environments

Article

Examining the Relationship between Phytoplankton Community Structure and Water Quality Measurements in Agricultural Waters: A Machine Learning Application

Jaclyn E. Smith [1,2,3], Jennifer L. Wolny [4], Robert L. Hill [2], Matthew D. Stocker [1,2,3] and Yakov Pachepsky [1,*]

[1] Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, ARS-USDA, Beltsville, MD 20705, USA
[2] Department of Environmental Science and Technology, University of Maryland, College Park, MD 20742, USA
[3] Oak Ridge Institute for Science and Education, Oak Ridge, TN 37831, USA
[4] Center for Food Safety and Applied Nutrition, Office of Regulatory Science, FDA, College Park, MD 20740, USA
* Correspondence: yakov.pachepsky@usda.gov

1. Supplementary Material

Citation: Smith, J.E.; Wolny, J.L.; Hill, R.L.; Stocker, M.D.; Pachepsky, Y. Examining the Relationship between Phytoplankton Community Structure and Water Quality Measurements in Agricultural Waters: A Machine Learning Application. *Environments* **2022**, *9*, 142. https://doi.org/10.3390/environments9110142

Academic Editor: Naresh Singhal

Received: 30 September 2022
Accepted: 9 November 2022
Published: 12 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

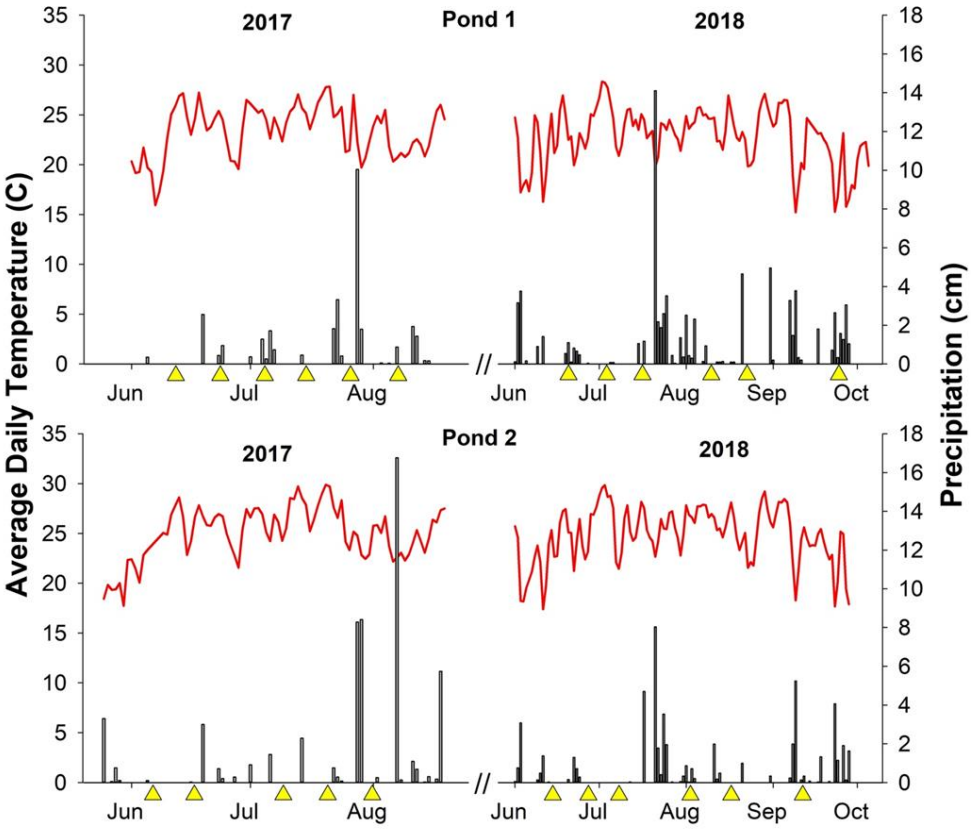

Figure S1. Weather data for Pond 1 and Pond 2 for 2017 and 2018. Average daily temperature readings are represented by the red line and daily precipitation is represented by black bars. Sampling dates are indicated with a yellow triangle at the x axis.

Table S1. Time series data of water quality parameters for Pond 1 and Pond 2 for 2017, 2018, and 2019.

Date		Temp	DO	SPC	pH	NTU	Phyco	CHL	*f*DOM	CDOM
					Pond 1					
	6/7/17	22.7	10.5	156.4	8.5	2.9	0.2	1.9	3.4	n/a
	6/21/17	27.1	9.1	168.9	8.7	4.8	0.3	2.3	2.2	n/a
	7/5/17	28.0	11.9	170.2	9.4	4.1	0.9	6.3	4.8	n/a
	7/18/17	29.1	11.0	175.8	8.9	7.4	2.3	7.6	6.2	n/a
	8/2/17	25.4	10.2	150.7	8.6	6.3	1.2	1.7	36.7	n/a
2017	8/16/17	25.4	9.6	153.1	8.8	5.5	1.9	3.3	25.0	n/a
					Pond 2					
	5/31/17	24.5	16.9	151.5	9.4	4.8	0.9	8.7	32.6	n/a
	6/13/17	30.7	14.5	160.1	8.8	8.4	2.9	15.1	25.0	n/a
	7/11/17	29.8	12.0	174.8	8.2	3.9	1.6	12.3	23.6	n/a
	7/25/17	29.6	10.3	179.1	8.1	32.2	11.7	15.3	28.3	n/a
	8/8/17	22.2	5.4	141.6	6.5	4.8	0.3	2.4	42.2	n/a
					Pond 1					
	6/20/18	26.6	11.6	141.4	8.5	8.8	2.5	2.4	21.4	85.0
	7/5/18	29.6	7.9	156.4	7.3	3.8	0.6	1.1	7.1	43.4
	7/19/18	28.3	9.0	166.9	7.3	2.6	0.3	1.0	1.9	32.3
	8/15/18	26.8	10.5	146.5	7.5	2.5	0.6	4.6	29.9	95.3
	8/29/18	27.7	11.2	132.8	8.1	4.0	0.9	6.9	25.6	102.6
	10/4/18	23.0	10.2	149.0	7.4	3.6	0.8	0.7	38.4	102.2
2018					**Pond 2**					
	6/14/18	25.5	20.7	136.4	8.8	9.3	6.1	36.6	39.6	211.7
	6/26/18	27.1	16.8	146.3	8.5	6.0	4.2	26.3	30.0	166.2
	7/10/18	28.6	11.1	162.1	7.5	10.2	4.5	12.9	27.5	133.3
	8/7/18	31.9	14.1	124.8	8.5	18.7	4.7	17.0	40.8	203.3
	8/23/18	25.9	11.5	139.6	7.4	18.5	5.0	13.7	38.8	188.5
	9/20/18	26.2	15.3	146.4	8.3	20.5	4.7	13.3	32.8	159.0
					Pond 1					
	6/6/19	24.6	9.3	122.5	7.5	1.8	0.1	0.4	8.7	41.5
	7/2/19	28.1	16.8	150.2	9.7	4.1	1.0	7.2	15.9	110.2
2019	8/8/19	27.6	8.8	158.1	8.3	7.2	0.9	5.9	11.2	74.3
					Pond 2					
	6/4/19	24.7	14.0	170.3	8.3	8.9	3.6	17.2	27.7	127.7
	6/26/19	28.1	14.7	170.8	8.2	7.5	5.4	31.7	28.4	128.2

Table S2. Summary statistics for all measured parameters and phytoplankton groups for 2017, 2018, and 2019.

Minimum, maximum, mean, and median values of measured phytoplankton functional groups and water quality data for 2017 and 2018									
Variable	**Units**	**Pond 1**				**Pond 2**			
2017		**Min**	**Max**	**Mean**	**Median**	**Min**	**Max**	**Mean**	**Median**
Diatoms Grouped	Log cells/L	4.19	7.59	5.94	6.00	4.19	7.77	6.32	6.46
Green Algae Grouped	Log cells/L	6.03	8.08	7.10	7.14	4.97	8.18	6.92	7.08
Cyanobacteria Grouped	Log cells/L	4.74	7.91	5.98	6.11	4.67	8.69	6.14	5.81
TEMP	°C	22.33	29.84	26.31	26.34	21.81	33.74	27.37	29.16

DO	mg/L	8.14	15.44	10.39	10.24	2.74	24.87	11.82	12.01
SPC	uS/cm	149.50	178.30	162.52	162.15	138.90	183.20	161.41	160.15
pH		6.90	9.56	8.82	8.84	6.18	9.90	8.18	8.49
NTU		1.81	23.78	5.16	5.05	1.52	68.61	10.82	4.73
Phyco	RFU	0.12	8.43	1.14	0.93	-0.27	34.18	3.49	0.83
CHL	RFU	0.87	46.54	3.83	2.21	0.84	123.31	10.74	5.83
FDOM	ppb	-0.05	39.53	13.06	5.29	9.39	47.63	30.34	29.11
2018									
Diatoms Grouped	Log cells/L	4.45	7.22	5.61	5.75	4.49	7.42	5.63	5.79
Green Algae Grouped	Log cells/L	5.49	7.78	6.70	6.72	4.89	8.08	7.10	7.15
Cyanobacteria Grouped	Log cells/L	4.19	7.95	5.64	4.19	5.29	8.38	7.08	7.25
TEMP	°C	22.34	30.10	26.99	27.06	24.00	36.28	27.54	26.95
DO	mg/L	7.04	12.68	10.07	10.41	5.83	27.18	14.91	14.94
SPC	uS/cm	1.70	168.10	148.83	149.05	10.10	164.10	142.59	145.70
pH		6.47	8.67	7.65	7.46	6.46	9.31	8.17	8.37
NTU		1.20	24.00	4.20	3.28	1.00	178.46	13.86	9.85
Phyco	RFU	0.10	10.65	0.95	0.61	0.45	35.11	4.86	3.32
CHL	RFU	0.51	11.12	2.80	1.79	1.37	166.42	19.97	10.43
FDOM	ppb	0.49	41.25	20.72	23.60	9.09	50.64	34.90	35.58
CDOM	ug/L	29.94	113.15	76.80	91.38	93.54	826.40	177.00	161.97
Submerged Light - 0cm		7.00	1809.00	712.72	793.00	55.00	2128.00	1113.02	1195.50
Submerged Light - 7.5cm		7.00	1578.00	624.42	653.50	7.00	1946.00	810.60	848.50
Submerged Light - 15cm		6.00	1479.00	540.88	546.50	5.00	1933.00	578.13	607.00
PAR		4.00	1278.00	950.46	928.00	48.00	2200.00	1312.86	1373.00
2019									
Diatoms Grouped	Log cells/L	4.66	6.22	5.39	5.29	5.30	6.97	6.19	6.27
Green algae Grouped	Log cells/L	6.11	7.74	7.31	7.52	6.39	7.53	7.04	7.02
Cyanobacteria Grouped	Log cells/L	0.00	7.22	0.74	0.00	6.09	7.85	6.82	6.81
TEMP	°C	23.80	29.60	26.71	27.14	23.46	30.12	26.39	26.96
DO	mg/L	7.70	17.18	11.58	9.31	3.02	20.53	14.32	14.80
SPC	uS/cm	39.30	158.60	143.58	150.40	99.00	178.90	170.55	171.60
pH		7.30	9.70	8.49	8.30	6.83	9.22	8.27	8.45
NTU		1.41	12.49	4.39	3.28	3.23	16.48	8.20	7.75
Phyco	RFU	0.00	1.42	0.69	0.88	0.65	12.41	4.47	3.55
CHL	RFU	0.28	9.96	4.59	5.72	1.23	67.11	24.45	15.38
FDOM	ppb	4.38	20.37	12.00	11.32	22.08	31.85	28.05	28.41
CDOM	ug/L	39.13	121.08	75.62	74.76	98.21	218.89	127.94	113.50
Submerged Light - 0cm		370.00	1773.00	971.82	888.00	359.00	1895.00	1249.25	1158.00
Submerged Light - 7.5cm		269.00	1499.00	834.47	679.00	52.00	1459.00	886.75	792.50
Submerged Light - 15cm		275.00	1417.00	700.70	541.50	16.00	1175.00	629.95	620.50
PAR		518.00	2124.00	1308.65	1185.50	997.00	2073.00	1657.53	1823.50

Table S3. Pearson correlation coefficients between water quality parameters in 2017 and 2018.

Variable pair	*2017-P1*	*2017-P2*	*2018-P1*	*2018-P2*
TEMP-SPC	0.804**	0.803**		
TEMP-FDOM		−0.869**		
DO-pH	0.606*	0.937**	0.767**	0.893**
DO-FDOM			0.590*	
pH-FDOM		0.576*		
SPC-FDOM	−0.738**	−0.710**		
NTU-Phyco	0.682*	0.640*	0.763**	0.537*
NTU-CHL	0.589¶			
Phyco-CHL	0.756**	0.654*		0.819**

Moderate ($0.5 < r < 0.7$, *) and strong ($r > 0.7$, **) coefficients.

Table S4. Spatial root-mean-squared errors (RMSEs) calculated by sampling location for 2017+2018 calculated for green algae, diatoms, and cyanobacteria using set A and AB parameters in Pond 1 and Pond 2.

Location	Input set A			Input set AB		
	Green Algae	Diatoms	Cyanobacteria	Green Algae	Diatoms	Cyanobacteria
Pond 1						
1	0.292	0.409	0.647	0.271	0.397	0.636
2	0.286	0.532	0.656	0.229	0.533	0.612
3	0.231	0.456	0.657	0.244	0.455	0.699
4	0.230	0.403	0.596	0.243	0.438	0.616
5	0.295	0.537	0.625	0.274	0.559	0.664
6	0.241	0.407	0.538	0.228	0.440	0.531
7	0.275	0.529	0.605	0.256	0.513	0.583
8	0.311	0.527	0.579	0.280	0.494	0.621
9	0.299	0.451	0.631	0.287	0.455	0.632
10	0.267	0.350	0.599	0.268	0.361	0.607
11	0.194	0.421	0.411	0.221	0.464	0.480
12	0.194	0.382	0.438	0.210	0.429	0.472
13	0.229	0.411	0.611	0.225	0.411	0.668
14	0.203	0.353	0.599	0.194	0.370	0.563
15	0.261	0.450	0.523	0.259	0.443	0.515
16	0.449	0.504	0.500	0.449	0.507	0.518
17	0.355	0.355	0.570	0.328	0.334	0.579
18	0.262	0.484	0.510	0.252	0.477	0.537
19	0.280	0.463	0.583	0.253	0.437	0.597
20	0.234	0.427	0.507	0.212	0.443	0.538
21	0.251	0.391	0.582	0.230	0.399	0.544
22	0.323	0.521	0.520	0.267	0.532	0.551
23	0.427	0.490	0.551	0.376	0.503	0.551
Pond 2						
1	0.414	0.588	0.686	0.394	0.606	0.668
3	0.297	0.550	0.669	0.333	0.609	0.665
5	0.345	0.620	0.656	0.373	0.621	0.743
7	0.435	0.644	0.657	0.452	0.650	0.665
9	0.494	0.698	0.799	0.480	0.679	0.761
11	0.419	0.707	0.892	0.432	0.652	0.798
13	0.463	0.764	0.766	0.514	0.784	0.761
15	0.468	0.537	0.843	0.476	0.528	0.833
17	0.505	0.477	0.794	0.522	0.478	0.771
19	0.351	0.499	0.732	0.350	0.563	0.697
21	0.395	0.601	0.716	0.381	0.577	0.688
23	0.304	0.581	0.561	0.313	0.571	0.606
25	0.288	0.668	0.644	0.282	0.639	0.651
26	0.312	0.546	0.686	0.324	0.587	0.768
27	0.220	0.500	0.586	0.262	0.469	0.570
28	0.289	0.551	0.642	0.300	0.548	0.592
29	0.374	0.590	0.574	0.361	0.610	0.562
30	0.242	0.568	0.620	0.261	0.583	0.618
31	0.272	0.425	0.628	0.310	0.516	0.617
32	0.268	0.445	0.634	0.268	0.496	0.549
33	0.298	0.592	0.531	0.318	0.540	0.522
34	0.381	0.571	0.728	0.415	0.568	0.719

Table S5. Percent of dates in which the coefficient of variation (CV) was larger for nearshore locations.

Parameter	Pond 1 %	Pond 2 %
Green algae	41.7	63.6

Diatoms	66.7	63.6
Cyanobacteria	66.7	54.6
TEMP	75.0	81.8
DO	100.0	90.9
SPC	75.0	63.6
pH	100.0	100.0
NTU	91.7	81.8
Phyco	91.7	72.7
CHL	75.0	81.8
FDOM	75.0	90.9