

**Title:** Evaluating the potential of Sentinel-2 and Landsat images for mapping open surface water body areas and water quality in Oklahoma

**Authors' Names and Affiliations:**

Zhenhua Zou, Ph.D. Student, 101 David L. Boren Blvd, Norman, OK 73019, (405)3437335, [zhua.zou@ou.edu](mailto:zhua.zou@ou.edu), University of Oklahoma, Department of Microbiology and Plant Biology and Center for Spatial Analysis.

Xiangming Xiao, Professor, 101 David L. Boren Blvd, Norman, OK 73019, (405) 325-8941, [xiangming.xiao@ou.edu](mailto:xiangming.xiao@ou.edu), University of Oklahoma, Department of Microbiology and Plant Biology and Center for Spatial Analysis.

**Start Date:** June 16, 2018

**End Date:** June 17, 2019

**Congressional District:** 4

Focus Category: SW, HYDROL

**Descriptors:** Surface Water, Mapping, Water Quality, Remote Sensing

**Students:**

Student Status	Number	Disciplines
Undergraduate	0	
M.S.	0	
Ph.D.	1	Remote sensing, Ecology
Post Doc	0	
Total	1	

**Principal Investigators:**

Xiangming Xiao, Professor, Department of Microbiology and Plant Biology and Center for Spatial Analysis, University of Oklahoma, Norman, OK.

Zhenhua Zou, Ph.D. Student, Department of Microbiology and Plant Biology and Center for Spatial Analysis, University of Oklahoma, Norman, OK.

**Publications and Presentations:**

Zou,Z. & X.Xiao. 2018. "Divergent trends in surface water area". Poster presentation at 2018 GIS Day at the University of Oklahoma, Norman, OK.

## **Problem and Research Objectives:**

Open surface water bodies (e.g., rivers, streams, lakes, reservoirs, ponds) are important water sources for agriculture, energy, commerce, industry, and public water supply. The numerous open surface water bodies across Oklahoma provide ~64% of the total fresh water withdrawal in Oklahoma (USGS, 2010). Existing surface water body maps are at 30-m spatial resolution (Feng et al., 2016; Homer et al., 2015; Pekel et al., 2016). Small streams and ponds, important for agricultural production and biodiversity conservation, are usually not shown on these maps. There is a need for annual maps of surface water body at 10-m spatial resolution. Thus, we used Sentinel-2 satellite images to generate a statewide 10-m water body frequency map of 2018.

Chlorophyll-a (Chl-a) concentration in surface water body is one of the most commonly used indicators of water quality. The Oklahoma Water Resource Board (OWRB) has traditionally used the Carlson's Trophic State Index (TSI), derived using Chlorophyll-a concentration (Carlson, 1977), to report lake trophic status. Based on 2014-2015 sampling data (OWRB, 2015), 6 lakes were hypereutrophic ( $TSI \geq 61$ ) and 22 lakes were eutrophic ( $60 \geq TSI \geq 51$ ), covering 7% and 88% of the total surface water areas sampled, respectively. While lab measurement of Chlorophyll-a concentration using water samples collected from the field can get the accurate results, this approach is time consuming and expensive. Thus, the Beneficial Use Monitoring Program (BUMP) collects water samples from the same sampling site in every year or every two to three years. As it covers only a small fraction of surface water bodies in Oklahoma at a given year, this strategy might miss the algal blooms and fail to capture the seasonality of Chlorophyll-a concentration. Landsat satellites scan the entire earth in every 16 days since 1980s while Sentinel-2 satellites scan the entire earth in every 10 days since June 2015. If robust relationships can be found between chlorophyll-a field measurements and satellite remote sensing data, we can estimate chlorophyll-a concentration directly using Landsat and Sentinel-2 images. Satellite-image based water quality detection has the potential to provide Chlorophyll-a concentration information of the entire state in high temporal frequencies and spatial resolution, which could provide an alternative and supplement to the expensive in-situ water sample collection and lab measurement.

## **Methodology:**

*Surface Water Body Area:* Terrestrial landscapes are often composed of water, soils and vegetation. The relationship between water and vegetation indices obtained from satellite imagery can be used to detect open surface water bodies (Xiao et al., 2006; Zou et al., 2017, Zou et al., 2018). We classify an observation in a image pixel as surface water body if its water signal is stronger than vegetation signal. First, we identify the water signal using the criteria  $mNDWI > EVI$  or  $mNDWI > NDVI$ . Second, we identify the non-vegetation signal (e.g., soils, water) using the criterion  $EVI < 0.1$ . Finally, those

observations in image pixels that met both criteria ((mNDWI > NDVI or mNDWI > EVI) and (EVI < 0.1)) were classified as open surface water body, while the remaining observations were classified as non-surface water body (Zou et al., 2017, Zou et al., 2018). We applied the above algorithms to all Sentinel-2 images of 2018 within the boundary of Oklahoma and generate a statewide annual water body frequency map at 10-m spatial resolution.

*Surface Water Quality:* We received 11,851 chlorophyll-a concentration field measurement records from the Oklahoma Water Resource Board. These measurements were taken from the major lakes, reservoirs, and rivers during 2002–2016 by the Beneficial Use Monitoring Program in Oklahoma Water Resource Board. Most of the water samples were collected from ~0.5m depth. There were 908 water sample collection sites, distributing across the entire Oklahoma.

For each chlorophyll-a measurement record, its geographic coordinates of water sampling site were added to google earth engine. The image data of a pixel at the water sampling site was extracted from a Landsat image, which has the closest acquisition date to the water sampling date compared to other Landsat images. In similar approach, a pixel was also extracted from a Sentinel 2 image (Table 1). Out of the 11,851 chlorophyll-a measurement records, we successfully extracted 11,369 Landsat pixels, of which 4938 were from Landsat 5, 5507 were from Landsat 7, and 924 were from Landsat 8. About 7% (816) of the extracted Landsat pixels were acquired by the satellites within 1 day of the field water sample collection, while ~80% (9082) of the extracted Landsat pixels were acquired by the satellites within 10 days of the field water sample collection (Table 2). In similar approach, we successfully extracted 176 Sentinel 2 pixels, of which 19% (34) were acquired by the satellites within 10 days of the field water sample collection. There were much less Sentinel 2 pixels because Sentinel 2A satellite was launched in Jun. 2015 and only a small portion of the chlorophyll-a measurements were acquired after that. In comparison, there are two Landsat satellites at work in most of the chlorophyll-a measurement period (2002–2016).

**Table 1** Band name and wavelength of Landsat 5/7/8 and Sentinel 2 images

Landsat 5/7		Landsat8		Sentinel 2A	
Band name	Wavelength (μm)	Band name	Wavelength (μm)	Band name	Wavelength (μm)
B1 (blue)	0.45-0.52	B1 (ultra-blue)	0.435-0.451	B1 (Aerosols)	0.4292-0.4562
B2 (green)	0.52-0.60	B2 (blue)	0.452-0.512	B2 (blue)	0.4434-0.5414
B3 (red)	0.63-0.69	B3 (green)	0.533-0.590	B3 (green)	0.5373-0.5823
B4 (near infrared)	0.77-0.90	B4 (red)	0.636-0.673	B4 (red)	0.6456-0.6836
B5 (shortwave infrared 1)	1.55-1.75	B5 (near infrared)	0.851-0.879	B5 (red edge 1)	0.6946-0.7136
B6 (brightness temperature)	10.40-12.50	B6 (shortwave infrared 1)	1.566-1.651	B6 (red edge 2)	0.7315-0.7495
B7 (shortwave infrared 2)	2.08-2.35	B7 (shortwave infrared 2)	2.107-2.294	B7 (red edge 3)	0.7688-0.7968
		B10 (brightness temperature 1)	10.60-11.19	B8 (near infrared)	0.7603-0.9053
		B11 (brightness temperature 2)	11.50-12.51	B8A (red edge 4)	0.8482-0.8812
				B11 (shortwave infrared 1)	1.5422-1.6852
				B12 (shortwave infrared 2)	2.0814-2.3234

**Table 2** Satellite and day range of the extracted Landsat pixels

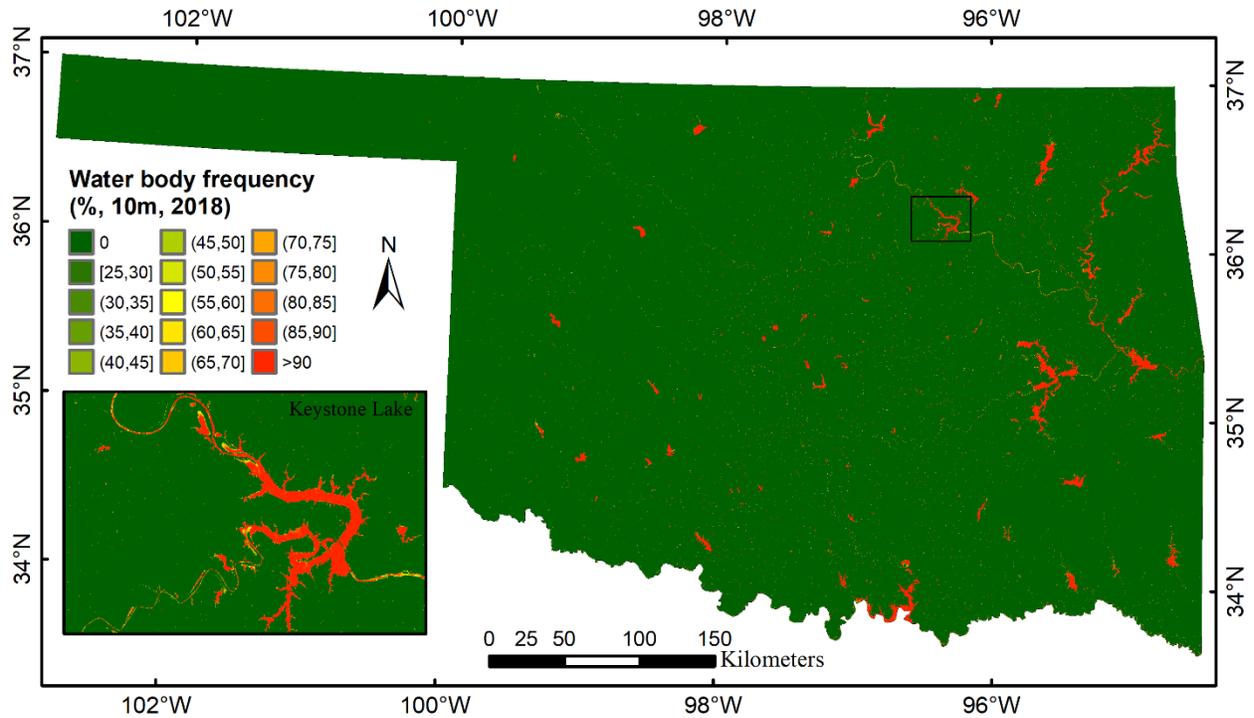
Satellite		Day range between satellite image and chlorophyll-a measurement	
Type	Num.	Range (day)	Num.
L5	4,938	0	816
L7	5,507	[1-5]	5,837
L8	924	[6-10]	2,429
		[11,15]	1,107
		[16,20]	614
		[21, 25]	233
		[25, 30]	143
		[31,94]	190
Sum	11,369	Sum	11,369

Chlorophyll-a concentration associated with algal blooms changes across the seasons, affected by nutrients, temperature, sunlight, water chemistry, etc. Thus, satellite pixels acquired closer with the chlorophyll-a field measurement date could better represent the actual chlorophyll-a concentration. This study classified the chlorophyll-a measurements with a corresponding satellite pixel within 10 days as measurements qualified for further analysis. Out of the 908 water sampling sites, 768 have at least one qualified measurement. There were 420 water sampling sites that have at least 10 qualified measurements and only these sites were included into the stepwise multiple regression analysis in MATLAB R2014a.

Considering there were only 34 chlorophyll-a field measurements with corresponding satellite pixels within 10 days. We included these 34 measurements, from different water sampling sites across Oklahoma, into the multiple regression models with all the reflectance bands of Sentinel 2 pixels as potential factors (Table 1).

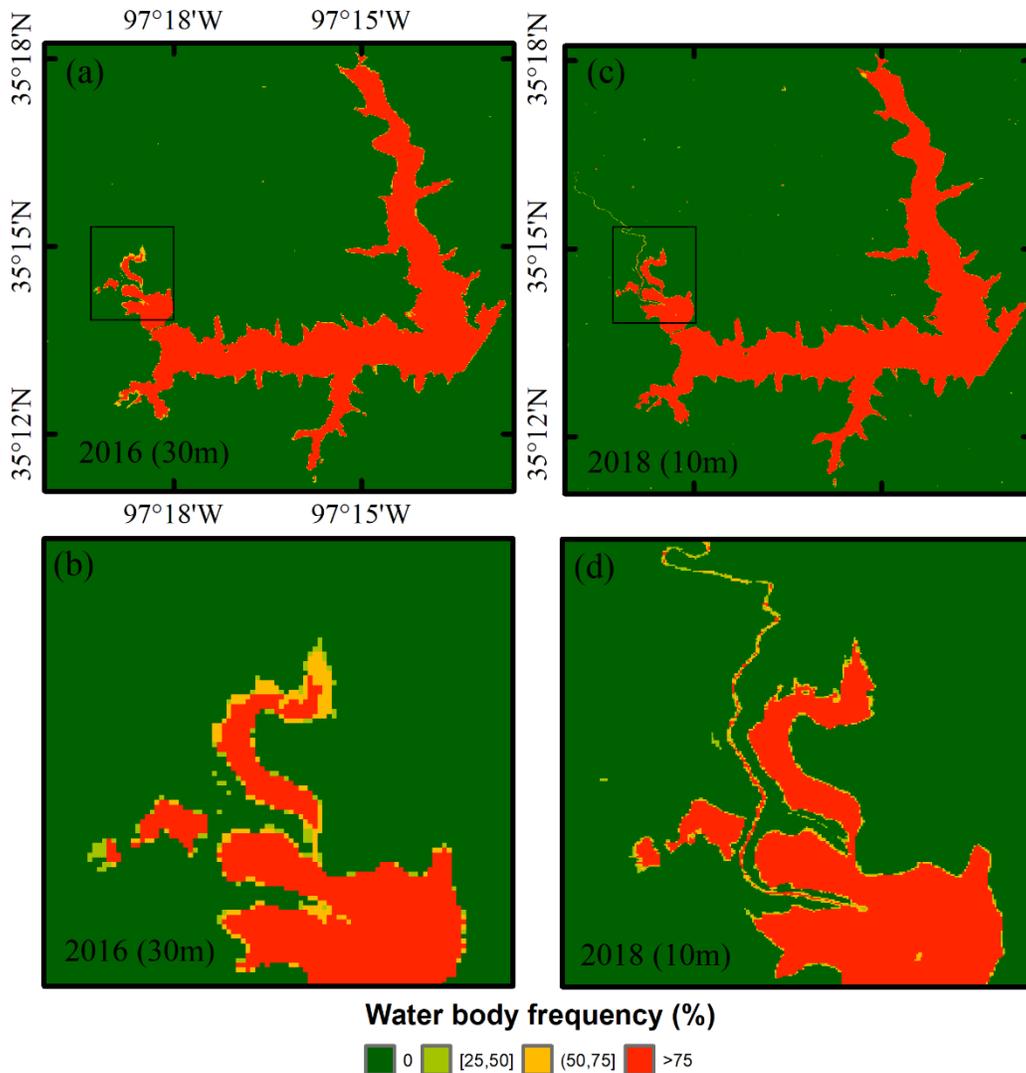
### **Principle Findings and Significance:**

We generated a statewide water body frequency map of 2018 at the spatial resolution of 10m (Fig. 1). Our annual water body frequency map provides the location and extent of surface water bodies and illustrates the stability of surface water resources. High frequency values represent consistent water bodies, while low frequency values represent discontinuous inundation, such as seasonal water bodies. According to this annual frequency map, there were ~ 820 km<sup>2</sup> seasonal water body area (25%≤water frequency<75%) and ~2670 km<sup>2</sup> year-long water body area (water frequency≥75%).



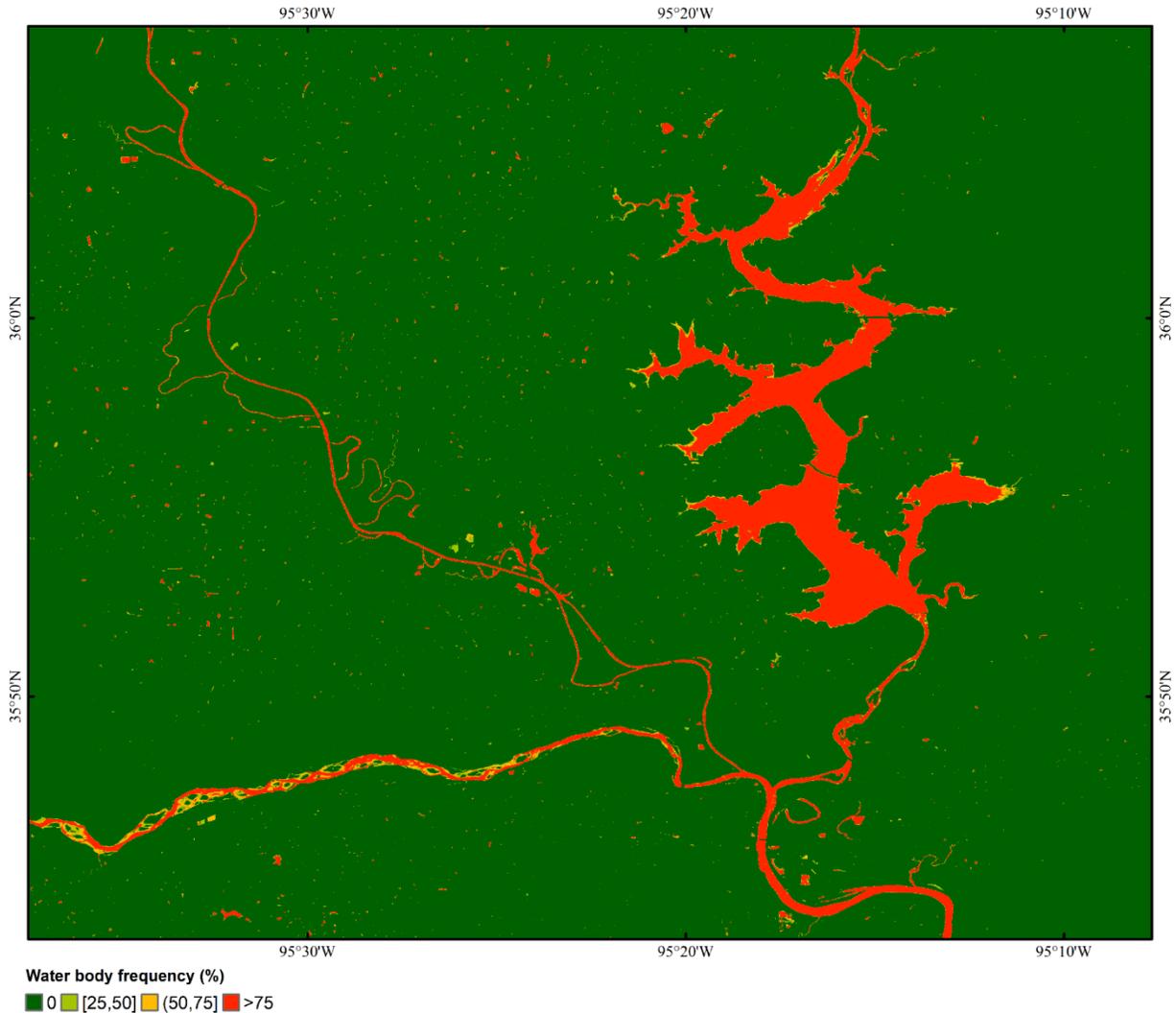
**Fig. 1.** Statewide annual water body frequency map of 2018 at the spatial resolution of 10m.

The 10-m resolution map of 2018 from this study can provide clearer water body boundaries while capturing small streams that were omitted in the 30-m map (Fig. 2). For example, the attribute river of Lake Thunderbird was not captured by the 30-m water body frequency map but clearly shown in the 10-m map we generated.



**Fig. 2.** Annual water body frequency maps of Lake Thunderbird. 30-m map based on Landsat 7 and 8 images in 2016 (a) and its zoom-in (b). 10-m map based on sentinel 1 and 2 images (c) and its zoom-in (d).

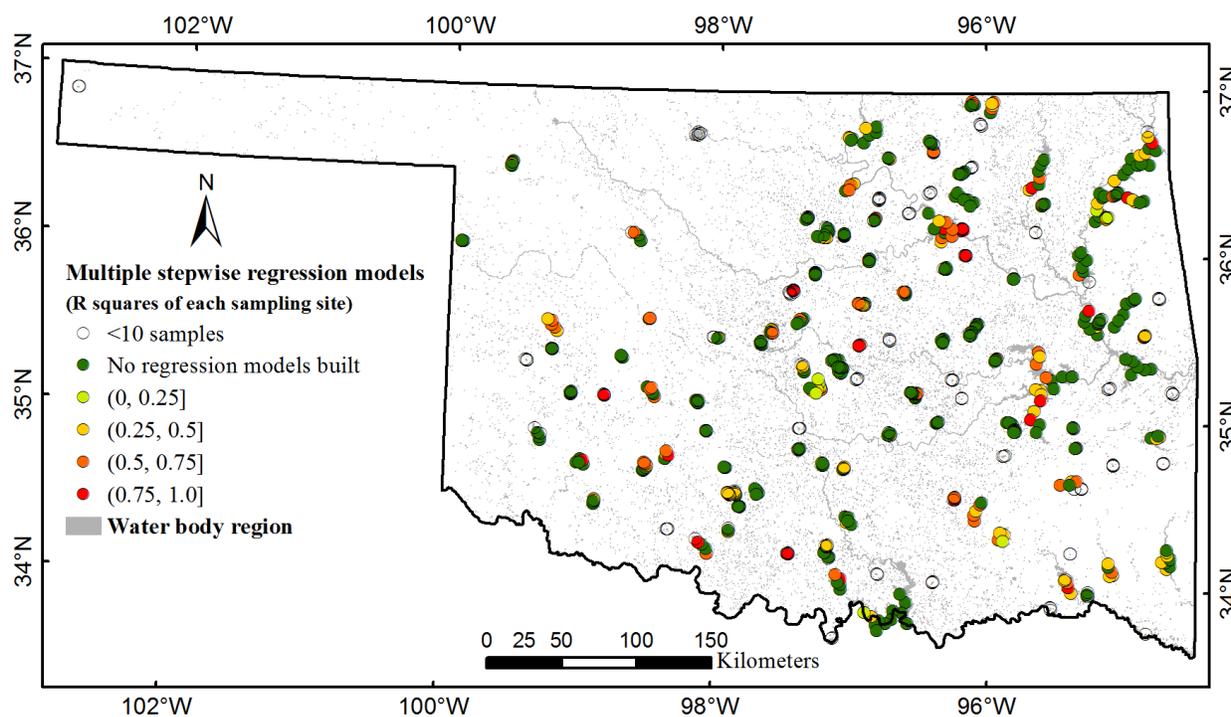
The 10-m annual water body frequency map can clearly show the river channels of Arkansas River, Verdigris River, and Neosho River (Fig. 3). It can also indicate the stability of streamflow. For example, in the upper reaches of Arkansas River, water only covers parts of the river channel. Parts of the river channels are only seasonally flooded.



**Fig. 3.** Annual water body frequency maps of Fort Gibson Lake, Arkansas River (lower part of figure), Verdigris River (upper part of figure), and Neosho River (connected to Fort Gibson Lake).

Among the 420 water sampling sites qualified for stepwise multiple regression analysis, only 165 successfully built the regression models. Nine (5%) regression models had R squares  $<0.25$ , 77 (47%) had R squares between 0.25 and 0.5, 53 (32%) had R squares between 0.5 and 0.75, and 26 (16%) had R squares  $>0.75$  (Fig. 4, Table S1). The performance of regression models varies substantially across Oklahoma, with relatively good performance in Eufaula Lake, Keystone Lake, Copan Lake, Hugo Lake, Foss Reservoir, and Atoka Reservoir. The performance even varies across different water sampling sites within a lake. For example, model performance in northern Grand Lake and southern Oologah Lake was much better than the other portions of these two lakes.

For different water sampling sites, the significant influencing bands selected by the multiple stepwise regression models were also different (Table S1). Among the 165 multiple regression models, blue band was selected by 24 models, green band 31 models, red band 36 models, Near infrared band 34 models, Shortwave infrared-1 band 38 models, brightness temperature band 51 models, and shortwave infrared-2 band 21 models. These phenomena were likely caused by the difference in nutrients, temperature, suspended matters, and water chemistry (Matsushita et al., 2015) among different water bodies across Oklahoma. However, the specific influencing factors of chlorophyll-a concentration in each water body of Oklahoma remains unknown. Further studies will be required to solve this problem. It is worth to mention that brightness temperature band was included in about one third of all the regression models, more than any other bands, indicating that water temperature is one of the most important factors of algal bloom in Oklahoma and that this band has great potential in water chlorophyll-a estimation.



**Fig.4** R squares of multiple stepwise regression models. 768 water sampling sites across the entire Oklahoma.

A regression model was built using chlorophyll-a measurement and sentinel 2 data (Equation 1). Only the reflectance of Red Edge-2 band (Table 1) was selected by stepwise regression model. The R square of the regression model is 0.313, standard error of the estimate is 5.85, and the F value is 14.61 (P=0.001).

$$\text{Chl. A} = 372.392 \times \rho_{\text{Red Edge 2}} - 3.617 \quad (\text{Equation 1})$$

Where Chl. A is Chlorophyll-a concentration (mg/m<sup>3</sup>),  $\rho_{\text{Red Edge 2}}$  is surface reflectance of Red Edge 2 band (0.7315-0.7495  $\mu\text{m}$ ).

The model performance is not very good ( $R^2 = 0.313$ ), which is likely affected by the various water optical properties among the different water sampling sites and different lakes across Oklahoma. On the other hand, the Red Edge-2 Band showed significant linear relationships with chlorophyll-a concentration across different water bodies, indicating its great potential in chlorophyll-a content estimation.

### References:

- Carlson, R.E. (1977), A trophic state index for lakes, *Limnology and oceanography*, 22(2), 361-369.
- Feng, M., J.O. Sexton, S. Channan, and J.R. Townshend (2016), A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic-spectral classification algorithm, *Int. J. Digit. Earth*, 9(2), 113-133, <http://dx.doi.org/10.1080/17538947.2015.1026420>.
- Homer, C., J. Dewitz, L.M. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown (2015), Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information, *Photogramm. Eng. Rem. S.*, 81(5), 345-354, <http://dx.doi.org/10.14358/Pers.81.5.345>.
- Matsushita, B., W. Yang, G.L. Yu, Y. Oyama, K. Yoshimura, and T. Fukushima (2015), A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters, *ISPRS J. Photogramm.*, 102, 28-37, <http://dx.doi.org/10.1016/j.isprsjprs.2014.12.022>.
- OWRB (2015), 2015 Oklahoma Lakes Report Beneficial Use Monitoring Program, pp. 1-216, Oklahoma Water Resources Board. [Available at [https://www.owrb.ok.gov/quality/monitoring/bump/pdf\\_bump/Reports/BUMP%20Lakes%20Report%202015.pdf](https://www.owrb.ok.gov/quality/monitoring/bump/pdf_bump/Reports/BUMP%20Lakes%20Report%202015.pdf).]
- Pekel, J.F., A. Cottam, N. Gorelick, and A.S. Belward (2016), High-resolution mapping of global surface water and its long-term changes, *Nature*, 540(7633), 418-422, <http://dx.doi.org/10.1038/nature20584>.
- USGS (2010), Water use in the United States, U.S. Geological Survey, Reston, Virginia. [Available at <http://water.usgs.gov/watuse/data/>. Accessed September 10, 2016.]
- Xiao, X.M., S. Boles, S. Frolking, C.S. Li, J.Y. Babu, W. Salas, and B. Moore (2006), Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images, *Remote Sens. Environ.*, 100(1), 95-113, <http://dx.doi.org/10.1016/j.rse.2005.10.004>.

- Zou, Z., J. Dong, M.A. Menarguez, X. Xiao, Y. Qin, R.B. Doughty, K.V. Hooker, and K. David Hambright (2017), Continued decrease of open surface water body area in Oklahoma during 1984-2015, *Sci. Total Environ.*, 595, 451-460, <http://dx.doi.org/10.1016/j.scitotenv.2017.03.259>.
- Zou, Z., X. Xiao, J. Dong, Y. Qin, R. B. Doughty, M. A. Menarguez, G. Zhang, J. Wang. (2018), Divergent trends of open surface water body area in the contiguous US during 1984-2016, *Proceedings of the National Academy of Sciences*, 115, 3810-3815, <https://doi.org/10.1073/pnas.1719275115>.

## Supplementary materials

**Table S1** Multiple stepwise regression models of 165 sampling sites

ID	Num.	R <sup>2</sup>	P	Inter.	Blue	Green	Red	NearIn	ShortIn1	TemC	ShortIn2
1	51	0.08	0.040	30.54	-298.93						
2	81	0.09	0.007	4.64						0.94	
3	85	0.22	0.000	10.66		-187.55				1.12	
4	20	0.22	0.039	2.82					60.61		
5	19	0.22	0.042	10.25					436.35		
6	35	0.23	0.004	-1.04			469.85				
7	46	0.23	0.003	5.78					-216.88	1.58	
8	20	0.24	0.030	3.00					99.93		
9	17	0.24	0.048	36.92							-840.22
10	16	0.25	0.048	25.06			-267.89				
11	16	0.25	0.048	7.83						-0.04	
12	17	0.25	0.039	8.45							663.43
13	18	0.25	0.033	12.84					614.41		
14	16	0.26	0.045	37.72	-507.59						
15	17	0.26	0.037	2.01					75.60		
16	19	0.26	0.024	10.46						-0.04	
17	17	0.26	0.035	0.87						0.15	
18	19	0.27	0.024	59.37			-480.68				
19	16	0.27	0.038	23.88			-162.89				
20	15	0.27	0.045	14.82	-172.77						
21	84	0.28	0.000	54.39	-751.93	-347.60		735.15			
22	17	0.28	0.030	3.95						0.01	
23	86	0.28	0.000	45.53		-501.63			457.89		
24	16	0.29	0.030	15.06			-80.80				
25	16	0.30	0.028	73.52			-575.11				
26	17	0.30	0.022	15.00	-169.46						
27	14	0.31	0.038	6.18					500.79		
28	14	0.32	0.036	21.38						-0.06	
29	13	0.33	0.039	3.56				113.52			

30	32	0.34	0.003	33.92	-428.91	315.67
31	16	0.34	0.018	5.81		-0.04
32	12	0.35	0.044	24.14	-223.72	
33	18	0.35	0.010	35.48	-414.20	
34	15	0.36	0.019	66.68		-862.70
35	16	0.36	0.014	7.59		516.27
36	34	0.36	0.001	27.85	-369.86	292.04
37	17	0.37	0.010	26.73		-0.19
38	13	0.37	0.028	26.32		-382.82
39	18	0.37	0.007	-1.40		346.81
40	16	0.37	0.012	63.72	-940.70	
41	19	0.37	0.005	0.58	70.59	
42	13	0.38	0.026	-0.06		0.65
43	13	0.38	0.026	13.88		-0.07
44	14	0.38	0.018	12.06		549.30
45	17	0.39	0.007	0.05		0.41
46	11	0.39	0.039	63.44	-1007.45	
47	20	0.39	0.014	7.52	-91.91	153.47
48	12	0.39	0.029	10.36		-0.06
49	11	0.39	0.038	0.14		1.08
50	15	0.40	0.012	30.08	-283.59	
51	18	0.40	0.005	2.68		-0.02
52	13	0.40	0.020	10.43		264.93
53	13	0.40	0.020	38.47	-319.28	
54	13	0.40	0.020	59.64	-616.21	
55	14	0.40	0.015	1.35		168.50
56	19	0.41	0.003	-0.35		257.45
57	27	0.41	0.002	7.70	-129.10	237.60
58	11	0.41	0.034	10.52		-101.13
59	14	0.41	0.013	1.77		0.78
60	11	0.41	0.033	5.60		83.69
61	18	0.42	0.004	45.41	-438.98	

62	12	0.42	0.022	34.93	-429.43		
63	17	0.43	0.004	15.57		-91.88	
64	15	0.43	0.008	5.64			1.75
65	19	0.43	0.011	26.41	-433.58	276.36	
66	13	0.44	0.014	62.01	-475.79		
67	14	0.44	0.009	-1.18			1.03
68	32	0.45	0.000	51.98	-645.54	373.54	
69	14	0.45	0.009	38.34	-376.47		
70	17	0.45	0.003	29.38			-0.20
71	11	0.46	0.022	22.88	-111.25		
72	11	0.47	0.021	44.11	-492.73		
73	15	0.47	0.005	14.85			1.94
74	16	0.48	0.015	39.81	-470.10	273.97	
75	13	0.48	0.009	7.96			460.53
76	17	0.48	0.002	22.16	-132.73		
77	35	0.48	0.000	40.14	-477.09	275.59	
78	15	0.49	0.004	27.47			-0.21
79	17	0.49	0.002	35.03	-295.55		
80	14	0.49	0.005	19.17			0.06
81	18	0.49	0.006	28.96	-353.92	240.10	
82	19	0.49	0.004	9.22	-118.88		201.10
83	18	0.49	0.006	49.72	-602.81	455.25	
84	19	0.50	0.001	50.33	-451.64		
85	18	0.50	0.006	12.96	-122.83		319.30
86	12	0.50	0.010	8.67			402.16
87	14	0.50	0.004	11.74			508.30
88	10	0.51	0.021	2.19			98.66
89	17	0.51	0.001	1.68			0.77
90	18	0.51	0.004	6.51	-183.11		339.58
91	13	0.52	0.006	3.74			855.63
92	22	0.52	0.001	56.72	-2108.12	1734.11	
93	11	0.53	0.012	1.35			85.30



125	14	0.65	0.000	28.07				-417.35		
126	13	0.66	0.005	27.46			-416.33	381.38		
127	12	0.66	0.001	11.70						-0.10
128	12	0.68	0.001	21.35						-0.09
129	10	0.69	0.003	-0.90						2.32
130	11	0.70	0.001	2.55					199.74	
131	15	0.70	0.000	0.19					517.15	
132	10	0.71	0.002	3.52						-0.02
133	12	0.71	0.004	15.18			-218.23	283.59		
134	16	0.71	0.000	15.30			-119.76		204.61	
135	16	0.71	0.001	8.41					-	1036.41 0.31
136	17	0.72	0.001	16.33			-270.48	482.31		-311.86
137	16	0.72	0.000	-15.51	355.94					-238.26
138	12	0.73	0.003	-0.98			64.52			-0.01
139	11	0.74	0.001	15.06			-70.68			
140	13	0.77	0.000	30.99					-520.84	
141	12	0.77	0.001	11.99			-88.00		324.64	
142	10	0.77	0.001	4.64						-0.03
143	12	0.77	0.001	16.84			-158.46			1000.63
144	10	0.77	0.001	-12.16					482.98	
145	13	0.78	0.000	-2.05						0.83
146	16	0.79	0.000	61.78			-792.72		-481.28	
147	12	0.79	0.004	9.49			-85.01		659.14	-542.67
148	11	0.79	0.000	3.62					252.65	
149	11	0.80	0.002	2.45					111.11	-0.02
150	23	0.80	0.000	17.67						-0.26 -712.63
151	17	0.81	0.000	15.16			-96.30			134.64
152	10	0.83	0.000	-11.88					973.36	
153	12	0.85	0.000	1.24					41.95	0.00
154	17	0.85	0.000	7.07			-23.27		1432.00	-520.33
155	14	0.86	0.000	32.43			-403.33		680.73	
156	11	0.86	0.000	10.56	-230.10					0.65

157	10	0.87	0.000	-6.72				0.96
158	12	0.91	0.000	0.73			-236.81	0.95
159	10	0.92	0.001	5.24	-197.75	115.81		0.00
160	13	0.94	0.000	11.40	1136.62	-1198.21	328.67	
161	12	0.95	0.000	18.83	-636.14	416.63		0.30
162	10	0.95	0.000	18.06	-555.81	147.34	188.00	
163	10	0.96	0.000	8.13			338.37	0.51
164	11	0.97	0.000	-	2382.2	42498.97	-	3093.11
								35746.46
165	10	0.97	0.000	54.22	-1349.40		824.83	-657.72

ID is sampling site ID. Num. is short for sample number at each site.  $R^2$  and P are from regression model summary. Inter. is short for intercept. Blue, Green, Red, NearIn, ShortIn1, TemC, and ShortIn2 are Landsat 5/7/8 bands of Blue, Green, Red, Near infrared, Shortwave infrared 1, brightness temperature, and shortwave infrared 2, respectively.