

# MODEL Klorofila-*a* ZA AKUMULACIJU BUTONIGA U ISTRI

**izv. prof. dr. sc. Goran Volf**

Sveučilište u Rijeci,  
Građevinski fakultet  
Radmile Matejčić 3, Rijeka, Hrvatska  
[goran.volf@uniri.hr](mailto:goran.volf@uniri.hr)

**doc. dr. sc. Petar Žutinić**

Sveučilište u Zagrebu,  
Prirodoslovno-matematički fakultet  
Horvatovac 102a, Zagreb, Hrvatska

**izv. prof. dr. sc. Marija Gligora Udovič**

Sveučilište u Zagrebu,  
Prirodoslovno-matematički fakultet  
Horvatovac 102a, Zagreb, Hrvatska

**dr. sc. Antonija Kulaš**

Sveučilište u Zagrebu,  
Prirodoslovno-matematički fakultet  
Horvatovac 102a, Zagreb, Hrvatska

**prof. dr. sc. Perica Mustafić**

Sveučilište u Zagrebu,  
Prirodoslovno-matematički fakultet  
Horvatovac 102a, Zagreb, Hrvatska

Butoniga pripada relativno malim i plitkim akumulacijama te je kao takva jako osjetljiva na procese degradacije i eutrofikacije uzrokovane kako klimatskim promjenama tako i antropogenim aktivnostima u svojem slivu. Klorofil-*a* kao široko primjenjiv ekološki pokazatelj primarne proizvodnje, odnosno biomase fitoplanktona, i procesa eutrofikacije u akumulacijama, ovisi o dostupnosti hranjivih tvari, pH vrijednosti, svjetlosti, temperaturi, kao i njihovim međudodnosima. Tehnika strojnog učenja za izradu modela u obliku regresijskih stabala i pravila primijenjena je na skup mjernih podataka u svrhu boljeg razumijevanja procesa eutrofikacije i ekološkog stanja akumulacije Butoniga. Upotrebom alata za strojno učenje izrađena su dva modela, (1) opisni model klorofila-*a* u obliku regresijskih stabala kojim se opisuju promjene koncentracije klorofila-*a* te utvrđuju najznačajniji čimbenici tih promjena, te (2) simulacijski model klorofila-*a* u obliku pravila koji služi za predviđanje koncentracije klorofila-*a* s obzirom na promatrane vrijednosti izmjerenih okolišnih čimbenika. Izrađeni modeli doprinose boljem razumijevanju ekosustava akumulacije Butoniga, a također se mogu koristiti u svrhu upravljanja akumulacijom.

**Ključne riječi:** akumulacija Butoniga, klorofil-*a*, strojno učenje, modeli, fitoplankton

## 1. UVOD

Koncentracija klorofila-*a*, kao pokazatelj brojnosti i biomase fitoplanktona u vodenim sustavima, ovisi o fizičkim i kemijskim svojstvima vodenog stupca koji uključuju niz čimbenika poput dostupnosti hranjivih tvari, svjetlosti, pH i temperature, kao i odnosima između navedenih čimbenika (Bernát i sur. 2020). Povećana koncentracija klorofila-*a* u jezerskim ekosustavima predstavlja neposredan biološki odgovor na procese degradacije, povećane primarne produkcije te eutrofikacije (Latif i sur. 2003). Samim time klorofil-*a* se može koristiti kao ključni čimbenik prilikom utvrđivanja vanjskih utjecaja na ekosustav jezera i akumulacija, posebice opterećenja hranjivim tvarima. Stoga, pouzdano predviđanje koncentracije klorofila-*a* omogućuje rano otkrivanje potencijalnog cvjetanja algi i usmjerene mjere ublažavanja i kontrole eutrofikacije,

kao i sveukupno upravljanje vodenim ekosustavima (Wang i sur. 2013).

Iako je veza između klorofila-*a* i hranjivih tvari jasna i neposredna, trenutna i ponekad ograničena količina dostupnih podataka onemogućava donošenje odgovarajućih upravljačkih odluka, a pouzdanost bilo koje prognoze uvjetovana je različitim utjecajima na klorofil-*a* koji su još uvijek relativno nepoznati zbog zamršenih odnosa unutar jezerskih ekosustava (Wang i sur. 2013).

Tehnike strojnog učenja pružaju moćan skup alata izrađenih za rješavanje najkompleksnijih problema u ekologiji, posebice problema povezanih s vodenim ekosustavima. Iako možda nikada neće dati potpuno objašnjenje i ponuditi odgovore na sva pitanja koja se postavljaju pred okolišne podatke, pristupi strojnog učenja se zbog brzog razvoja uspješno i intenzivno primjenjuju u različitim znanstvenim istraživanjima te

okolišnim i monitoring studijama (Flach 2001; Crisci i sur. 2012). To se posebno očituje u velikom broju publikacija koje su se pojavile posljednjih godina, a u kojima se primjenjuju različite tehnike strojnog učenja za rješavanje najrazličitijih problema koji uključuju modeliranje brojnosti fitoplanktona i zooplanktona te opisi predviđanje cvjetanja algi, kako u slatkovodnim tako i u morskim ekosustavima. Neke primjene ovih tehnika u vodnim ekosustavima uključuju opisne i modele predviđanja fitoplanktona primjenom regresijskih stabala i modela u obliku pravila (Volf i sur. 2011), opis i predviđanje cvjetanja algi upotrebom regresijskih stabala (Volf i sur. 2015), detekciju morske trave usporedbom algoritama strojnog učenja (Effrosynidis i sur. 2018), modeliranje dinamike fitoplanktona korištenjem metode Lagrange (Atanasova i sur. 2008), modeliranje dinamike zooplanktona (Gal i sur. 2013) te modeliranje učinaka višestrukih stresora na disanje i mikrobnu biomasu korištenjem odgovarajućih pravila u obliku stabala (Mori i sur. 2019).

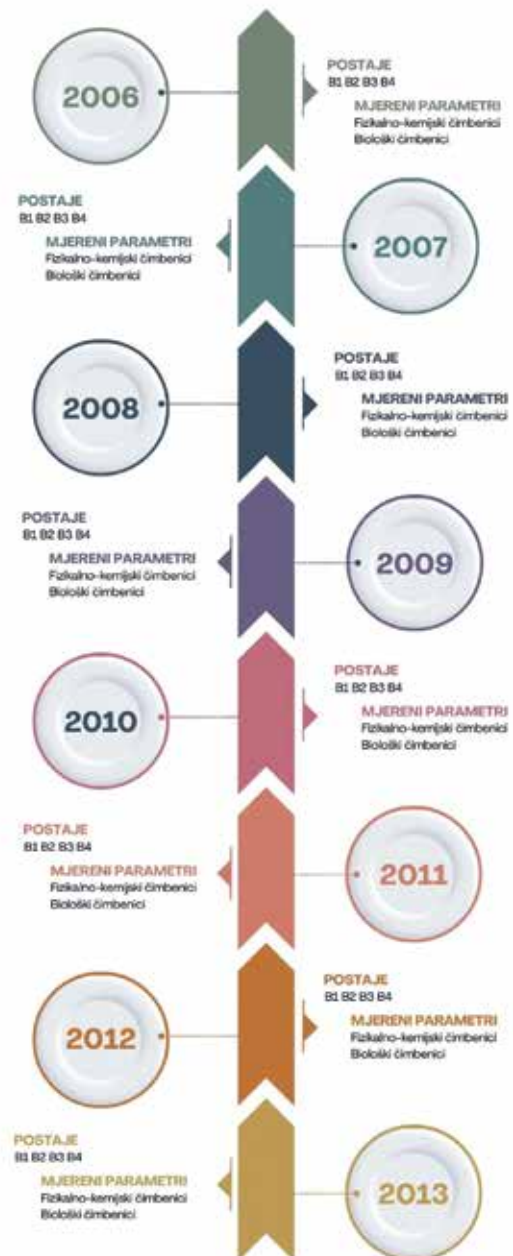
Ciljevi istraživanja u ovome radu su izgraditi modele za: (1) opisni model klorofila-*a* u obliku regresijskih stabala kojim se nastoji objasniti mijenjanje koncentracija klorofila-*a* u akumulaciji Butoniga te odrediti najvažnije čimbenike tih promjena, te (2) simulacijski model klorofila-*a* u obliku pravila koji služi za predviđanje koncentracije klorofila-*a* s obzirom na promatrane vrijednosti izmjerenih okolišnih čimbenika i koji se može koristiti u svrhu upravljanja akumulacijom Butoniga. U svrhu postizanja navedenih ciljeva primijenjene su metode strojnog učenja temeljene na (1) regresiji u obliku stabala odluka i (2) modelima zasnovanim na pravilima za izgradnju iz skupa podataka prikupljenih s akumulacije Butoniga.

## 2. AKUMULACIJA BUTONIGA

Akumulacija Butoniga (slika 1) izgrađena je 1987. godine s dva osnovna cilja: (1) zaštita od štetnog djelovanja voda, te (2) vodoopskrba pitkom vodom. Površina sliva akumulacije iznosi otprilike 73 km<sup>2</sup>,



Slika 1: Akumulacija Butoniga s pozicijama mjernih postaja (B1, B2, B3, B4).



Slika 2: Slikovni prikaz ulaznih podataka (mjerjenja) po postajama prema vremenskim serijama.

s razlikama u nadmorskim visinama od 40 do 500 m.n.v. Volumen akumulacije iznosi 19,5 milijuna m<sup>3</sup>, s vodenom površinom od oko 2,5 km<sup>2</sup>. Srednja dubina akumulacije iznosi 7,5 m dok je maksimalna 17,5 m. Tlo u slivu akumulacije čini pretežno fliš u kojem se izmjenjuju primjese pješčenjaka, lapora, breča i konglomerata, a koji su skloni eroziji i pogodni su za apsorpciju raznih zagađivala zbog čega dolazi do njihovog taloženja u zaplavnom prostoru akumulacije. Kao mala i relativno plitka, akumulacija je jako osjetljiva na procese eutrofikacije i degradacije prouzročene klimatskim promjenama i ljudskim aktivnostima u slivu. Poznati i uobičajeni pritisci u slivu uključuju eroziju i ispiranje hranjivih tvari s poljoprivrednih zemljišta, kao i nepročišćene otpadne vode iz naselja, koja se najčešće

kroz crne jame ili kanalizacijsku mrežu dreniraju u akumulaciju (Hajduk Černeha 2017).

Kvaliteta vode u akumulaciji Butoniga prvenstveno ovisi o godišnjem dobu te količini vode u samoj akumulaciji. Akumulacija je termički stratificirana od travnja do listopada što značajno utječe na fizikalne i kemijske karakteristike vode u akumulaciji. Najveći problem kvalitete vode u akumulaciji povezan je s povećanim koncentracijama hranjivih tvari, prvenstveno fosfora koji ima pozitivan učinak na biološku proizvodnju te na taj način pospješuje procese eutrofikacije. Hranjive tvari najvećim dijelom dotječu u akumulaciju iz slivnog područja te dijelom putem remobilizacije iz sedimenta zbog reduciranih uvjeta u donjim slojevima akumulacije (Hajduk Černeha 2017).

### 3. OPIS PODATAKA

Set podataka sastoji se od fizikalnih, kemijskih i bioloških čimbenika mjerenih jednom mjesečno od svibnja do listopada/studenog tijekom vegetacijske sezone u razdoblju od 2006. do 2013. godine. Na slici 2 prikazan je slikovni prikaz ulaznih podataka, odnosno mjerenja po postajama prema vremenskim serijama. Čimbenici su određeni za svakih 5 metara dubine od površine do dna akumulacije na 4 postaje (B1, B2, B3 i B4; slika 1). Fizikalno-kemijski čimbenici koji uključuju temperaturu (Temp), pH i koncentraciju otopljenog kisika (DO), mjereni su s WTW Multi 3630 prijenosnim multimetrom (Xylem Analytics, Njemačka), dok je prozirnost vode (Secchi) procijenjena korištenjem Secchijevega diska. Za kemijsku analizu vode sljedeći čimbenici mjereni su standardnim analitičkim metodama prema važećim hrvatskim normama i smjernicama (HRN ISO 5667-4:2016, HRN EN ISO 5667-3:2018): amonijak ( $\text{NH}_4^+\text{-N}$ ), kalijev permanganat ( $\text{KMnO}_4$ ), ukupni fosfor (TP) i ortofosfat ( $\text{PO}_4^{3-}\text{-P}$ ). Biološki čimbenik klorofil-*a* (Chl-*a*) analiziran je pomoću HACH DR 4000 spektrofotometra (HACH, S.A.D.).

U tablici 1 dan je prikaz osnovne statističke analize [srednja vrijednost (AVG), minimum (MIN), maksimum (MAX) te standardna devijacija (STDEV)] mjerenih parametara po postajama (B1, B2, B3 i B4) za promatrani period od 2006. do 2013. godine.

Za potrebe modeliranja podaci su osrednjeni po dubini, odnosno iznad eufotičke dubine i uzeti kao jedan sloj kako je pojašnjeno u poglavlju 5.

### 4. REGRESIJSKA STABLA I MODELI U OBLIKU PRAVILA

Linearna regresija je metoda koja ima za cilj izraziti zavisnu varijablu (ciljana varijabla ili klasa) kao linearnu kombinaciju nezavisnih varijabli (atributi ili deskriptori) iz danih mjerenja (primjera). Primjeri mogu biti predstavljeni u obliku tabele gdje svaki red (primjer) ima oblik ( $a_1, a_2, \dots, a_n, x$ ), gdje su  $a_i$  vrijednosti  $N$  atributa (također nezavisne

varijable ili deskriptori), a  $x$  je vrijednost klase. Zadatak jednostavne linearne regresije je izraziti vrijednost klase u obliku:

$$x = a_1 * w_1 + a_2 * w_2 + \dots + a_n * w_n = \sum a_i * w_i$$

gdje su  $w_i$  težine koje su naučene (izračunate) iz seta za treniranje (učenje) modela.

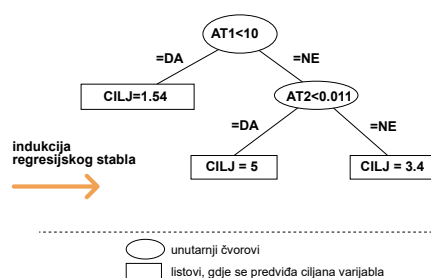
Dok jednostavna linearna regresija računa jednu jednadžbu za cijeli skup podataka, regresija strukturirana u obliku stabla dijeli skup podataka na nekoliko podskupova na kojima se može primijeniti jedinstvena vrijednost klase odnosno linearna jednadžba. Podjela na podskupove temelji se na testovima/provjerama vrijednosti ulaznih atributa koji su stavljani kao čvorovi u regresijskom stablu. Dakle, regresijska stabla su hijerarhijske strukture sastavljene od čvorova i grana, gdje unutarnji čvorovi vrše provjere na ulaznim atributima. Svaka grana unutarnje provjere odgovara ishodu te provjere, dok su predviđanja za vrijednost ciljane varijable (klase) pohranjena u listovima koji označavaju krajnje čvorove stabla. Ako list sadrži jedinstvenu vrijednost za predviđanje klase tada se govori o regresijskom stablu, a ako je u listu linearna jednadžba tada se govori o modelskom stablu (Quinlan 1992; Witten i Frank 2000). Na slici 3 prikazan je primjer procedure koja se koristi prilikom izrade regresijskih stabala.

Jedan od najčešće korištenih algoritama za konstrukciju regresijskih stabala je M5 algoritam (Quinlan 1992), temeljen na algoritmu za izradu stabala za odlučivanje odozgo prema dolje (engl.: "top-down induction of decision trees-TDIDT"; Quinlan 1986). Za ovo istraživanje korištena je inačica M5 algoritma pod nazivom M5P, koja se nalazi u alatu za strojno učenje Weka (Witten i Frank 2000).

Modeli u obliku pravila za numeričku predikciju spadaju u vrstu modela koji se interpretiraju kao set "IF THEN" pravila, gdje je svako pravilo povezano s multivarijantnim linearnim modelom. Pravilo ukazuje da, kad god pojedini slučaj zadovoljava sve uvjete, linearni model je prikladan za predviđanje vrijednosti ciljnog atributa. Algoritmi za indukciju pravila uglavnom predstavljaju različite inačice M5 algoritma. Za modeliranje u ovom radu primijenjen je algoritam razvijen u programskom paketu WEKA (Witten i Frank 2000), u kojem je inačica M5 algoritma (M5P)

SET PODATAKA (PRIMJERI)

AT1	AT2	AT3	...	CILJ (KLASA)
3.67	8.500	0.005	...	2.133
4.15	7.207	0.005	...	2.601
5.32	8.357	0.011	...	3.718
7.80	7.929	0.005	...	3.481
8.11	7.096	0.005	...	1.791
9.36	7.804	0.005	...	1.128
10.87	6.018	0.005	...	1.471
11.10	7.400	0.006	...	1.521
10.23	5.457	0.011	...	0.869
8.39	5.486	0.014	...	0.535
7.42	5.486	0.013	...	1.034
4.06	8.307	0.005	...	1.636
...	...	...	...	...



Slika 3: Primjer izrade regresijskog stabla pomoću seta podataka.

	B1			B2			B3			B4		
	AVG	MIN-MAX	STDEV	AVG	MIN-MAX	STDEV	AVG	MIN-MAX	STDEV	AVG	MIN-MAX	STDEV
2006	DO (mgO <sub>2</sub> /L)	9.30	5.32-17.57	3.56	7.26-14.18	2.62	10.85	6.69-19.00	3.79	10.00	1.27-17.00	4.64
	Temp (°C)	20.2	13.9-27.0	3.9	16.4-27.3	4.1	22.2	16.6-27.6	4.1	22.0	16.9-27.2	3.9
	Secchi (m)	2.7	1.2-3.6	0.9	1-3.4	1.0	2.6	0.9-3.4	0.9	3.4	1-5.5	1.6
	pH	7.98	7.44-8.26	0.19	7.62-8.13	0.16	8.04	7.87-8.2	0.10	8.07	7.97-8.20	0.09
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.160	0.048-0.363	0.068	0.135-0.95	0.343	0.174	0.067-0.287	0.077	0.176	0.115-0.251	0.051
	KMnO <sub>4</sub> (mg/L)	3.06	1.58-7.62	1.67	1.26-3.55	0.83	2.78	1.24-4.08	0.97	3.11	1.50-3.95	0.93
	TP (mgP/L)	0.035	0-0.066	0.022	0.009-0.067	0.025	0.046	0.006-0.102	0.038	0.030	0-0.053	0.018
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	0.015	0-0.035	0.008	0-0.017	0.006	0.009	0-0.026	0.008	0.006	0-0.014	0.005
	Chl-a (µg/L)	1.20	0-4.46	1.17	0.55-5.58	1.78	1.63	0-3.9	1.30	1.05	0-3.350	1.13
		8.23	0.05-12.36	3.35	0.20-12.02	3.90	9.09	2.34-11.86	3.79	9.57	1.73-13.24	4.24
2007	DO (mgO <sub>2</sub> /L)	18.4	6.6-26.9	7.0	6.2-26.8	6.8	24.4	12.8-26.8	4.2	24.0	12.5-26.6	4.0
	Temp (°C)	2.5	0.9-4.0	1.0	0.9-2.8	0.7	3.0	0.8-4	1.0	2.7	0.8-3.5	0.9
	Secchi (m)	8.03	7.40-8.25	0.20	7.73-8.20	0.18	8.11	8.05-8.18	0.04	8.18	8.04-8.30	0.07
	pH	0.295	0.01-2.473	0.384	0-0.661	0.215	0.140	0-0.629	0.161	0.267	0.017-1.461	0.404
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	2.27	1.13-3.55	0.59	1.64-3.29	0.57	2.48	1.42-3.71	0.67	2.50	1.6-3.55	0.61
	KMnO <sub>4</sub> (mg/L)	0.033	0.000-0.106	0.031	0.009-0.165	0.053	0.039	0.003-0.084	0.027	0.058	0.000-0.239	0.068
	TP (mgP/L)	0.012	0.001-0.07	0.014	0.007-0.100	0.029	0.009	0.006-0.013	0.002	0.014	0.005-0.039	0.012
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	2.04	0.55-4.46	1.08	0.55-4.46	1.16	1.84	1.11-4.46	0.90	1.78	1.11-3.35	0.74
	Chl-a (µg/L)	8.15	3.53-11.24	2.83	4.54-11.13	2.88	10.33	9.79-10.73	0.48	9.25	8.12-10.57	1.27
		14.6	11.1-19.2	3.1	11.6-19.4	3.3	17.4	15.1-19.1	2.1	16.1	13.5-19.0	2.9
2008	DO (mgO <sub>2</sub> /L)	2.0	2.0-2.0	0.0	1.8-1.8	0.0	1.6	1.6-1.6	0.0	1.2	1.2-1.2	0.0
	Temp (°C)	7.81	7.15-8.23	0.29	7.02-8.25	0.30	7.96	7.86-8.16	0.10	7.95	7.77-8.17	0.14
	Secchi (m)	0.482	0.088-1.379	0.389	0.042-4.008	1.075	0.253	0-0.686	0.176	0.193	0-0.53	0.183
	pH	1.71	0.8-2.38	0.40	0.76-2.52	0.54	1.69	0.8-2.78	0.57	1.90	0.88-2.6	0.45
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.054	0.010-0.167	0.049	0.015-0.078	0.018	0.037	0.009-0.095	0.027	0.036	0.016-0.124	0.031
	KMnO <sub>4</sub> (mg/L)	0.012	0-0.026	0.007	0.006-0.039	0.010	0.018	0.008-0.042	0.011	0.019	0.006-0.044	0.012
	TP (mgP/L)	1.34	0.56-2.79	0.70	0.56-2.79	0.69	1.64	1.12-3.35	0.68	1.65	0.56-3.35	0.83
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	7.04	0.15-11.90	4.00	0.14-10.65	3.71	8.71	4.12-11.10	2.05	8.37	2.01-13.42	3.21
	Chl-a (µg/L)	19.7	9.9-27.0	6.2	10.3-27.0	5.9	21.8	14.4-26.5	3.3	21.6	13.6-26.7	3.7
		3.2	2.3-4.0	0.6	2-4.6	0.8	2.5	2.1-3.6	0.5	2.6	2.1-3.4	0.5
2009	DO (mgO <sub>2</sub> /L)	7.97	7.35-8.38	0.34	7.36-8.36	0.37	8.16	7.80-8.34	0.17	8.04	7.36-8.33	0.29
	Temp (°C)	0.216	0.203-0.157	0.182	0-0	0.000	0.901	0.833-0.277	0.609	0.190	0.211-0.092	0.170
	Secchi (m)	1.92	1.12-2.72	0.45	1.19-2.46	0.31	1.96	1.67-2.46	0.25	2.24	1.51-2.92	0.50
	pH	0.056	0.005-0.218	0.061	0.007-0.078	0.021	0.034	0.003-0.160	0.044	0.023	0.007-0.053	0.015
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.009	0.001-0.031	0.007	0.002-0.035	0.010	0.009	0.003-0.031	0.008	0.010	0.002-0.029	0.008
	KMnO <sub>4</sub> (mg/L)	1.14	0.55-2.79	0.72	0.55-2.79	0.58	1.97	1.11-3.35	0.81	1.51	0.55-2.79	0.71
	TP (mgP/L)											
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)											
	Chl-a (µg/L)											

Tablica 1: Prikaz osnovne statističke analize mjerenih parametara po postajama (B1, B2, B3 i B4) za promatrani period od 2006-2013 godine.

2010	DO (mgO <sub>2</sub> /L)	8.06	0.23-13.26	3.75	885	0.29-13.06	2.82	9.59	6.26-14.11	2.45	9.51	4.29-17.10	3.23
	Temp (°C)	21.3	10.4-27.4	5.5	22.2	11.6-27.8	5.5	23.2	16.9-27.3	3.8	22.3	8.2-26.7	5.1
	Secchi (m)	3.8	2.5-5.0	0.9	3.5	2.2-6	1.4	3.5	2.0-5.0	1.0	3.2	2.0-4.5	0.7
	pH	8.05	7.60-8.41	0.25	8.14	7.60-8.42	0.23	8.10	7.83-8.39	0.20	8.32	7.92-8.95	0.29
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.229	0.04-0.795	0.239	0.154	0.053-0.462	0.117	0.152	0.04-0.353	0.094	0.158	0.057-0.377	0.105
	KMnO <sub>4</sub> (mg/L)	2.11	1.61-2.50	0.20	2.41	1.82-4.02	0.67	2.38	1.86-2.66	0.22	2.60	1.98-4.02	0.61
	TP (mgP/L)	0.024	0.007-0.109	0.026	0.017	0.005-0.025	0.006	0.017	0.004-0.028	0.006	0.019	0.008-0.033	0.008
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	0.015	0.005-0.049	0.012	0.014	0.003-0.02	0.006	0.014	0.006-0.023	0.006	0.016	0.005-0.029	0.007
	Chl-a (µg/L)	1.47	0.55-2.79	0.59	1.48	0.55-2.23	0.57	1.81	1.11-2.23	0.50	1.40	0.55-2.79	0.62
2011	DO (mgO <sub>2</sub> /L)	10.62	0.13-12.86	8.58	10.61	0.24-13.07	7.90	13.18	6.03-12.33	7.38	11.53	4.91-10.64	3.59
	Temp (°C)	16.7	8.89-28.73	6.2	18.0	9.3-28.42	6.2	19.5	11.15-28.37	5.7	19.2	11.43-28.37	5.8
	Secchi (m)	2.8	1.15-5	1.4	2.7	1.1-6.0	1.6	2.8	1.1-5.0	1.3	2.7	1.2-4.5	1.2
	pH	8.04	7.48-8.35	0.29	8.09	7.58-8.36	0.25	8.18	7.95-8.38	0.13	8.18	8.00-8.38	0.12
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.380	0.061-1.205	0.342	0.268	0.048-1.123	0.257	0.237	0.051-1.027	0.281	0.189	0.078-0.321	0.070
	KMnO <sub>4</sub> (mg/L)	2.19	1.10-2.90	0.41	2.19	1.34-4.38	0.52	2.23	1.72-3.58	0.48	2.40	1.53-3.56	0.44
	TP (mgP/L)	0.017	0.000-0.036	0.009	0.018	0.008-0.035	0.008	0.021	0.005-0.093	0.016	0.018	0.009-0.044	0.009
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	0.005	0-0.014	0.004	0.005	0-0.012	0.003	0.004	0-0.012	0.003	0.006	0.001-0.01	0.003
	Chl-a (µg/L)	1.57	0.056-7.26	1.54	1.53	0.5584-8.93	1.72	1.69	0-7.26	1.28	2.44	0.5584-13.4	3.34
2012	DO (mgO <sub>2</sub> /L)	5.71	0.25-11.36	4.07	6.20	1.02-11.60	3.57	10.38	8.31-15.11	1.71	9.24	7.65-12.25	1.47
	Temp (°C)	20.1	11.8-28.0	4.7	22.2	13.7-28.4	4.1	24.1	17.6-28.0	3.4	24.0	16.2-27.9	4.3
	Secchi (m)	1.7	0.5-2.95	1.0	1.2	0.6-2.0	0.5	1.3	0.9-1.6	0.3	1.1	0.8-1.4	0.2
	pH	8.08	7.54-8.19	0.28	8.19	7.70-8.46	0.20	8.46	8.21-8.77	0.20	8.34	7.87-8.48	0.17
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.576	0.104-3.067	0.830	0.424	0.112-0.837	0.191	0.358	0.113-0.632	0.137	0.253	0.116-0.593	0.126
	KMnO <sub>4</sub> (mg/L)	2.47	0.70-3.04	0.44	2.42	1.56-4.08	0.72	2.50	1.96-3.36	0.47	2.30	1.84-3.14	0.39
	TP (mgP/L)	0.018	0.006-0.037	0.008	0.025	0.003-0.053	0.017	0.024	0.007-0.062	0.022	0.015	0.006-0.026	0.007
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	0.004	0.001-0.007	0.002	0.005	0.002-0.009	0.002	0.004	0.002-0.01	0.003	0.003	0.002-0.005	0.001
	Chl-a (µg/L)	1.57	0.55-3.35	0.75	1.89	0.55-3.9	1.12	2.52	0.55-5.58	1.84	2.02	0.55-4.46	1.51
2013	DO (mgO <sub>2</sub> /L)	9.58	0.69-16.38	5.00	9.87	1.23-15.84	4.77	11.29	7.68-14.93	3.64	11.13	7.72-15.01	3.62
	Temp (°C)	17.8	9.5-26.7	5.5	18.1	10.3-26.7	5.2	21.1	12.6-26.2	4.7	20.0	10.7-25.9	4.8
	Secchi (m)	3.6	1.8-5.8	1.3	3.4	1.6-5.0	1.2	3.1	1.4-4.0	0.8	3.2	2.3-4.8	1.0
	pH	7.91	7.38-8.19	0.22	7.93	7.43-8.17	0.20	8.02	7.91-8.15	0.07	8.02	7.80-8.16	0.09
	NH <sub>4</sub> <sup>+</sup> -N (mgN/L)	0.407	0.052-1.598	0.368	0.312	0.101-1.100	0.206	0.210	0.114-0.493	0.093	0.228	0.102-0.633	0.130
	KMnO <sub>4</sub> (mg/L)	2.04	1.04-3.47	0.57	2.08	1.38-4.35	0.62	2.02	1.22-2.72	0.42	2.30	1.64-2.64	0.29
	TP (mgP/L)	0.011	0.000-0.072	0.010	0.011	0.003-0.046	0.010	0.011	0.006-0.019	0.004	0.010	0.004-0.019	0.004
	PO <sub>4</sub> <sup>3-</sup> -P (mgP/L)	0.002	0-0.017	0.003	0.003	0-0.013	0.003	0.004	0-0.01	0.004	0.002	0-0.008	0.002
	Chl-a (µg/L)	0.98	0-2.23	0.49	0.95	0-2.23	0.64	0.97	0-2.23	0.58	1.01	0-2.23	0.68

poboljšana korištenjem metode učenja temeljenog na modelu i metode učenja temeljenog na primjerima (Quinlan 1993).

Nakon izrade stabla i modela u obliku pravila iz seta podataka za treniranje (učenje), potrebno je procijeniti kvalitetu modela, odnosno točnost predviđanja. To se može učiniti simulacijom modela na podacima za testiranje te usporedbom predviđene vrijednosti sa stvarnim vrijednostima. Druga, češće korištena mogućnost je procjena modela primjenom metode unakrsne provjere (validacije). Zadani set podataka (za treniranje modela) podijeli se na odabrani broj dijelova ( $n$ ) te se svaki dio koristi za unakrsno testiranje, a ostatak ( $n-1$  dijelova) za treniranje modela. Konačna pogreška u modelu predstavlja prosječnu pogrešku svih dijelova modela tijekom cijelog postupka. U ovome radu korištena je metoda 10-dijelne unakrsne validacije, gdje je set podataka podijeljen u dvije grupe. Jednu grupu čini set za treniranje, a drugu grupu set za testiranje modela te se nakon postupka primjene klasifikatora izračunava prosjek performansa za dani model. Veličinu pogreške između stvarnih i simuliranih (predviđenih) vrijednosti moguće je izračunati koristeći nekoliko metoda za procjenu točnosti modela: korijen srednjih kvadratnih grešaka (RMSE), srednja apsolutna greška (MAE), korijen relativnih kvadratnih grešaka (RRSE), relativna apsolutna greška (RAE) te koeficijent korelacije ( $R$ ). Za izračunavanje točnosti modela u radu, korištene su sve mjere točnosti modela.

## 5. METODOLOGIJA IZRADE MODELA

Podaci za izradu modela prikazani su u [tablici 2](#). Podaci su prethodno obrađeni s obzirom na zadane ciljeve modeliranja i istraživanja. Za oba modela, opisni model klorofila-*a* (1. model) i simulacijski model klorofila-*a* (2. model), korišten je cijeli raspon mjerenih podataka na sve četiri mjerne postaje od 2006. do 2013. godine. Za potrebe modeliranja podaci su osrednjeni po dubini, odnosno iznad eufotičke dubine i uzeti kao jedan sloj. Opisni model klorofila-*a* prikazuje promjene u promatranom

ekosustavu tijekom promatranog razdoblja, dok se drugi model koristi za simulaciju klorofila-*a*.

Za izradu 1. modela klorofil-*a* je postavljen kao ciljane (zavisna) varijabla, dok su mjesec i godina sakupljanja podataka te vrijednosti čimbenika pH,  $\text{NH}_4^{+}\text{-N}$ ,  $\text{KMnO}_4$ , TP,  $\text{PO}_4^{3-}\text{-P}$ , DO, temperature i prozirnosti vode ([tablica 2](#)) postavljeni kao nezavisne varijable (deskriptori) preko kojih se modelira klorofil-*a*. Navedeni čimbenici korišteni su jer najbolje opisuju dijelove ekosustava o kojima ovisi ciljane varijabla. Za izradu 2. modela klorofil-*a* je postavljen kao ciljane (zavisna) varijabla, dok su vrijednosti čimbenika pH,  $\text{NH}_4^{+}\text{-N}$ ,  $\text{KMnO}_4$ , TP,  $\text{PO}_4^{3-}\text{-P}$ , DO, temperature i prozirnosti vode ([tablica 2](#)) postavljeni kao nezavisne varijable (deskriptori) preko kojih se simulira koncentracija klorofil-*a*.

Cilj je modela primjenjivost na čitavom promatranom području, odnosno pružanje što pouzdanije simulacije u odnosu na podatke s ostalih mjernih postaja. Za postizanje navedenog cilja, prilikom izrade 1. modela korištena je metoda 10-dijelne unakrsne validacije, kao najčešća metoda za konstruiranje i testiranje modela. Konačna pogreška prikazana je kao prosječna pogreška svih dijelova modela tijekom cijelog postupka izrade modela. Za izradu 2. modela također je korištena ista metodologija. Model koji daje najpouzdaniju simulaciju prema metodi 10-dijelne unakrsne validacije i koji ima dobru opisnu moć, odabran je kao reprezentativan model klorofila-*a* za promatrano područje. Za provjeru statističke značajnosti, odnosno pouzdanosti modela, u oba slučaja korišteni su korijen srednjih kvadratnih grešaka (RMSE), srednja apsolutna greška (MAE), korijen relativnih kvadratnih grešaka (RRSE), relativna apsolutna greška (RAE) te koeficijent korelacije ( $R$ ) između modeliranih i mjerenih vrijednosti koncentracije klorofila-*a*.

## 6. REZULTATI I DISKUSIJA

### 6.1. OPISNI MODEL KLOROFILA-*a*

Koeficijent korelacije za odabrani opisni model ([slika](#)

Tablica 2. Fizikalno-kemijski i biološki čimbenici korišteni za izradu modela.

Simbol/čimbenik	Opis čimbenika	Mjerna jedinica	Korišteno u modelu
GODINA	Godina sakupljanja podataka	-	(1)
MJESEC	Mjesec sakupljanja podataka	-	(1)
Chl- <i>a</i>	Klorofil- <i>a</i>	$\mu\text{g/L}$	(1), (2)
DO	Koncentracija otopljenog kisika	$\text{mgO}_2/\text{L}$	(1), (2)
Temp	Temperatura	$^{\circ}\text{C}$	(1), (2)
Secchi	Prozirnost vode	m	(1), (2)
pH	pH	/	(1), (2)
$\text{NH}_4^{+}\text{-N}$	Amonijak	$\text{mgN/L}$	(1), (2)
$\text{KMnO}_4$	Kalijev permanganat	$\text{mg/L}$	(1), (2)
TP	Ukupni fosfor	$\text{mgP/L}$	(1), (2)
$\text{PO}_4^{3-}\text{-P}$	Ortofosfat	$\text{mgP/L}$	(1), (2)

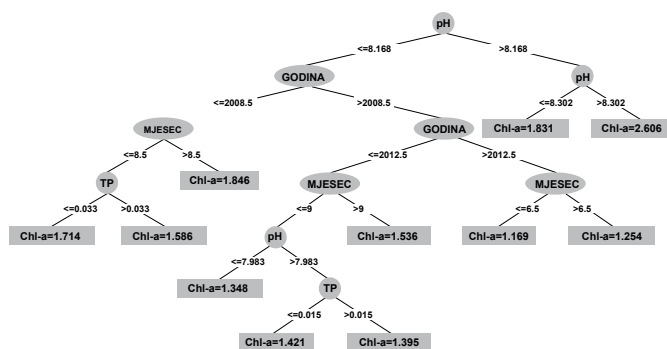
Tablica 3: Pravila u simulacijskom modelu koncentracije klorofila-a za akumulaciju Butoniga.

Br. pravila	Pravilo	Jednadžba	Br. slučajeva / % za RMSE
Pravilo 1	pH ≤ 8.168 TP ≤ 0.012 pH ≤ 8.021 TP > 0.008	Chl-a = 0.9785 * pH - 12.5839 * TP - 0.0164 * DO - 0.0275 * Temp - 5.8635	13/24.04 %
Pravilo 2	pH ≤ 8.185 Temp > 22.005 Temp ≤ 24.77	Chl-a = 0.332 * pH - 0.019 * Temp - 0.9286	14/21.25 %
Pravilo 3	pH > 8.2 TP ≤ 0.023	Chl-a = 1.0541 * pH - 0.0629 * Temp - 5.3099	13/51.28 %
Pravilo 4	pH > 8.141 pH ≤ 8.302	Chl-a = 4.6205 * pH - 0.1874 * Secchi - 35.3143	14/27.33 %
Pravilo 5	pH ≤ 8.11 DO > 8.909 TP > 0.026 pH > 7.777	Chl-a = 1.3015 * pH - 8.4381	14/48.32 %
Pravilo 6	pH ≤ 8.117 TP ≤ 0.018 TP ≤ 0.012	Chl-a = 1.0366 * pH + 20.8577 * TP - 0.0296 * Temp - 6.3785	14/19.94
Pravilo 7	pH ≤ 8.127 TP ≤ 0.018	Chl-a = 1.5034 * pH + 10.8034 * PO <sub>4</sub> <sup>3-</sup> -P - 10.4232	13/37.08
Pravilo 8	pH ≤ 8.137 TP ≤ 0.038 PO <sub>4</sub> <sup>3-</sup> -P > 0.007 TP ≤ 0.031	Chl-a = 1.6862 * pH + 1.4648 * TP + 2.9775 * PO <sub>4</sub> <sup>3-</sup> -P - 12.1216	12/15.06
Pravilo 9	pH ≤ 8.137 TP > 0.029	Chl-a = 2.3495 * pH + 1.9994 * TP - 17.3215	14/22.62
Pravilo 10	pH ≤ 8.222	Chl-a = 5.1277 * pH - 39.7769	13/17.67 %
Pravilo 11	-	Chl-a = + 6.0912	4/100 %

4) iznosi 0,76, ali je, unatoč relativno nižoj vrijednosti te uzimajući u obzir kompleksnost rješavanja domene i manjak podataka odgovarajuće točnosti i frekvencije, model postigao pouzdan opis dinamike klorofila-a. Ostale mjere točnosti za dani opisni model su MAE (0,522), RMSE (1,012), RAE (75,34 %) te RRSE (77,54 %). Na slici 4 može se uočiti kako koncentracija klorofila-a u akumulaciji Butoniga tijekom promatranog razdoblja prvenstveno ovisi o vrijednostima pH, koje je strojno učenje istaknulo kao najznačajniji čimbenik u konstrukciji modela. Od ostalih čimbenika, koncentracija klorofila-a ovisila je i o koncentracijama ukupnog fosfora (TP), dok model nije izdvojio koncentraciju otopljenog kisika, prozornost vode, temperaturu te koncentracije NH<sub>4</sub><sup>+</sup>-N, KMnO<sub>4</sub> i PO<sub>4</sub><sup>3-</sup>-P kao bitne čimbenike. Na slici 4 također se mogu opaziti vremenske promjene u koncentracijama klorofila-a koje su u modelu definirane čimbenicima kao što su mjesec i godina.

Promatrajući desnu stranu modela možemo uočiti kako više koncentracije klorofila-a ovisе o

višoj vrijednosti pH (> 8,168). Promatrajući lijevu stranu modela mogu se opaziti promjene u koncentracijama klorofila-a sredinom 2008. godine (2008,5). Prije 2008. godine koncentracije ovisе o mjesecu uzorkovanja (8,5), odnosno do 8. mjeseca koncentracija klorofila-a ovisi i o koncentraciji ukupnog fosfora (TP). Između 2008. i 2012. godine koncentracije klorofila-a ovisе o mjesecu



Slika 4: Opisni model klorofila-a za akumulaciju Butoniga u istraživanom razdoblju.

uzorkovanja te koncentracijama pH i ukupnog fosfora. Nakon 2012. godine koncentracije klorofila-*a* povezane su s mjesecom uzorkovanja (6,5). Tako je istraživanjem tijekom 2009. godine na postaji B1 također utvrđeno kako je najviša koncentracija klorofila-*a* zabilježena u listopadu (2,79 mg L<sup>-1</sup>), a najniža tijekom lipnja, srpnja i kolovoza (< 0,5 mg L<sup>-1</sup>), dok se istodobno pH-vrijednost vode postupno povećavala od lipnja prema listopadu (Božinović 2010).

## 6.2. SIMULACIJSKI MODEL KLOROFILA-*a*

Simulacijski model koncentracije klorofila-*a* sastoji se od 11 pravila, od kojih je svako povezano s odgovarajućom linearnom jednadžbom koja se koristi za izračun koncentracije klorofila-*a*. Nakon odabira pravila primjenjuje se odgovarajuća linearna jednadžba (tablica 3). Koeficijent korelacije za odabrani simulacijski model klorofila-*a* iznosi 0,87, što predstavlja visoku razinu korelacije s obzirom na kompleksnost rješavanja domene. Ostale mjere točnosti za dani simulacijski model su MAE (0,336), RMSE (0,781), RAE (48,33 %) te RRSE (51,27 %). Na slici 5. prikazano je ponašanje modela (tablica 3), odnosno usporedba modeliranih i izmjerenih vrijednosti koncentracija klorofila-*a* za svaku pojedinu postaju. Na slici se može uočiti visoka pouzdanost modela u usporedbi s izmjerenim podacima, te relativno dobro predviđanje vršnih vrijednosti koncentracija klorofila-*a*. Najveći koeficijent korelacije pri testiranju simulacijskog modela zasebno na pojedinim postajama, dobiven je za postaju B2 (0,89), a slijede postaje B3 (0,83), B4 (0,80) i B1 (0,73).

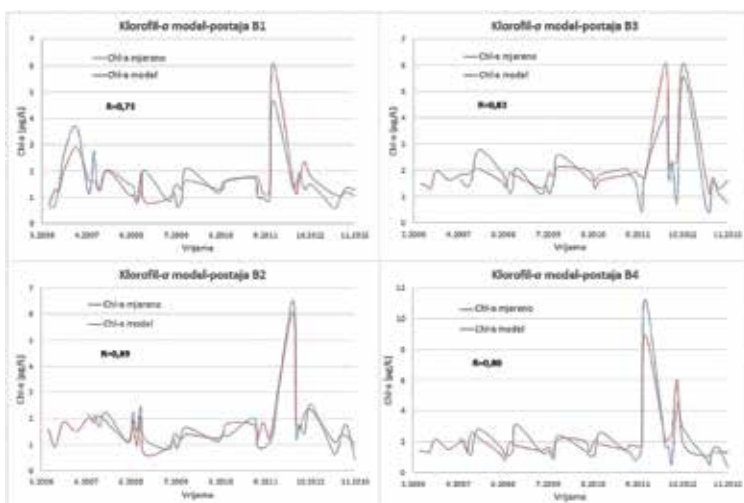
Kako bi se dodatno provjerila točnost dobivenog modela, model je testiran na način da je 80 % podataka korišteno za treniranje, a zatim preostalih 20 % za testiranje dobivenog simulacijskog modela. Kako se ova metoda inače koristi kod velikih setova podataka, dobiveni koeficijent korelacije je nešto manji te iznosi 0,78 prilikom treniranja modela i 0,69 prilikom testiranja modela. Ostale mjere točnosti za provedenu dodatnu

provjeru točnosti dobivenog simulacijskog modela su sljedeće te prilikom treniranja iznose MAE (0,513), RMSE (0,766), RAE (73,28 %) te RRSE (61,43 %), odnosno prilikom testiranja na preostalim 20 % podataka iznose MAE (0,467), RMSE (0,662), RAE (69,30 %) te RRSE (77,03 %).

U okviru ovog istraživanja pomoću alata za strojno učenje po prvi je puta razvijen jednostavan te učinkovit model koji simulira kretanje koncentracije klorofila-*a* u akumulaciji Butoniga. S obzirom na to da u svojoj strukturi model uključuje osam okolišnih čimbenika koji igraju važnu ulogu u simulaciji i predviđanju kretanja koncentracije klorofila-*a* u ovom ekosustavu, model se može koristiti kao alat u svrhu upravljanja akumulacijom. Upravo modeliranje koncentracije klorofila-*a* u ovisnosti o hranjivim tvarima, prvenstveno ukupnom fosforu te drugim okolišnim čimbenicima poput temperature vode, pH, i dr. omogućuje uspješnu kontrolu eutrofikacije i upravljanje mnogih sustava stajaćica i tekućica (Pan i sur. 2009; Li i sur. 2017; Deng i sur. 2019). S obzirom na dobivenu visoku pouzdanost simulacije kretanja koncentracije klorofila-*a* te ponašanja na neviđenim podacima, model se može uspješno primijeniti kao alat za upravljanje akumulacijom u obliku samostojeće simulacije klorofila-*a* ili može biti ugrađen u složenije modele sliva koji bi obuhvaćali aktivnosti poput predviđanja dotoka i kontrole opterećenja hranjivim tvarima u slivu.

## 7. ZAKLJUČCI

U predstavljenom istraživanju primijenjene su dvije metode strojnog učenja na višegodišnjim mjernim podacima prikupljenim iz akumulacije Butoniga s ciljem izrade (1) opisnog modela klorofila-*a* kojim se opisuju promjene koncentracije klorofila-*a* u akumulaciji Butoniga i određuju najznačajniji čimbenici tih promjena, te (2) simulacijskog modela klorofila-*a* koji daje procjenu koncentracije klorofila-*a* u akumulaciji s obzirom na promatrane vrijednosti izmjerenih okolišnih čimbenika. Prvi model predstavlja opisni model klorofila-*a* koji objašnjava promjene koncentracije klorofila-*a* tijekom promatranog razdoblja te identificira čimbenike tih promjena, kao što su utjecaj promjene vrijednosti pH i koncentracije ukupnog fosfora. Drugi model simulira kretanje koncentracije klorofila-*a* s obzirom na promatrane vrijednosti osam okolišnih čimbenika te se kao takav može koristiti kao alat u svrhu upravljanja akumulacijom. Potrebno je naglasiti relativno visoku pouzdanost modela u usporedbi s izmjerenim vrijednostima te relativno dobro predviđanje vršnih vrijednosti koncentracija klorofila-*a* na pojedinim mjernim postajama. Budući rad fokusiran je na procjenu utjecaja vode iz akumulacije Butoniga na rad postrojenja za kondicioniranje vode kako bi se dodatno optimizirao rad samog uređaja. ■



Slika 5: Rezultati modela za simuliranje koncentracije klorofila-*a* za akumulaciju Butoniga na mjernim postajama B1, B2, B3 i B4 u istraživanom razdoblju od 2006. do 2013. godine.



## 8. ZAHVALA

Ovaj rad podržan je projektima "Održivo upravljanje riječnim slivom primjenom inovativnih metodologija, pristupa i alata" (uniri-tehnic-18-129) financiranog od strane Sveučilišta u Rijeci, "Utjecaj ljetnih požara na tlo i kakvoću vode" koji je financiran od strane Hrvatske zaklade za znanost (IP-2018-01-1645) te "Istraživanje

i optimizacija ihtiocenoze u svrhu smanjenja trofije akumulacije Butoniga od 2006. do 2013. godine". Također, ovaj rad je financiran u okviru projektne linije ZIP UNIRI Sveučilišta u Rijeci, za projekte ZIP-UNIRI-1500-3-22 i ZIP-UNIRI-1500-2-22. Dio istraživanja provedenog u ovom radu je predstavljen na 8. Hrvatskoj konferenciji o vodama.

## LITERATURA

- Atanasova, N., Todorovski, L., Džeroski, S., Kompare, B. 2008. Application of automated model discovery from data and expert knowledge to a real-world domain: Lake Glumsø. *Ecological Modelling*. 212, 92-98.
- Bernát, G., Boross, N., Somogyi, B., Vörös, L., G.-Tóth, L., Boros, G. 2020. Oligotrophication of Lake Balaton over a 20-year period and its implications for the relationship between phytoplankton and zooplankton biomass. *Hydrobiologia*. 847. 10.1007/s10750-020-04384-x.
- Božinović, K. 2010. Zooplankton akumulacije Butoniga. Diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet. <https://urn.nsk.hr/urn:nbn:hr:217:642704>.
- Crisci, C., Ghattas, B., Perera, G. 2012. A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*. 240, 113-122.
- Deng, J., Chen, F., Hu, W., Lu, X., Xu, B., Hamilton, D.P. 2019. Variations in the distribution of Chl-a and simulation using a Multiple Regression Model. *International Journal of Environmental Research and Public Health*. 16, 4553. 10.3390/ijerph16224553.
- Effrosynidis, D., Arampatzis, A., Sylaios, G. 2018. Seagrass detection in the Mediterranean: A supervised learning approach. *Ecological Informatics*. 48, 158-170.
- Flach, P., 2001. On the state of the art in machine learning: a personal review. *Artificial Intelligence*. 13 (1), 199-222.
- Gal, G., Škerjanec, M., Atanasova, N. 2013. Fluctuations in water level and the dynamics of zooplankton: a data-driven modelling approach. *Freshwater Biology*. 58 (4), 800-816.
- Hajduk Černeha, B., 2017. Butoniga reservoir - Loads in the basin and water protection, Znanstveno stručni skup, Upravljanje jezerima i akumulacijama u Hrvatskoj, Rubinić (ed), pp 86-95, Biograd na Moru.
- HRN ISO 5667-4:2016 Kvaliteta vode - Uzorkovanje - 4. dio: Upute za uzorkovanje vode prirodnih i umjetnih jezera (ISO 5667-4:2016)
- HRN EN ISO 5667-3:2018 Kvaliteta vode - Uzorkovanje - 3. dio: Čuvanje i rukovanje uzorcima vode (ISO 5667-3:2018; EN ISO 5667-3:2018)
- Latif, Z., Tasneem, M. A., Javed, T., Butt, S., Fazil, M., Ali, M., Sajjad, M.I. 2003. Evaluation of water-quality by chlorophyll and dissolved oxygen. In: *Water resources in the south: present scenario and future prospects*. Commission on Science and Technology for Sustainable Development in the South, Islamabad, Pakistan, pp. 122-135.
- Li, X., Sha, J., Wang, Z.-L. 2017. Chlorophyll-a prediction of lakes with different water quality patterns in China based on Hybrid Neural Networks. *Water*. 9, 524. <https://doi.org/10.3390/w9070524>
- Mori, N., Debeljak, B., Škerjanec, M., (...), Kanduč, T., Brancelj, A. 2019. Modelling the effects of multiple stressors on respiration and microbial biomass in the hyporheic zone using decision trees. *Water Research*. 149, 9-20.
- Pan, B., Wang, H., Liang, X., Wang, H. 2009. Factors influencing chlorophyll *a* concentration in the Yangtze-connected lakes. *Fresenius Environmental Bulletin*. 18(10), 1894-1900.
- Quinlan, J. R., 1986. Induction of Decision Trees. *Machine Learning*, 1, 81-106.
- Quinlan, J.R., 1993. Combining instance-based and model-based learning. In: *Proceedings ML'93*, Utgoff (ed), Morgan Kaufmann.
- Quinlan, J.R., 1992. Learning with continuous classes. *Proc. AI'92, 5th Australian Joint Conference on Artificial Intelligence*, Adams & Sterling (eds.), Singapore, World Scientific, 343-348.
- Volf, G., Atanasova, N., Kompare, B., Precali, R., Ožanić, N., 2011. Descriptive and prediction models of phytoplankton in the northern Adriatic. *Ecological Modelling*. 222, 2502-2511.
- Volf, G., Atanasova, N., Kompare, B., Precali, R., Ožanić, N., 2015. Study of the impact of TIN/PO<sub>4</sub> ratio on mucilage formation in the northern Adriatic using regression trees. *Acta Adriatica*. 56 (2), 207-222.
- Wang, F., Wang, X., Chen, B., Zhao, Y., Yang, Z., 2013. Chlorophyll *a* simulation in a lake ecosystem using a Model with Wavelet Analysis and Artificial Neural Network. *Environmental Management*. 51, 1044-1054.
- Witten, I.H., Frank, E., 2000. *Data mining – Practical machine learning tools and techniques with java implementations*. Academic Press, USA.

## MODEL OF CHLOROPHYLL-A FOR THE BUTONIGA RESERVOIR IN ISTRIA

**Abstract:** The Butoniga reservoir is relatively small and shallow, and, as such, very sensitive to degradation and eutrophication processes caused both by climate change and by anthropogenic activities in its basin. Chlorophyll-a is a widely used ecological indicator of primary production, i.e. phytoplankton biomass and the process of eutrophication in reservoirs, depending on the availability of nutrients, pH value, light, temperature, as well as their interrelations. The machine learning technique for creating models in the form of regression trees and rules was applied to a set of measurement data in order to better understand the eutrophication process and the ecological state of the Butoniga reservoir. Using machine learning tools, two models were created: (1) a descriptive model of chlorophyll-a in the form of regression trees, which describes changes in chlorophyll-a concentrations and determines the most significant factors causing these changes; and (2) a simulation model of chlorophyll-a in the form of rules that serves to predict the chlorophyll-a concentration in relation to the observed values of the measured environmental factors. The created models contribute to a better understanding of the ecosystem of the Butoniga reservoir, and can be used for the reservoir management as well.

**Keywords:** Butoniga reservoir, chlorophyll-a, machine learning, models, phytoplankton

## CHLOROPHYLLMODELL FÜR DEN BUTONIGA-STausee IN ISTRIEN

**Abstrakt:** Butoniga gehört zu den relativ kleinen und flachen Stauseen und reagiert daher sehr empfindlich auf die Degradations- und Eutrophierungsprozesse, die sowohl durch Klimaveränderungen als auch durch anthropogene Aktivitäten in seinem Becken verursacht werden. Chlorophyll ist ein weit verbreiteter ökologischer Indikator der Primärproduktion, also der Phytoplankton-Biomasse, und des Prozesses der Eutrophierung in Stauseen, der von der Verfügbarkeit von Nährstoffen, dem pH-Wert, dem Licht, der Temperatur sowie deren Wechselbeziehungen abhängt. Die Technik des maschinellen Lernens zur Erstellung von Modellen in Form von Klassifikations- und Regressionsbäumen wurde auf eine Reihe von Messdaten angewendet, um den Prozess der Eutrophierung und den ökologischen Zustand des Butoniga-Stausees besser zu verstehen. Mithilfe von Tools des maschinellen Lernens wurden zwei Modelle entwickelt: (1) eine Modellbeschreibung von Chlorophyll in Form von Regressionsbäumen, das Änderungen in der Konzentration von Chlorophyll beschreibt und die wichtigsten Faktoren dieser Änderungen bestimmt, und (2) ein Simulationsmodell von Chlorophyll in Form von Regeln, die dazu dienen, die Konzentration von Chlorophyll im Hinblick auf die beobachteten Werte der gemessenen Umweltfaktoren vorherzusagen. Die erstellten Modelle tragen zu einem besseren Verständnis des Ökosystems des Butoniga-Stausees bei und können auch für dessen Bewirtschaftung genutzt werden.

**Schlüsselwörter:** Butoniga-Stausee, Chlorophyll, maschinelles Lernen, Modelle, Phytoplankton