018, 2011; Huang et al, 2019; Kallio et al, 2008; Kutser, 2012; Liu et al, 2021; Mannino et al, 2014; Shang et al, 2021; Xu et al, 2018; Zhu et al, 2014).

Numerous remote sensing studies on CDOM have been conducted using multispectral and hyperspectral reflectance in marine waters (Brezonik et al, 2015; Cao et al, 2018; D'Sa and Miller, 2003; Griffin et al, 2018; Keith et al, 2016; Li et al, 2017; Morel and Gentili, 2009; Tehrani et al, 2013) and inland waters (Kutser et al, 2005; Zhu et al, 2011, 2014). In addition, CDOM studies have used satellite imagery from Sentinel-2 and Landsat to map CDOM. However, satellite imagery has a low spectral resolution (5–10 and 3–4 visible light bands for Sentinel and Landsat, respectively), making estimating CDOM accurately challenging. Thus, hyperspectral imagery has been presented as an innovative alternative for remote sensing as it can elucidate physical, biological, and chemical water properties with high spectral resolution (Keller et al, 2018).

The input wavelength to retrieve CDOM through remote sensing appears in the blue spectrum (less than550 nm) because the a_{CDOM} is strong in that wavelength range. In inland water systems, this blue region is greatly affected by relatively high concentrations of phytoplankton and total suspended solids, which can cause optical interference in estimating CDOM (Xu et al, 2018). Hence, extended wavelengths in the green and red region (500–700 nm) have been proposed for CDOM retrieval in turbid inland water (Kutser et al, 1998; Matthews, 2011). However, guidelines to select reflectance bands for CDOM retrieval have not yet been developed. Several methods to estimate CDOM (e.g., semi-analytical/quasi-analytical method, matrix inversion method (MIM), random forest (RF), and artificial neural network) have been proposed (Brando and Dekker, 2003; Keller et al, 2018; Lee et al, 2002; Ruescas et al, 2018; Sandidge and Holyer, 1998; Sun et al, 2011; Wang et al, 2005; Zhu et al, 2014). Among the methods, machine learning-based algorithms have exhibited the potential to handle complicated optical features of inland water with hyperspectral imagery (Keller et al, 2018; Ruescas et al, 2018). In particular, decision tree-based methods such as random forests have been preferred over other machine learning techniques (Zhang et al, 2021b). The random forest has a high level of precision, robustness, and ease of use. It is suitable for processing hyperspectral data because it can model complex relationships of data and process many input variables based on the application characteristics of decision rules (Carranza and Laborte, 2015; Wang et al, 2018).

In this study, optimal reflectance wavelengths in terms of different CDOM absorption coefficients were determined using a random forest model. The specific steps were as follows: (1) collect CDOM and characterize its values in the Baekje reservoir (BJR), (2) determine the optimal combination of input reflectance wavelengths to retrieve CDOM absorption coefficient at wavelengths using backward elimination and iterative runs of a random forest model, (3) Determine the best CDOM absorption coefficient using a random forest model with the optimal input reflectance wavelengths, (4) evaluate the overall detection performance of CDOM using accuracy index and map of the spatiotemporal distribution of CDOM by processing hyperspectral imagery in the BJR.

2. Materials and methods

2.1. Site description

The Geum River is in the midwestern part of South Korea, with a total length of 396 km and a watershed area of 9,810 km^2 (Fig. 1). The Baekje Reservoir (BJR; 36°32′N, 126°94′E) is an artificial and the southernmost among three reservoirs (Sejong, Gongju, and Baekje) in the Geum River. BJR is 311 m in length and 5.3 m in height and has a 4.20 m management water level, 7,976 km^2 watershed area, and 24 million m^3 total storage capacity. BJR was constructed starting in 2009 for agricultural and daily use and was completed in June 2012. The distance between BJR and the Geum River estuary is approximately 70 km, and the Gongju and Sejong reservoirs are 22.8 km and 41.3 km away, respectively.

Since 2012, an algal bloom caused by cyanobacteria has been a critical issue in BJR. In particular, large-scale algal blooms have

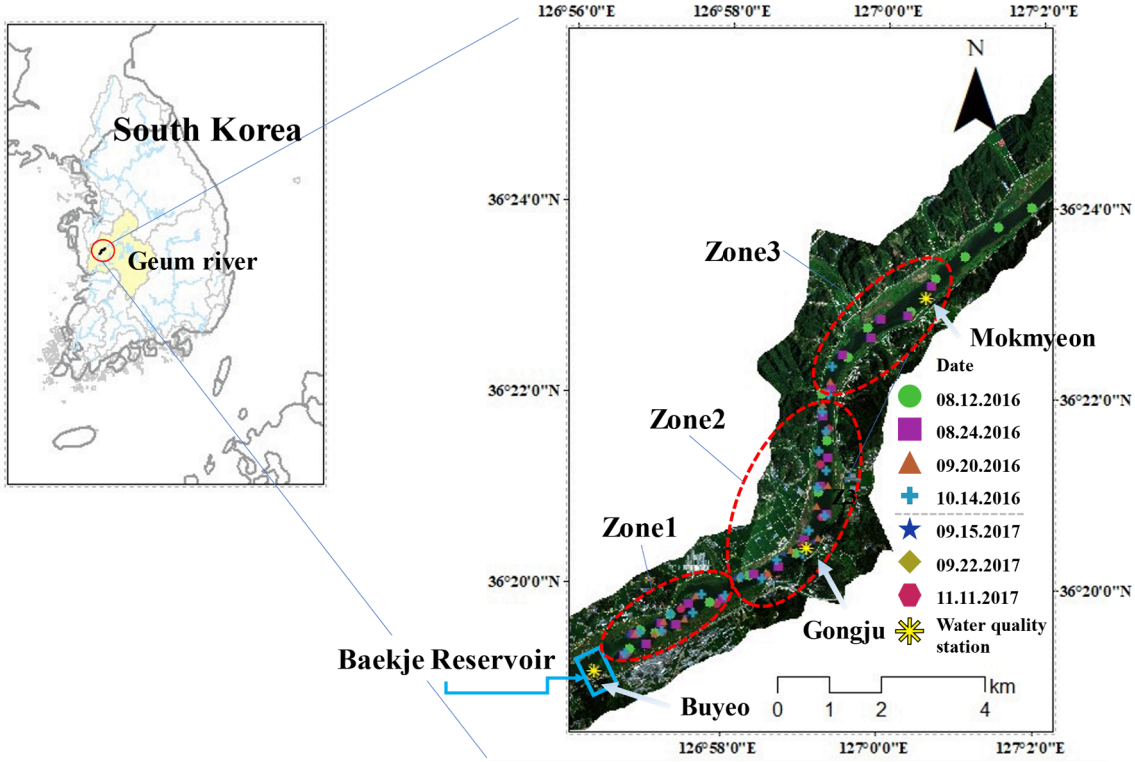


Fig. 1. Map of the Baekje reservoir (BJR) with sampling points in each monitoring event.

occurred mainly due to rainfall decreases, solar radiation increases, and an increase in water retention time in 2014 (Pyo et al, 2018). The large-scale cyanobacterial bloom in the reservoir increased the organic matter load inside the water body, which in turn increased the biochemical oxygen demand (BOD) and TOC concentrations (Kim et al, 2019). There are three water quality monitoring stations (i.e., Buyeo, Gongju, and Mokmyeon) managed by the Ministry of Environment (ME) (Fig. 1). We assessed the spatial CDOM distribution and water quality variations across the three zones of the reservoir, as indicated in Fig. 1.

2.2. Data acquisition

2.2.1. Hyperspectral imagery

Seven field campaigns including hyperspectral imaging and water sampling were performed from 2016 to 2017 (on August 12, August 24, September 20, and October 14, 2016; and September 15, September 22, and November 11, 2017), which covered both summer and fall during the algal bloom season. Water samples were collected at 108 sampling stations, with 10–20 data points per sampling event (Fig. 2). Hyperspectral images were taken with the AisaFENIX hyperspectral sensor (AISA Aero Survey Co., ltd) installed in a single-engine turboprop utility aircraft, and water samples were simultaneously obtained from the reservoir. The AisaFENIX hyperspectral sensor has a 400–970 nm spectral resolution at intervals of 4–5 nm, 127 bands in total, and a spatial resolution of 2 m. To retrieve CDOM, 66 visible bands (400–700 nm) were referenced. In situ reflectance data were collected from the water surface via a FieldSpec Handheld2 spectroradiometer (ASD Inc. Boulder, CO, USA).

Atmospheric correction of the hyperspectral imagery was performed using MODTRAN to retrieve water surface reflectance for the purpose of estimating CDOM concentration. MODTRAN, a program developed by Spectral Science Inc. and Air Force Research Laboratory, solves the radiative transfer equation through atmospheric correction parameter transmittance, solar flux, path radiance, and spherical albedo (Berk et al., 2014). MODTRAN 6 was used in this study because it was observed by Pyo et al. (2018) to yield reasonable atmospheric correction results, with an NSE value of 0.8 and RMSE of 0.0034 sr^{-1}, implying that the water surface reflectance measurements corrected by MODTRAN could be utilized to estimate CDOM concentrations. Several studies have also performed atmospheric correction of multispectral imagery using MODTRAN to calculate CDOM (Bagheri and Peters, 2004; Concha and Gerace, 2012; Ford and Vodacek, 2020; Jeon et al., 2019; Koponen et al., 2007;).

2.2.2. CDOM measurements

The CDOM sample was filtered through a Millipore polycarbonate membrane pre-rinsed with a 10 % HCI solution (0.22-μm pore size, Φ47 mm). Using a Cary 5000 UV–vis-NIR spectrophotometer, the a_{CDOM} of the filtered sample solution was estimated from 350 to 800 nm with a 1 nm interval, and the a_{CDOM} was calculated using Eq. (1).

$$a_{CDOM}(\lambda) = 2.303 \times A(\lambda)/L, \quad (1)$$

where $A(\lambda)$ is the absorption of filtered water at a specific wavelength measured across pathlength L. A quartz cuvette with 0.1 m pathlength was used in this study.

2.3. Selection of optimal input hyperspectral bands

Hyperspectral images contain many features (66 bands corresponding to the 400–700 nm visible light wavelength range), and it is necessary to efficiently find a band suitable for a CDOM search. In this case, each hyperspectral wavelength should be reviewed and reduced through band selection or feature extraction (Sun et al, 2019). When the number of variables equals or exceeds the number of observations, multicollinearity can be reduced by combining classical methods, such as backward elimination (Peerbhay et al, 2014; Richter et al, 2016). Band selection using multiple linear regression has been successfully completed in previous studies (Abdel-Rahman et al, 2013; Mutanga et

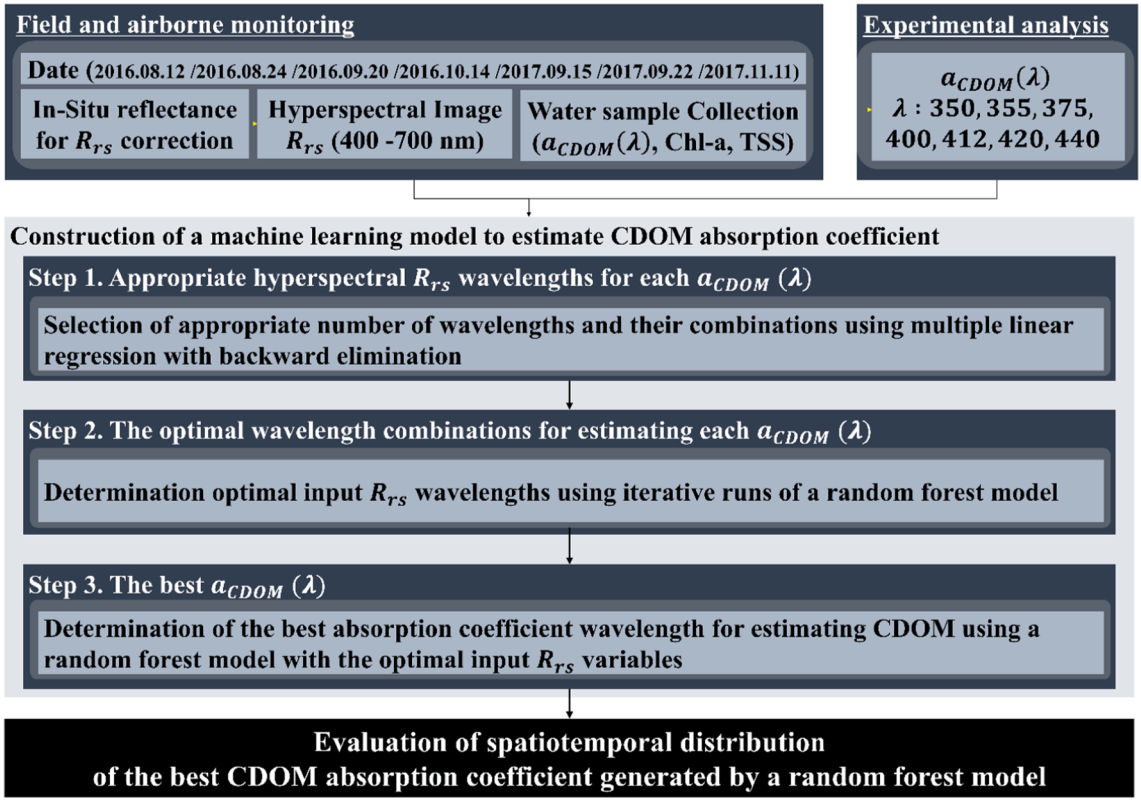


Fig. 2. Logical flow for constructing a machine learning model to estimate CDOM absorption coefficient using hyperspectral remote sensing reflectance.

al, 2004). In this study, the appropriate input variables for a CDOM algorithm were determined based on the statistical dependence of each a_{CDOM} on the 66 hyperspectral remote sensing reflectance (R_{rs}) wavelengths. Multiple linear regression with backward elimination was applied to analyze the statistical dependence between variables using R software (see step 1 in Fig. 2). The major R_{rs} wavelengths were selected by comparing the regression correlation coefficients (see Appendix C in the Supplementary Material).

Numerous combinations of the major R_{rs} wavelengths for each a_{CDOM} at wavelengths were considered as input sets in a random forest model (see step 2 in Fig. 2); the combination was calculated by $\sum_r^n nCr$, where r is the minimum number of wavelengths and n is the number of the major wavelengths. A random forest model to estimate each a_{CDOM} at wavelengths was built using the best wavelength combination in the python-based sklearn library. Hyperparameters such as ntree, max_depth, mtry, and min_sample_split were optimized GridSearchCV in sklearn. Detailed information about definition of the model including hyperparameters and optimization is presented in Appendix A in the Supplementary Material. The best wavelength combination in each a_{CDOM} was evaluated by comparing the performance of the model. The 108 data points were randomly divided into 80 % and 20 % for training and test datasets, respectively, and a 5-fold cross-validation was performed to evaluate model training. The model performance was evaluated using determination coefficient (R^2), root mean square error (RMSE), and percent bias (PB). Detailed information about the indices is described in Appendix B in the Supplementary Material.

2.4. Selection of effective CDOM absorption coefficient

The absorption coefficient at wavelengths has been used to indicate CDOM concentration (Bricaud et al. 1981). Previous studies have used satellite imagery to retrieve CDOM absorption coefficients at various wavelengths, such as 350 nm ($a_{CDOM}(350)$), 355 nm ($a_{CDOM}(355)$), 375 nm ($a_{CDOM}(375)$), 400 nm ($a_{CDOM}(400)$), 412 nm ($a_{CDOM}(412)$), 420 nm ($a_{CDOM}(420)$), and 440 nm ($a_{CDOM}(440)$) (Brezonik et al, 2015; Griffin et al, 2018, 2011; Huang et al, 2019; Kallio et al, 2008; Kutser et al, 2012; Liu et al, 2021; Mannino et al, 2014; Shang et al, 2021; Xu et al, 2018; Zhu et al, 2014). In this study, seven representative wavelengths were selected and tested to determine which wavelength can be most accurately retrieved the effective CDOM absorption coefficient using a random forest in the BJR (see step 3 in Fig. 2).

3. Results

3.1. Variations in CDOM absorption coefficient and hyperspectral reflectance spectra

Table S1 shows the absorption coefficient ($a_{CDOM}(\lambda)$) values, total suspended solid (TSS) concentration, and chlorophyll-a (Chl-a) concentration in samples measured at 108 points during the sampling period. In BJR, $a_{CDOM}(350)$ had a wide range of 2.2–11.6 m^{-1} with a mean value of 4.3 m^{-1}, and the magnitude of the a_{CDOM} value decreased exponentially from $a_{CDOM}(350)$ to $a_{CDOM}(440)$. Relatively high mean $a_{CDOM}(350)$ values were observed in August, with values of 5.2 m^{-1} and 6.4 m^{-1} on August 12 and August 24, respectively. The temporal trends of the $a_{CDOM}(355)$ values were relatively consistent with those of the observed TOC values at the three monitoring stations (see the variations in TOC (closed circles) and $a_{CDOM}(355)$ (cross) in Fig. S3). Chl-a averaged 30.9 mg/m^3 and ranged from 11.8 to 66.2 mg/m^3 during seven sampling events. The observed Chl-a concentrations irregularly fluctuated for two years (see the variations in Chl-a in Fig. S3). In 2016, the maximum Chl-a concentration was 66.2 mg/m^3 on August 12, and the maximum average Chl-a concentration was 37.2 mg/m^3 on August 24. In 2017, maximum Chl-a concentration was 47.4 mg/m^3 on September 15, while a relatively low concentration (17.6 mg/m^3) was observed on September 22. Moreover, TSS ranged between 6.0 and 23.3 mg/L, with an average value of 13.4 mg/L during the entire sensing period. The TSS concentration at the three monitoring stations fluctuated, and their variations were relatively consistent with the field survey data of this study (see TSS in Fig. S3). The observed TSS at the maximum Chl-a concentration, $a_{CDOM}(350)$ was also at its highest. However, in September 2016, although Chl-a concentration was low, the $a_{CDOM}(350)$ average ranged from 3.4 to 4.4 m^{-1}. Similarly, when the a_{CDOM} absorption coefficient was low on September 20, 2016 and October 14, 2016, TSS concentration was high at 15.1 mg/L and 16.9 mg/L, respectively. However, when a_{CDOM} concentration was high on August 12 and August 24, 2016, TSS concentration was 10.2 mg/L and 13.9 mg/L, respectively, making it difficult to establish a concrete relationship.

Fig. 3 presents the spatially averaged hyperspectral reflectance spectra for seven sampling events. Over a total of seven events in BJR, the peak hyperspectral reflectance occurred between 548 and 590 nm, while an absorption wavelength band appeared around 675 nm, followed by a peak at 700 nm. The important role of particle backscattering is corroborated by the fact that a maximum spectral reflectance peak of 570 nm was observed in most of the water spectra (Lubac and Loisel, 2007), while the peak reflectance of 700 nm was caused by low total absorption and chlorophyll fluorescence (Sun et al, 2011). However, at 675 nm, the reflectance valley was caused by chlorophyll-a absorption. In August 2016, when the a_{CDOM} was relatively high, the peaks around 548 nm exhibited lower reflectance than during other periods (Brezonik et al, 2015). In particular, during October 2016, a high a_{CDOM} value appeared as a nearly flat reflectance spectrum over the entire visible light wavelength. When a_{CDOM} was low, the peak near 590 nm caused by backscattering exhibited higher reflectance and reflectance width compared to that in other periods in 2017. In contrast, when a_{CDOM} was low in September 2016, it had a lower reflectance compared to that in other periods. However, the peaks near 590 nm and the reflectance widths were larger than those in August 2016, which likely had different effects due to high TSS concentrations (Brezonik et al, 2015).

3.2. Construction of CDOM retrieving model using different combinations of hyperspectral reflectance bands in each CDOM absorption coefficient

3.2.1. Selection of hyperspectral bands using backward elimination

Table S2 shows the regression coefficients and p-values of the bands at each CDOM absorption coefficient (i.e., $a_{CDOM}(350)$, $a_{CDOM}(355)$, $a_{CDOM}(375)$, $a_{CDOM}(400)$, $a_{CDOM}(412)$, $a_{CDOM}(420)$, and $a_{CDOM}(440)$). Fig. S1. shows variations in regression coefficient values based on ranks concerning the strength of the correlation between R_{rs} wavelengths and each CDOM absorption coefficient. Ten $\boldsymbol{R_{rs}}$ wavelengths were selected as input variables for a random forest model (Fig. S1 and Table 1). Detailed information on how to determine the number of $\boldsymbol{R_{rs}}$ wavelengths was described in Appendix C in the Supplementary Material.

3.2.2. Performance of random forest model in estimating CDOM absorption coefficient

Table 2 presents the best combination of R_{rs} wavelengths for each $a_{CDOM}(\lambda)$ obtained from 100 iterations of a random forest model. For $a_{CDOM}(350)$, the training R^2, RMSE, and PB were 0.92, 0.62, 0.28 %, respectively, and 0.89, 0.65, and 0.18 %, respectively in $a_{CDOM}(355)$. Overall, the best performance occurred at 350–355 nm. However, in test R^2, the estimation results of $a_{CDOM}(355)$ were 0.85, 0.70, and 4.73 %, respectively, resulting in a much more reasonable result than that of $a_{CDOM}(350)$. The test R^2 for $a_{CDOM}(400)$, $a_{CDOM}(412)$, $a_{CDOM}(420)$, and $a_{CDOM}(440)$ after $a_{CDOM}(440)$ was 0.67, 0.66, 0.64, and 0.47, respectively, and the RMSE was 1.02, 0.23, 0.40, and 0.20, respectively, indicating low model reliability. In addition, the PB of $a_{CDOM}(400)$ was the largest (22.65%), and the PB deviation of $a_{CDOM}(440)$ was the largest (±4.16 %). The selected band combinations varied from 3 to 6 bands. To estimate $a_{CDOM}(355)$, the combination of 660 nm, 475 nm, and 497 nm provided good performance in both train and test. When all band

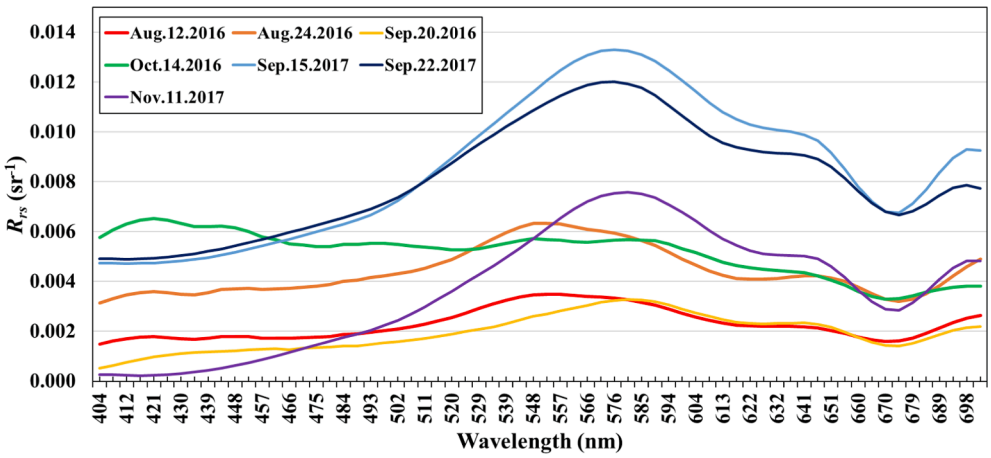


Fig. 3. Spatially averaged hyperspectral reflectance spectra for seven sampling events in BJR; each colored solid line indicates a sampling event.

Table 1
Ten hyperspectral R_{rs} wavelengths for each a_{CDOM} wavelength selected by multiple linear regression with backward elimination; R_{rs} wavelengths were ranked with respect to the strength of the correlation.

$a_{CDOM}(\lambda)$	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10
350	608	412	552	571	660	543	665	475	497	434
355	608	412	543	660	552	571	475	434	590	497
375	660	608	571	543	590	539	552	412	655	580
400	608	543	660	539	590	493	571	580	665	412
412	539	543	608	493	660	497	534	552	571	641
420	543	660	608	497	665	539	493	412	693	571
440	660	539	618	608	571	543	552	641	665	497

Table 2
Overall performance of 100 iteration model runs using absorption coefficient at each wavelength and each selected R_{rs} bands. The ± sign separates the mean performance value and standard deviation obtained from 100 iterations.

$a_{CDOM}(\lambda)$	Selected R_{rs}	Train. R^2	Test. R^2	Train. RMSE	Test. RMSE	Train. PB (%)	Test. PB (%)
350	R_{rs}608, R_{rs}660, R_{rs}543, R_{rs}665, R_{rs}497	0.92	0.77	0.62	1.54	0.28	10.82
		(±0.03)	(±0.04)	(±0.20)	(±0.34)	(±1.07)	(±1.68)
355	R_{rs}660, R_{rs}475, R_{rs}497	0.89	0.85	0.65	0.70	0.18	4.73
		(±0.05)	(±0.05)	(±0.26)	(±0.29)	(±1.26)	(±2.31)
375	R_{rs}608, R_{rs}571, R_{rs}590, R_{rs}655, R_{rs}580	0.87	0.73	0.48	0.43	−0.26	2.64
		(±0.05)	(±0.11)	(+0.15)	(±0.14)	(±1.18)	(±2.81)
400	R_{rs}608, R_{rs}539, R_{rs}590, R_{rs}493, R_{rs}580, R_{rs}665	0.91	0.67	0.25	1.02	−0.40	22.65
		(±0.04)	(±0.05)	(±0.09)	(±0.40)	(±1.63)	(±3.63)
412	R_{rs}543, R_{rs}493, R_{rs}660, R_{rs}497, R_{rs}641	0.82	0.66	0.39	0.23	0.51	4.29
		(±0.07)	(±0.12)	(±0.13)	(±0.10)	(±2.08)	(±3.46)
420	R_{rs}660, R_{rs}493, R_{rs}412, R_{rs}693	0.83	0.64	0.35	0.40	0.55	4.45
		(±0.06)	(±0.05)	(±0.09)	(±0.13)	(±2.18)	(±3.31)
440	R_{rs}660, R_{rs}608, R_{rs}543, R_{rs}552, R_{rs}665	0.80	0.47	0.32	0.20	0.25	6.48
		(±0.07)	(±0.17)	(±0.09)	(±0.06)	(±3.11)	(±4.16)

combinations were considered, model performance did not increase significantly.

3.2.3. *Determining optimal hyperspectral bands and effective CDOM absorption coefficient*

For the best performance out of 100 iterations, Fig. 4 compares the observed and simulated values for the best case and Table S3 shows the optimization parameters (ntree, max_depth, mtry, min_samples_split). In $a_{CDOM}(355)$, R^2 values were 0.91 and 0.78 and RMSE values were 0.61 and 0.95 for train and test, respectively. In addition, PB was low (0.97 % and 6.78 % for train and test, respectively) compared to other a_{CDOM}. Moreover, $a_{CDOM}(375)$ exhibited high performance in train and test with R^2 values of 0.86 and 0.80 and RMSE of 0.54 and 0.51, respectively. $a_{CDOM}(412)$ also produced good performance with R^2 of 0.81 and 0.71 and RMSE of 0.51 and 0.533, respectively. In both $a_{CDOM}(375)$ and $a_{CDOM}(412)$, PB was low (4.66 % and 4.11 %, respectively) but was weak

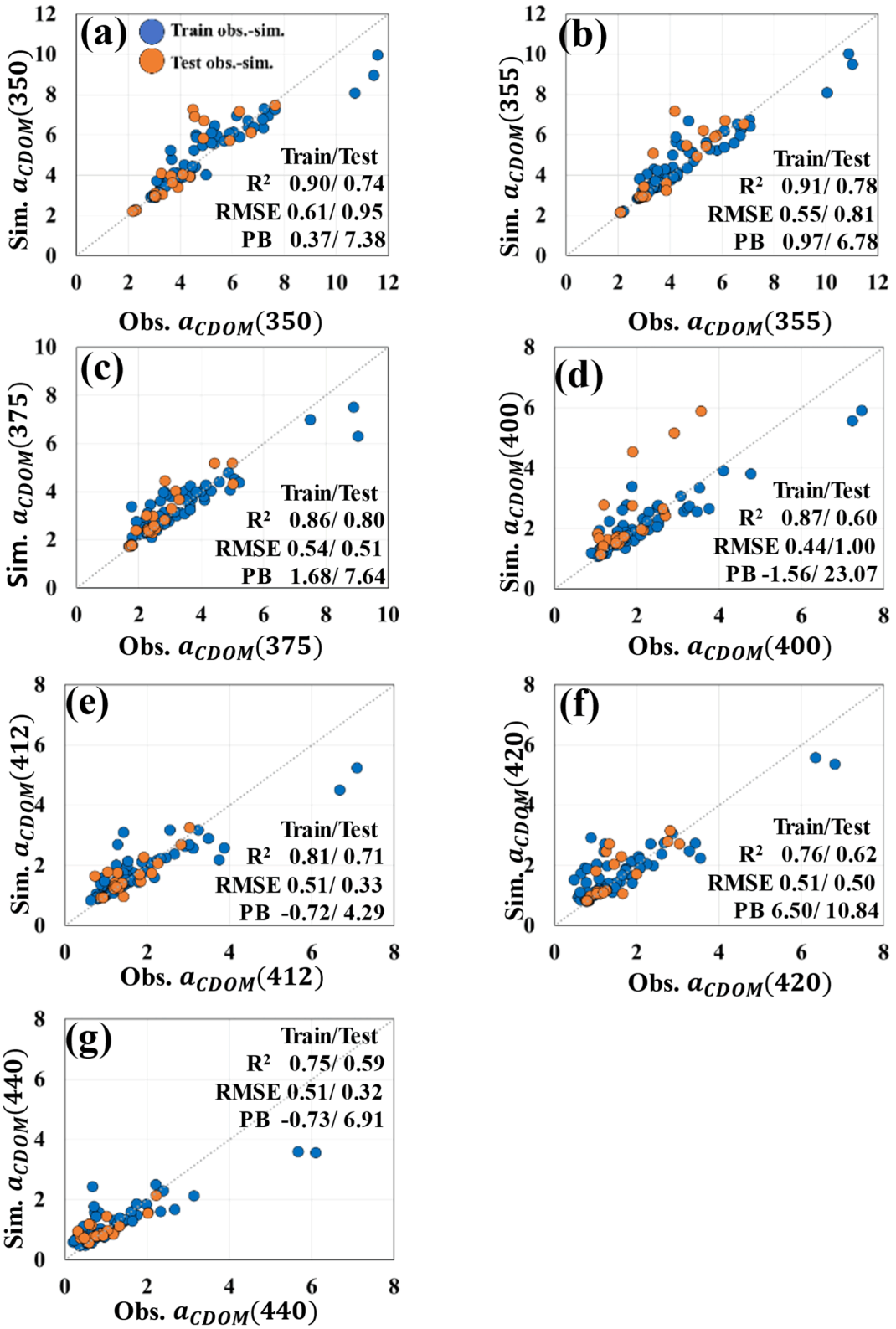


Fig. 4. Comparison of a_{CDOM} between observed and simulated values by the best random forest model for each a_{CDOM} wavelength; (a) $a_{CDOM}(350)$, (b) $a_{CDOM}(355)$, (c) $a_{CDOM}(375)$, (d) $a_{CDOM}(400)$, (e) $a_{CDOM}(412)$, (f) $a_{CDOM}(420)$, (g) $a_{CDOM}(440)$. The best random forest model was chosen from 100 iterations in Table 2. Obs. and Sim. indicate observed and simulated absorption coefficients.

when estimated at relatively high CDOM concentrations. However, $a_{CDOM}(400)$ and $a_{CDOM}(420)$ demonstrated significantly low estimation performance by having R^2 and PB values of 23.07 % and 10.84 % lower in test compared to train. In test $a_{CDOM}(440)$, PB was 6.91 %, but R^2 were 0.59, indicating that the model explanatory power was slightly inferior.

3.3. Generating the spatial and temporal distribution of CDOM absorption coefficients using the best random forest model

Three reflectance wavelengths, 475 nm, 497 nm, and 660 nm, were selected to estimate $a_{CDOM}(355)$ using the random forest model. Fig. 5 presents the spatiotemporal patterns of the a_{CDOM} with the minimum, maximum, and mean values. In terms of spatial distribution, the CDOM of Zone 1 adjacent to the reservoir on August 12, 2016, was the highest at 5.9 m^{-1}, and this area was where the actual algal blooms occurred (Fig. 5a-g). The highest mean CDOM absorption, 6.3 m^{-1}, was found on August 24, 2016 (range: 4.8–10 m^{-1}), centered on Zone 3 (Fig. 5o-u). The CDOM absorption decreased on September 20, 2016 but increased again on October 14, 2016. The overall spatial variation of CDOM was high in 2016 but primarily low throughout 2017.

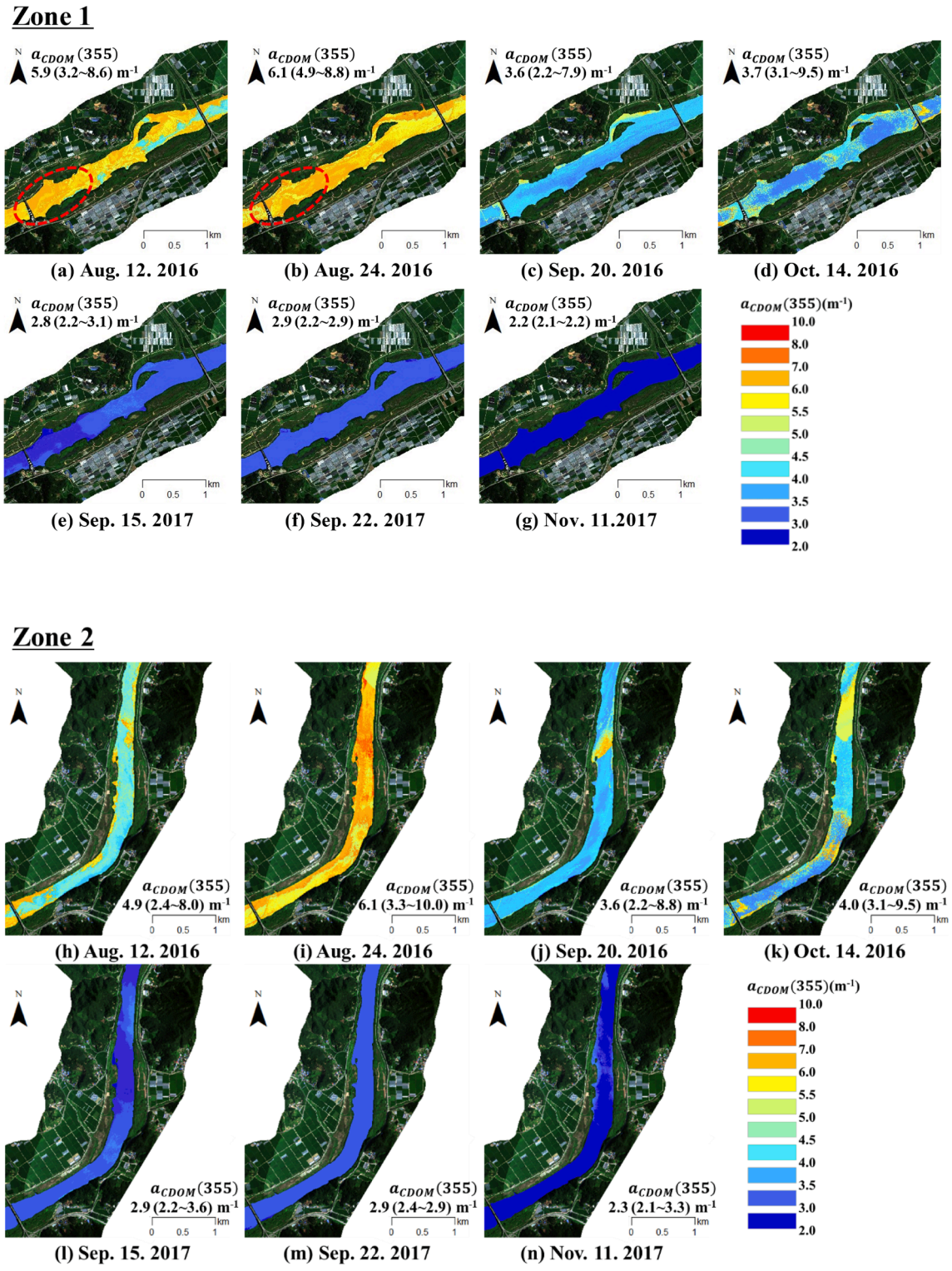


Fig. 5. Spatiotemporal distribution of CDOM absorption coefficients. Distribution of CDOM for each time in (a)~(g) Zone 1, (h)~(n) Zone 2, (o)~(u) Zone 3.

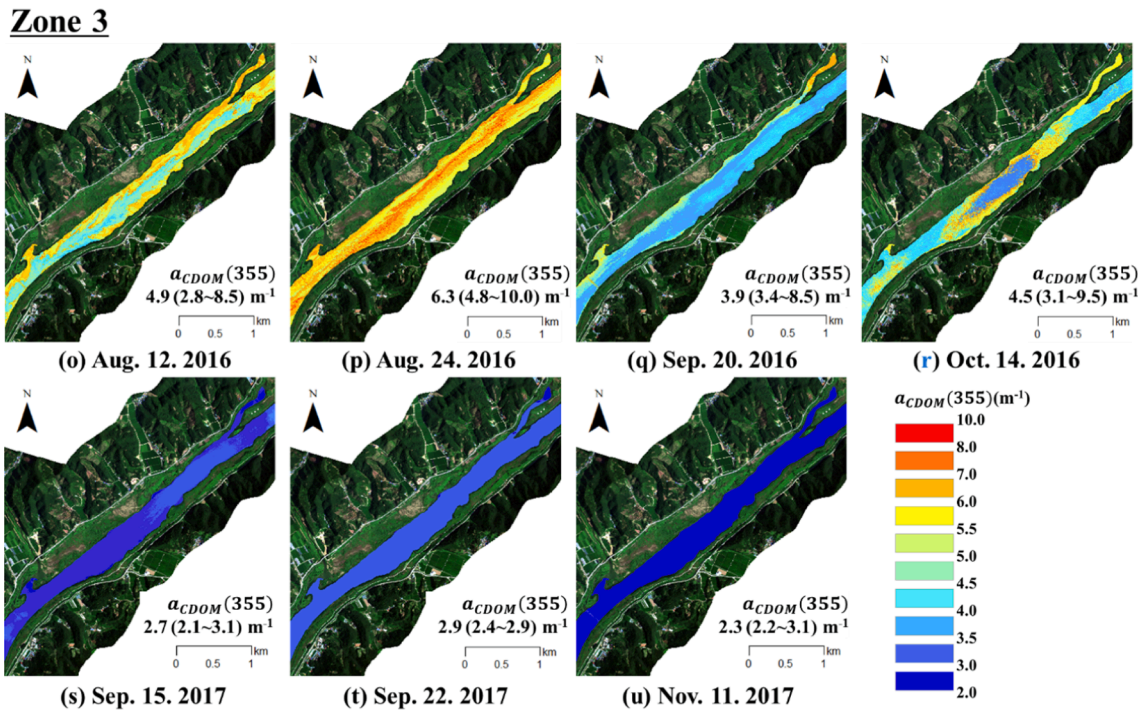


Fig. 5. (*continued*).

4. Discussion

4.1. The effective wavelength for CDOM absorption coefficient

Although CDOM estimation using remote sensing has been widely conducted, CDOM remote sensing extraction is challenging in inland turbid water because there is no accurate answer for representative wavelength selection (Brezonik et al, 2015). Table S4 summarizes studies that used remote sensing methods and wavelength combination to determine best performance for each a_{CDOM}. $a_{CDOM}(375)$, $a_{CDOM}(412)$, and $a_{CDOM}(440)$ have been more commonly retrieved from the ocean than inland waters (Castillo et al, 2000; D'sa and Miler, 2003; Gitelsen et al, 1993; Koponen et al, 2007; Kowalczuk et al, 2006; Mannino et al, 2014). Studies that used $a_{CDOM}(375)$ as the reference wavelength primarily focused on the Arctic Ocean or rivers flowing into it and used the remote sensing application program of the LOADEST (USGS load estimator) model (Griffin et al, 2018; Huang et al, 2019). $a_{CDOM}(412)$ has been mainly used by marine scientists to detect CDOM (Castillo et al, 1999; D'sa and Miler, 2003; Mannino et al, 2014). $a_{CDOM}(440)$ was mainly used because satellite ocean color sensors detect the spectrum near 440 nm (Olmanson et al, 2016; Zhu et al, 2011). $a_{CDOM}(440)$ can often appear near or less than zero for clear water because of low device detection limits (Zhang et al, 2021a).

$a_{CDOM}(350)$ is located at the center of various reference wavelength ranges (i.e., 254–440 nm) and is primarily implemented to estimate CDOM concentration and carbon cycle in rivers and lakes (Lambert et al, 2016; Zhang et al, 2021a). Overall, recent work has suggested that reference wavelength between 350 nm and 355 nm should be applied to characterize CDOM concentration (Xu et al, 2018; Liu et al, 2021; Zhang et al, 2021a). In this study, the average R^2 of the train and test values were relatively high in both $a_{CDOM}(350)$ and $a_{CDOM}(355)$. The average R^2 of the train and test of $a_{CDOM}(355)$ were 0.89 and 0.87 overall and 0.91 and 0.78 in the best case, demonstrating the best performance among all absorption coefficients (Fig. 4).

4.2. Optimal input reflectance bands to estimate CDOM in a random forest model

In this study, hyperspectral images were collected with high spatial and spectral resolution through a hyperspectral sensor mounted on an aircraft. As shown in Table S4, various studies have estimated CDOM absorption coefficients using field-measured hyperspectral devices, such as CASI, Spectrameter, hyperSAS and hyperOCR), or data from satellites, such as SeaWiFS, Landsat, and Sentinel. In addition, in this study, two input bands, 475 nm and 497 nm, appeared to be important reflectance wavelengths for CDOM retrieval. The two wavelengths are not distinguished by Landsat (Band 1 = 450–520 nm) or by Sentinel-2 (band 2 is centered on 492 nm with 66 nm bandwidth) (Brezonik et al, 2015; Griffin et al, 2011; Liu et al, 2021). Based on the previous studies, it was found that even if CDOM strongly absorbed light at a short wavelength, it may not be used to retrieve CDOM due to atmospheric scattering or interference of Chl-a (Brezonik et al, 2015; D'sa and Miler, 2003; Kutser et al, 2005;). However, Mannino et al. (2014) estimated CDOM using short wavelength reflectance, such as 412, 443, and 490 nm, taken by the SeaWiFS and MODIS-Aqua. Their multiple linear regression model using 443 nm and 547 nm showed satisfactory performance compared to other regression models using multiple wavelengths between 443 nm and 670 nm. Moreover, Brezonik et al. (2005) and Griffin et al. (2011) reported that multiple linear regression models using Landsat-based reflectance, such as B1, B2, B3, and B4, showed the best performance on estimating CDOM. Brezonik et al. (2005) found that the regression model using B1 and ratio of B1 to B4 outperformed other regression models. The previous studies inferred that reflectance below 500 nm may be used to retrieve CDOM with reflectance.

Concurrently, in the literature mentioned in Table S4, the CDOM was estimated using the peak in the mid-500 nm region. The peak in the mid-500 nm region may be determined following slopes of 475 nm and 497 nm. In Fig. S2, when the slope between the two wavelength regions was large, a peak mainly occurred between 570 nm and 580 nm. but the CDOM signal at 570 nm may vary due to the covariance of TSS (Matthews, 2011). A detailed description about a peak is mentioned in Appendix D in the Supplementary Material.

4.3. Comparing performance of machine learning approaches to retrieve CDOM absorption coefficient

CDOM detection using machine learning is straightforward to implement because it may not need further knowledge of the water quantity and/or quality parameters. In this study, the random forest exhibits strengths in alleviating the overfitting and selecting the important variable among many input variables (Ruescas et al, 2018). In research using random forests, Keller et al (2018) obtained an R^2 of 0.914 when hyperspectral data measured by on-site multi-sensor Biofish monitors were combined with principal component analysis (PCA). Moreover, Zhang et al (2021b) observed an R^2 value of 0.76 using hyperspectral data obtained with an on-site fieldspec spectroradiometer (Analytical Spectral Devices, Inc.).

Deep learning methods such as convolutional neural network and long short-term memory are useful in remote sensing research because they can model complex nonlinear effects of different parameters on water reflectance (Pahlevan et al, 2020). In predicting phycocyanin and Chl-a, Pyo et al (2019) obtained 12–62 % better accuracy using the convolutional neural network model compared to a bio-optical model in the same region. However, there are few cases for which deep learning techniques have been applied for CDOM and compared to previous techniques for estimating water quality parameters. Thus, there is room for additional data to be collected and analyzed to estimate CDOM using deep learning models. Moreover, there are various machine learning models, and a comparison of the performance between models may be considered for future research.

4.4. Spatial and temporal distribution of CDOM absorption coefficients

Compared to Brezonik et al. (2015) and Olmanson et al. (2020), the variation in CDOM was high on seasonal and annual time scales and influenced by hydrological factors and precipitation. In this study, when the streamflow was less than 100 m^3/s in summer when the water temperature was high like in 2016, relatively high CDOM values were obtained as in August 12 and August 24. On the other hand, on September 20, 2016, and on September 15 and 22, 2017, the water temperature was high, but the CDOM decreased due to the previous rainfall and discharge from the reservoir. Because of the short data periods examined here, there were limitations to understanding the long-term temporal distribution of CDOM in the BJR. Especially, The Geum river basin (including BJR) had a large CDOM source area with forests and rice paddy composing 62 % and 15 % of the catchment area, respectively (Lee et al, 2019). Long-term hyperspectral images are needed to track CDOM effects from terrestrial carbon, such as land cover changes and human land use activities (Butman and Raymond, 2011; Li et al, 2018; Yallop and Clutterbuck, 2009).

On August 12, an algal bloom hot spot was formed due to the gate operation of the hydroelectric power plant in the middle of Zone 1 (Fig. 5a-5 g) (Park et al, 2017; Pyo et al, 2018). Unlike an algal bloom hot spot in the Zone 1, the CDOM absorption coefficient was relatively low in the center and high around the edge of the reservoir. CDOM can break down into low molecular weight organic molecules, inorganic carbon, rich phosphorus, and rich nitrogen compounds that promote the growth of aquatic microorganisms when inorganic nutrients are reduced (Vähätalo and Zepp, 2005; Ylöstalo et al., 2016).

5. Conclusions

This study investigated the spatial distribution of CDOM absorption coefficients that were estimated from the reflectance of hyperspectral imagery and a random forest model. We compared seven reference wavelengths between 350 and 440 nm for CDOM retrieval, using the 10 most correlated reflectance wavelengths for each reference wavelength. In the overall test, the performance of $a_{CDOM}(355)$ was the best with R^2 of 0.87, NSE of 0.71, and PB of 2.46. The best combinations were $R_{rs}660$, $R_{rs}475$, and $R_{rs}497$. In addition, CDOM spatial distribution variability was well-captured by the trained random forest model. Thus, this study identified the capabilities of hyperspectral image-based random forests in understanding CDOM in optically complex inland water. These data from deep learning models could be applied in future to investigate the causes of temporal and spatial fluctuations in CDOM distributions, along with water quality factors such as phycocyanin and chlorophyll-a. Meanwhile, using drone- and aircraft-based hyperspectral images in regular monitoring may be quite challenging because of efficiency issues on cost, labor, time, and spatial coverage. A comprehensive study on application of hyperspectral imaging may be needed in a practical perspective.

CRediT authorship contribution statement

Jinuk Kim: Methodology, Investigation, Writing – original draft. **Wonjin Jang:** Software, Formal analysis. **Jin Hwi Kim:** Writing – review & editing. **Jiwan Lee:** Investigation. **Kyung Hwa Cho:** Conceptualization, Resources. **Yong-Gu Lee:** Formal analysis. **Kangmin Chon:** Formal analysis. **Sanghyun Park:** Resources. **JongCheol Pyo:** Writing – review & editing. **Yongeun Park:** Writing – review & editing, Supervision, Funding acquisition. **Seongjoon Kim:** Conceptualization, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This paper was supported by Konkuk University in 2019.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2022.103053.

References

Abdel-Rahman, E.M., Ahmed, F.B., Ismail, R., 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. International Journal of Remote Sensing. 34 (2), 712–728. https://doi.org/10.1080/01431161.2012.713142.

Bagheri, S., Peters, S., 2004. Retrieval of marine water constituents using atmospherically corrected AVIRIS hyperspectral data. In: In ProceedIngs of the 12th JPL Airborne Earth Science Workshop. https://doi.org/10.1080/0143116042000274023.

Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., & Van Den Bosch, J. (2014, June). MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code. In 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-4). IEEE. https://doi.org/10.1109/WHISPERS.2014.8077573.

Brando, V.E., Dekker, A.G., 2003. Satellite hyperspectral remote senseing for estimating estuarine and coastal water quality. IEEE Transactions on Geoscience and Remote Sensing. 41, 1378–1387. https://doi.org/10.1109/TGRS.2003.812907.

Brezonik, P.L., Menken, K.D., Bauer, M., 2005. Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). Lake and Reservoir Management. 21 (4), 373–382. https://doi.org/10.1080/07438140509354442.

Brezonik, P.L., Olmanson, L.G., Finlay, J.C., Bauer, M.E., 2015. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. Remote Sensing of Environment. 157, 199–215. https://doi.org/10.1016/j.rse.2014.04.033.

Bricaud, A., Morel, A., Prieur, L., 1981. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. Limnology and Oceanography. 26 (1), 43–53. https://doi.org/10.4319/lo.1981.26.1.0043.

Butman, D., Raymond, P.A., 2011. Significant efflux of carbon dioxide from streams and rivers in the United States. Nature Geoscience. 4, 839–842. https://doi.org/10.1038/ngeo1294.

Cao, F., Tzortziou, M., Hu, C.M., Mannino, A., Fichot, C.G., Del Vecchio, R., Najjar, R.G., Novak, M., 2018. Remote sensing retrievals of colored dissolved organic matter and dissolved organic carbon dynamics in north American estuaries and their margins. Remote Sensing of Environment. 205, 151–165. https://doi.org/10.1016/j.rse.2017.11.014.

Carranza, E.J.M., Laborte, A.G., 2015. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). Computers & Geosciences. 74, 60–70. https://doi.org/10.1016/j.cageo.2014.10.004.

Castillo, C.E., Coble, P.G., Morell, J.M., López, J.M., Corredor, J.E., 1999. Analysis of the optical properties of the Orinoco river plume by absorption and fluorescence spectroscopy. Marine Chemistry. 66 (1–2), 35–51. https://doi.org/10.1016/S0304-4203(99)00023-7.

Castillo, C.E., Gilbes, F., Coble, P.G., Müller-Karger, F.E., 2000. On the dispersal of riverine colored dissolved organic matter over the west florida shelf. Limnology Oceanography. 45, 1425–1432. https://doi.org/10.4319/lo.2000.45.6.1425.

Chang, N.B., Vannah, B.W., 2012. Monitoring the total organic carbon concentrations in a lake with the integrated data fusion and machine-learning (IDFM) technique. In: Proc. SPIE 8513, Remote Sensing and Modeling of Ecosystems for Sustainability IX, p. 851307. https://doi.org/10.1117/12.927632.

Coble, P.G., 2007. Marine optical biogeochemistry: the chemistry of ocean color. Chemical Reviews. 107, 402–418. https://doi.org/10.1021/cr050350+.

Colombo, R., Mernoni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., Panigada, C., 2008. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. Remote Sensing of Environment. 112 (4), 1820–1834. https://doi.org/10.1016/j.rse.2007.09.005.

Concha, J. A., Gerace, A. D. 2012. Atmospheric compensation for WorldView-2 satellite and in-water component retrieval. Proc. SPIE 8390, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, 83900W. https://doi.org/10.1117/12.918962.

D'Sa, E.J., Miller, R.L., 2003. Bio-optical properties in waters influenced by the Mississippi River during low flow conditions. Remote Sensing of Environment. 84, 538–549. https://doi.org/10.1016/S0034-4257(02)00163-3.

Ford, R.T., Vodacek, A., 2020. Determining improvements in Landsat spectral sampling for inland water quality monitoring. Science of Remote Sensing 1, 100005. https://doi.org/10.1016/j.srs.2020.100005.

Gholizadeh, M.H., Melesse, A.M., Reddi, L., 2016. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. Sensors. 16 (8), 1298. https://doi.org/10.3390/s16081298.

Griffin, C.G., Frey, K.E., Rogan, J., Holmes, R.M., 2011. Spatial and interannual variability of dissolved organic matter in the Kolyma River, East Siberia, observed using satellite imagery. Journal of Geophysical Research. 116, G3. https://doi.org/10.1029/2010JG001634.

Griffin, C.G., McClelland, J.W., Frey, K.E., Fiske, G., Holmes, R.M., 2018. Quantifying CDOM and DOC in major Arctic rivers during ice-free conditions using Landsat TM and ETM+ data. Remote Sensing of Environment. 209, 395–409. https://doi.org/10.1016/j.rse.2018.02.060.

Huang, J., Wu, M., Cui, T., Yang, F., 2019. Quantifying DOC and Its Controlling Factors in Major Arctic Rivers during Ice-Free Conditions using Sentinel-2 Data. Remote Sensing. 11 (24), 2904. https://doi.org/10.3390/rs11242904.

Jeon, E.I., Kang, S.J., Lee, K.Y., 2019. Estimation of chlorophyll-a concentration with semi-analytical algorithms using airborne hyperspectral imagery in Nakdong river of South Korea. Spatial Information Research 27 (1), 97–107. https://doi.org/10.1007/s41324-018-0204-0.

Kallio, K., Attila, J., Härmä, P., Koponen, S., Pulliainen, J., Hyytiäinen, U.M., Pyhälahti, T., 2008. Landsat ETM+ images in the estimation of seasonal Lake water quality in Boreal River basins. Environmental Management. 42, 511–522. https://doi.org/10.1007/s00267-008-9146-y.

Keith, D., Lunetta, R., Schaeffer, B., 2016. Optical models for remote sensing of colored dissolved organic matter absorption and salinity in New England, Middle Atlantic and Gulf Coast Estuaries USA. Remote Sensing. 8 (4), 283. https://doi.org/10.3390/rs8040283.

Keller, S., Maier, P.M., Riese, F.M., Norra, S., Holbach, A., Borsig, N., Wilhelms, A., Moldaenke, C., Zaake, A., Hinz, S., 2018. Hyperspectral Data and Machine Learning for Estimating CDOM, Chlorophyll a, Diatoms, Green Algae and Turbidity. Environmental Research and Public Health. 15 (9), 1881. https://doi.org/10.3390/ijerph15091881.

Kim, S.J., Chung, S.W., Park, H.S., Cho, Y.C., Lee, H.S., Park, Y.J., 2019. Analysis of Environmental Factors Associated with Cyanobacteria Dominance in Baekje Weir and Juksan Weir. Journal of Korean Society on Water Environment. 35 (3), 257–270. https://doi.org/10.15681/KSWE.2019.35.3.257.

Koponen, S., Attila, J., Pulliainen, J., Kallio, K., Pyhälahti, T., Lindfors, A., Rasmus, K., Hallikainen, M., 2007. A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland. Continental Shelf Research. 27 (2), 228–244. https://doi.org/10.1016/j.csr.2006.10.006.

Kowalczuk, P., Stedmon, C.A., Markager, S., 2006. Modeling absorption by CDOM in the Baltic Sea from season, salinity and chlorophyll. Marine Chemistry. 101 (1–2), 1–11. https://doi.org/10.1016/j.marchem.2005.12.005.

Kutser, T., 2012. The possibility of using the Landsat image archive for monitoring long time trends in coloured dissolved organic matter concentration in lake waters. Remote Sensing of Environment. 123 (3), 334–338. https://doi.org/10.1016/j.rse.2012.04.004.

Kutser, T., Arst, H., Mäekivi, S., Kallaste, K., 1998. Estimation of the water quality of the Baltic Sea and lakes in Estonia and Finland by passive optical remote sensing measurements on board vessel. Lakes and Reservoirs: Research and Management. 3, 53–66. https://doi.org/10.1111/j.1440-1770.1998.tb00032.x.

Kutser, T., Pierson, D.C., Kallio, K.Y., Reinart, A., Sobek, S., 2005. Mapping lake CDOM by satellite remote sensing. Remote Sensing of Environment. 94 (4), 535–540. https://doi.org/10.1016/j.rse.2004.11.009.

Lambert, T., Teodoru, C.R., Nyoni, F.C., Bouillon, S., Darchambeau, F., Massicotte, P., Borges, A.V., 2016. Along-stream transport and transformation of dissolved organic matter in a large tropical river. Biogeosciences. 13 (9), 2727–2741. https://doi.org/10.5194/bg-13-2727-2016.

Laurion, I., Ventura, M., Catalan, J., Psenner, R., Sommaruga, R., 2000. Attenuation of ultraviolet radiation in mountain lakes: factors controlling the among-and within-lake variability. Limnology and Oceanography. 45 (6), 1274–1288. https://doi.org/10.4319/lo.2000.45.6.1274.

Lavonen, E.E., Kothawala, D.N., Tranvik, L.J., Gonsior, M., Schmitt-Kopplin, P., Köhler, S.J., 2015. Tracking changes in the optical properties and molecular composition of dissolved organic matter during drinking water production. Water Research. 85, 286–294. https://doi.org/10.1016/j.watres.2015.08.024.

Lee, Z.P., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. Applied Optics. 41, 5755–5772. https://doi.org/10.1364/AO.41.005755.

Lee, J.W., Jung, C.G., Kim, S.H., Kim, S.J., 2019. Assessment of Climate Change Impact on Future Groundwater-Level Behavior Using SWAT Groundwater-Consumption Function in Geum River Basin of South Korea. Water. 11 (5), 949. https://doi.org/10.3390/w11050949.

Li, J., Yu, Q., Tian, Y.Q., Becker, B.L., 2017. Remote sensing estimation of colored dissolved organic matter (CDOM) in optically shallow waters. ISPRS Journal of Photogrammetry and Remote Sensing. 128, 98–110. https://doi.org/10.1016/j.isprsjprs.2017.03.015.

Li, J., Yu, Q., Tian, Y.Q., Becker, B.L., Siqueira, P., Torbick, N., 2018. Spatio-temporal variations of CDOM in shallow inland waters from a semi-analytical inversion of Landsat-8. Remote Sensing of Environment. 218, 189–200. https://doi.org/10.1016/j.rse.2018.09.014.

Liu, G., Li, S., Song, K., Wang, X., Wen, Z., Kutser, T., Jacinthe, P.A., Shang, Y., Lyu, L., Fang, C., Yang, Y., Yang, Q., Zhang, B., Cheng, S., Hou, J., 2021. Remote sensing of CDOM and DOC in alpine lakes across the Qinghai-Tibet Plateau using Sentinel-2A imagery data. Journal of Environmental Management. 286, 112231 https://doi.org/10.1016/j.jenvman.2021.112231.

Lubac, B., Loisel, H., 2007. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. Remote Sensing of Environment. 110 (1), 45–58. https://doi.org/10.1016/j.rse.2007.02.012.

Mannino, A., Novak, M.G., Hooker, S.B., Hyde, K., Aurin, D., 2014. Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. Remote Sensing of Environment. 152, 576–602. https://doi.org/10.1016/j.rse.2014.06.027.

Massicotte, P., Asmala, E., Stedmon, C., Markager, S., 2017. Global distribution of dissolved organic matter along the aquatic continuum: Across rivers, lakes and oceans. Science of The Total Environment. 609, 180–191. https://doi.org/10.1016/j.scitotenv.2017.07.076.

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. International Journal of Remote Sensing. 32 (21), 6855–6899. https://doi.org/10.1080/01431161.2010.512947.

McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnology and Oceanography. 46 (1), 38–48. https://doi.org/10.4319/lo.2001.46.1.0038.

Morel, A., Gentili, B., 2009. A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data. Remote Sensing of Environment. 113, 998–1011. https://doi.org/10.1016/j.rse.2009.01.008.

Morris, D.P., Zagarese, H., Williamson, C.E., Balseiro, E.G., Hargreaves, B.R., Modenutti, B., Moeller, R., Queimalinos, C., 1995. The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. Limnology and Oceanography. 40 (8), 1381–1391. https://doi.org/10.4319/lo.1995.40.8.1381.

Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting In Situ Pasture Quality in Kruger National Park, South Africa, Using Continuum-Removed Absorption Features. Remote Sensing of Environment. 89 (3), 393–408. https://doi.org/10.1016/j.rse.2003.11.001.

Olmanson, L.G., Brezonik, P.L., Finlay, J.C., Bauer, M.E., 2016. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. Remote Sensing of Environment. 185, 119–128. https://doi.org/10.1016/j.rse.2016.01.007.

Olmanson, L.G., Page, B.P., Finlay, J.C., Brezonik, P.L., Bauer, M.E., Griffin, C.G., Hozalski, R.M., 2020. Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. Science of The Total Environment. 724, 138141 https://doi.org/10.1016/j.scitotenv.2020.138141.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Ha, N., Mastsushita, B., Moses, W., Greb, S., Lehmann, M., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment 240, 111604. https://doi.org/10.1016/j.rse.2019.111604.

Park, Y.E., Pyo, J.C., Kwon, Y.S., Cha, Y.K., Lee, H., Kang, T.G., Cho, K.H., 2017. Evaluating physico-chemical influences on cyanobacterial blooms using

hyperspectral images in inland water. Korea. Water Research. 126, 319–328. https://doi.org/10.1016/j.watres.2017.09.026.
Peerbhay, K.Y., Mutanga, O., Ismail, R. 2014. Does simultaneous variable selection and dimension reduction improve the classification of Pinus forest species?. Journal of Applied Remote Sensing, 8, 085194. https://doi.org/10.1117/1.jrs.8.085194.
Pyo, J.C., Ligaray, M., Kwon, Y.S., Ahn, M.H., Kim, K.H., Lee, H., Kang, T.G., Cho, S.B., Park, Y.E., Cho, K.H., 2018. High-Spatial Resolution Monitoring of Phycocyanin and Chlorophyll-a Using Airborne Hyperspectral Imagery. Remote Sensing. 10, 1180. https://doi.org/10.1016/j.rse.2019.111350.
Pyo, J.C., Duan, H., Baek, S.S., Kim, M.S., Jeon, T.G., Kwon, Y.S., Lee, H., Cho, K.H., 2019. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. Remote Sensing of Environment. 233, 111350 https://doi.org/10.1016/j.rse.2019.111350.
Richter, R., Reu, B., Wirth, C., Doktor, D., Vohland, M., 2016. The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. International Journal of Applied Earth Observation and Geoinformation 52, 464–474. https://doi.org/10.1016/j.jag.2016.07.018.
Ruescas, A.B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K., Camps-Valls, G. C., 2018. Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data. Remote Sensing. 10, 786. https://doi.org/10.3390/rs10050786.
Sandidge, J.C., Holyer, R.J., 1998. Coastal bathymetry from hyperspectral observations of water radiance. Remote Sensing of Environment. 65 (3), 341–352. https://doi.org/10.1016/S0034-4257(98)00043-1.
Shang, Y., Liu, G., Wen, Z., Jacinthe, P.A., Song, K., Zhang, B., Lyu, L., Li, S., Wang, X., Yu, X., 2021. Remote estimates of CDOM using Sentinel-2 remote sensing data in reservoirs with different trophic states across China. Journal of Environmental Management. 286, 112275 https://doi.org/10.1016/j.jenvman.2021.112275.
Stedmon, C.A., Markager, S., Kaas, H., 2000. Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. Estuarine, Coastal and Shelf Science. 51, 267–278. https://doi.org/10.1006/ecss.2000.0645.
Sun, D.Y., Li, Y.M., Wang, Q., Lu, H., Le, C.F., Huang, C.C., Gong, S.Q., 2011. A neural-network model to retrieve CDOM absorption from in situ measured hyperspectral data in an optically complex lake: Lake Taihu case study. International Journal of Remote Sensing. 32 (14), 4005–4022. https://doi.org/10.1080/01431161.2010.481297.
Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., Wang, P., 2019. Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. Computers and Electronics in Agriculture. 160, 153–159. https://doi.org/10.1016/j.compag.2019.03.004.
Tehrani, N.C., D'Sa, E.J., Osburn, C.L., Bianchi, T.S., Schaeffer, B.A., 2013. Chromophoric dissolved organic matter and dissolved organic carbon from sea-viewing wide field-of-view sensor (SeaWiFS), moderate resolution imaging spectroradiometer (MODIS) and MERIS sensors: case Study for the Northern Gulf of Mexico. Remote Sensing. 5, 1439–1464. https://doi.org/10.3390/rs5031439.
Tzortziou, M., Zeri, C., Dimitriou, E., Ding, Y., Jaffé, R., Anagnostou, E., Pitta, E., Mentzafou, A., 2015. Colored dissolved organic matter dynamics and anthropogenic influences in a major transboundary river and its coastal wetland. Limnology and oceanography 60 (4), 1222–1240. https://doi.org/10.1002/lno.10092.
Vähätalo, A.V., Zepp, R.G., 2005. Photochemical mineralization of dissolved organic nitrogen to ammonium in the Baltic Sea. Environmental Science & Technology. 39 (18), 6985–6992. https://doi.org/10.1021/es050142z.
Wang, P., Boss, E.S., Roesler, C., 2005. Uncertainties of inherent optical properties obtained from semianalytical inversions of ocean color. Applied Optics. 44, 4074–4085. https://doi.org/10.1364/AO.44.004074.
Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B.o., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. Applied Water Science. 8, 125. https://doi.org/10.1007/s13201-018-0742-6.
Xu, J., Fang, C., Gao, D., Zhang, H., Gao, C., Xu, Z., Wang, Y., 2018. Optical models for remote sensing of chromophoric dissolved organic matter (CDOM) absorption in Poyang Lake. ISPRS Journal of Photogrammetry and Remote Sensing. 142, 124–136. https://doi.org/10.1016/j.isprsjprs.2018.06.004.
Yallop, A.R., Clutterbuck, B., 2009. Land management as a factor controlling dissolved organic carbon release from upland peat soils 1: Spatial variation in DOC productivity. Science of The Total Environment 407 (12), 3803–3813.
Ylöstalo, P., Seppälä, J., Kaitala, S., Maunula, P., Simis, S., 2016. Loadings of dissolved organic matter and nutrients from the Neva River into the Gulf of Finland – Biogeochemical composition and spatial distribution within the salinity gradient. Marine Chemistry. 186, 58–71. https://doi.org/10.1016/j.marchem.2016.07.004.
Yu, Q., Tian, Y.Q., Chen, R.F., Liu, A., Gardner, G.B., Zhu, W., 2010. Functional linear analysis of in situ hyperspectral data for assessing CDOM in rivers. Photogrammetric Engineering and Remote Sensing. 76, 1147–1158. https://doi.org/10.14358/PERS.76.10.1147.
Zhang, Y., Zhou, L., Zhou, Y., Zhang, L., Yao, X., Shi, K., Jeppesen, E., Yu, Q., Zhu, W., 2021a. Chromophoric dissolved organic matter in inland waters: Present knowledge and future challenges. Science of The Total Environment. 759, 143550 https://doi.org/10.1016/j.scitotenv.2020.143550.
Zhang, Z., Zhu, W., Chen, J., Cheng, Q., 2021b. Remotely observed variations of reservoir low concentration chromophoric dissolved organic matter and its response to upstream hydrological and meteorological conditions using Sentinel-2 imagery and Gradient Boosting Regression Tree. Water Supply. 21 (2), 668–682. https://doi.org/10.2166/ws.2020.342.
Zhu, W., Yu, Q., Tian, Y.Q., Chen, R.F., Gardner, G.B., 2011. Estimation of chromophoric dissolved organic matter in the Mississippi and Atchafalaya river plume regions using above-surface hyperspectral remote sensing. Journal of Geophysical Research. 116, C2. https://doi.org/10.1029/2010JC006523.
Zhu, W., Yu, Q., Tian, Y.Q., Becker, B.L., Zheng, T., Carrick, H.J., 2014. An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments. Remote Sensing of Environment. 140, 766–778. https://doi.org/10.1016/j.rse.2013.10.015.