

<https://doi.org/10.15255/KUI.2022.033>

KUI-61/2022

Izvorni znanstveni rad

Prispjelo 31. ožujka 2022.

Prihvaćeno 27. lipnja 2022.

# Kauzalni ekološki model sjevernog Jadrana temeljem podataka EU projekta “LTER Northern Adriatic Sea”

Ž. Kurtanjek\*

Ovo djelo je dano na korištenje pod  
Creative Commons Attribution 4.0  
International License

Sveučilište u Zagrebu, Prehrambeno-biotehnološki fakultet, Pierrotijeva 6, 10 000 Zagreb

## Sažetak

Cilj ovog rada je pokazati mogućnosti primijene metodologije umjetne inteligencije i strukturnog kauzalnog modeliranja (engl. *Structural Causal Model*, SCM) s ciljem postizanja znanstvenog doprinosa utvrđivanjem kauzalne funkcionalne zakonitosti bioloških značajki o abiotičkim parametrima. Temeljna zadaća rada je istražiti model SCM za određivanje zavisnosti koncentracije klorofila o fizikalnim značajkama u području sjevernog Jadrana tijekom razdoblja od 1965. do 2015. godine. Eksperimentalni podatci rezultat su dugotrajnog i ekstenzivnog istraživanja u okviru EU projekta “LTER Northern Adriatic Sea” i dostupni su (putem EU znanstvene politike “Open Science”) u velikoj bazi podataka (engl. *Big Data*), koja sadrži 10 8687 uzoraka s 43 značajke. Predloženi je matematički model Bayesove mreže (engl. *Bayes Network*, BN) kao usmjereni neciklički graf (engl. *Directed Acyclic Graph*, DAG). Struktura grafa određena je primjenom testa uvjetne nezavisnosti (Hamilton-Schmidtova Conditional Independence test, HSCI) s razinom signifikantnosti  $\alpha = 0,05$ . SCM model pokazuje da su neposredni kauzalni utjecaji na koncentraciju klorofila: temperatura, salinitet, pH, dušik, fosfor i silicij. Primijenjena je metodologija d-razdvajanja BN grafa sa svrhom blokiranja interferencije (engl. *confounding*) za procjenu kauzalne funkcionalne zavisnosti bioloških značajki o abiotičkim parametrima. Funkcije kauzalnosti određene su kao rubne razdiobe (engl. *marginal distributions*) modeliranjem Bayesovom neuronskom mrežom (engl. *Bayes Neural Network*, BNN). Najveći neposredni negativni kauzalni učinak na klorofil A (*Chlorophyll A*) ima temperatura ( $-0,07 \mu\text{g}$  klorofila  $^{\circ}\text{C}$ ). Utvrđena je pozitivna kauzalna zavisnost između klorofila-A i otopljenog kisika ( $0,2 \text{ mg}$  otopljenog kisika  $\text{DO}_2/\mu\text{g}$  klorofila A). Također je provedena neparametarska usporedna analiza klorofila A i fizikalnih parametara hrvatskog dijela i podataka za cjelokupni sjeverni Jadran. Medijan koncentracije otopljenog kisika u hrvatskom dijelu Jadrana je  $5,8 \text{ mg O}_2/\text{l}$  a u sjevernom je  $5,5 \text{ mg O}_2/\text{l}$ , dok je medijan temperature u hrvatskom dijelu  $T = 14,6 \text{ }^{\circ}\text{C}$  u odnosu na  $T = 15,1 \text{ }^{\circ}\text{C}$  za sjeverni Jadran. Medijan broja stanica bičaća (*Dinoflagellate*) je u hrvatskom dijelu Jadrana 3 stanice/l, u odnosu na cijeli sjeverni Jadran, gdje je on od 5 stanica/l. Značajna je razlika u učestalosti i iznosu visokog broja bičaća. Medijani koncentracija klorofila A ne pokazuju značajnu razliku ( $0,65$  i  $0,90 \mu\text{g l}^{-1}$ ), ali u sjevernom Jadranu je znatno veći broj uzoraka koji po iznosu signifikantno odstupaju od normalne razdiobe (engl. *outliers or hotspots*). Utvrđena je i značajna razlika u razdiobi koncentracije silicija s velikim brojem uzoraka s visokim koncentracijama u zapadnom dijelu Jadrana. Primijenjeni su modeli “šume” stabala odlučivanja (engl. *random forest*) predikcije bioloških značajki na osnovi abiotičkih veličina. Validacije modela provedene su određivanjem relativne postotne pogreške predikcije primjenom simulacije “novih” podataka peterostrukom podjelom baze podataka. Postignute su sljedeće pogreške modela predikcije: za klorofil (engl. *chlorophyll*) 6,5 %; feopigment (*Pheopigment*) 17,4 %; diatomeje (*Diatom*) 18,8 %; dinoflagelat (*Dinoflagellate*) 17,4 %; i kokolitofore (*Coccolithophores*) 12,1 %. Za svaki od modela utvrđeni su ključni abiotički faktori za procjenu predikcija.

## Ključne riječi

Strukturni kauzalni model (SCM), biosfera mora (LTER), Jadran

## 1. Uvod

Mediteran je u idućih desetak godina prema predviđanjima ekoloških modela globalnih klimatskih promjena istaknuto kritično područje (engl. *climate hot spot*) zbog očekivanog smanjenja precipitacije od 40 do 60 % , uz godišnje povećanje prosječne temperature za  $2,5 \text{ }^{\circ}\text{C}$ . Specifičnost kritičnosti područja Mediterana je međusobni utjecaj naglog povećanja zagrijavanja kontinentalnog dijela sjeverne Afrike i poremećaja cirkulacije gornjih slojeva atmosfere. Zbog zatvorenosti Mediterana s afričkim i europskim kontinentom dolazi do posebno kritičnih promjena morske biosfere. Predikcije modela zagrijavanja Mediterana predviđaju moguće povećanje temperature mora tijekom ljetnog razdoblja od  $2$  do  $6 \text{ }^{\circ}\text{C}$ .<sup>1-3</sup> Područje sjeverni Jadran (engl. *North Adriatic Sea*, NAS) u okviru Mediterana ima istaknutu ulogu u hvatanju i pohrani  $\text{CO}_2$  s procjenom

brzine od  $3 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$ . Iako je NAS samo 0,15 % površine Mediterana, ima važnu ulogu u bilanci  $\text{CO}_2$ . Globalni klimatološki modeli predviđaju porast razine mora za 30 do 40 cm do 2100. Te promjene imat će posljedice za gradove uz morsku obalu i, posebice, kataklizmičke posljedice za agronomska polja uz deltu Neretve. Osim globalnih klimatoloških promjena, velik je specifični antropološki utjecaj na more Jadrana. Posljedice ljudske aktivnosti su mnogobrojne i djeluju na biološku ravnotežu ekološkog sustava Jadrana. Neposredno onečišćenje (bi-orazgradivi otpad, plastični otpad, ugljikovodici, svjetlost, buka), masovni turizam, agrikultura, ribarstvo, industrijska proizvodnja) zahtijeva sustav nadzora okoliša i dugoročno planiranje održivog gospodarskog razvoja. EU Integrirani sustav održivog razvoja planiranja opisan je u dokumentu EU komisije “Assessing and valuing ecosystem services in the Northern Adriatic”.<sup>5</sup> Ministarstvo gospodarstva i održivog razvoja RH donijelo je “Akcijski program strategije upravljanja morskim okolišem i obalnim područjem” s

\* Prof. dr. sc. Želimir Kurtanjek, u mirovini  
e-pošta: [zelimir.kurtanjek@gmail.com](mailto:zelimir.kurtanjek@gmail.com)

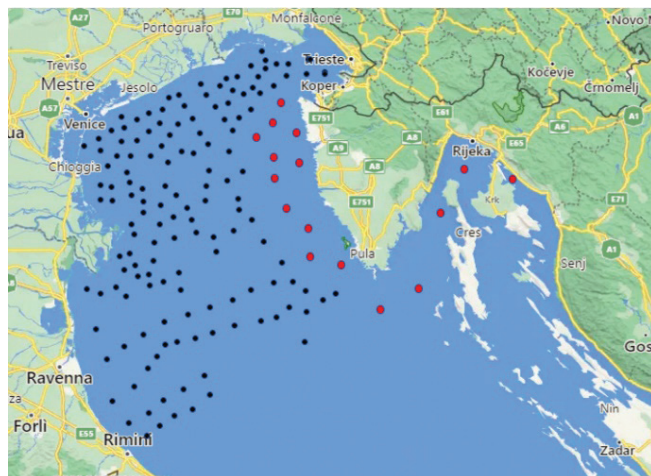
mjerama nadzora i stalnom procjenom stanja mora za razdoblje 2021. – 2026.<sup>6</sup> Oba dokumenta detaljno opisuju metodologije nadzora i deskriptivne statistike za integraciju podataka i grafičku vizualizaciju.

Zbog složenosti i nelinearnosti interakcija, predloženi sustavi nadzora ne omogućuju zaključivanje o uzročno posljedičnim vezama te i nisu u mogućnosti dati odgovore na pitanja o intervencijskim odlukama (politikama) u borbi s globalnim klimatskim promjenama, gospodarskim razvojem i održivosti okoliša. Napredak u razumijevanju i modeliranju ekoloških sustava očekuje se od primjene umjetne inteligencije (engl. *artificial intelligence*, AI) i strojnog učenja (engl. *machine learning*, ML) za analizu velikih baza podataka (engl. *Big Data*). Nedavna studija grupe norveških autora iz instituta istraživanja mora iz primjene područja umjetne inteligencije ističe važnost povezivanja metodologija računalnih znanosti i ekspertnog znanja iz biologije i ekologije.<sup>7</sup> Računalne znanosti omogućuju *on-line* analizu velikog broja podataka iz satelitskih mjernih sustava, analizu slike, morskih i zemaljskih mjernih stanica, meteoroloških mjerenja i modela predikcije te socioekonomskih podataka. Ključna metodologija je u primjeni algoritma dubokog učenja (engl. *deep learning*, DL) s neuronskim mrežama. Na primjer, algoritam DL primijenjen je za automatski nadzor biološke raznovrsnosti računalnim prepoznavanjem slika riba i audio-zapisa ptica i životinja. *On-line* automatizacija i DL analiza omogućuju prepoznavanje i razlučivanje slika u hidrodinamičkim uvjetima uz prostornu analizu biosfere oceana.<sup>9</sup> Temeljne jednadžbe fizike i bilance prijenosa tvari (numeričke aproksimacije Navier-Stokesovih parcijalnih diferencijalnih jednadžbi) povezuju se kao integrirane (proširene) funkcije cilja učenja neuronskih mreža (engl. *physically informed neural networks*, PINN) za modeliranje oceana i globalnih klimatskih promjena.<sup>10</sup> Primjenom analize osjetljivosti DL modela moguće je odrediti najvažnije ekološke varijable koje utječu na održivost biosfere i dati osnovu za upravljanje ekonomskih mjera održivosti ekoloških sustava.<sup>11</sup> Za dubinsko funkcionalno razumijevanje složenih sustava potrebno je primijeniti načela umjetne opće inteligencije (engl. *Artificial General Intelligence*, AGI) koja omogućuju razlučivanje u uvjetima složenih interferencija ekoloških parametara i primjenom strukturnih kausalnih modela (engl. *Structural Causal Model*, SCM) analizirati intervencijske odluke nužne za održivo gospodarstvo.<sup>12</sup> Cilj ovog rada je razviti SCM model i odrediti funkcionalne zavisnosti ekoloških parametara sjevernog Jadrana.

## 2. Eksperiment i baza podataka

Upotrijebljeni eksperimentalni podatci nadzora sjevernog Jadrana rezultat su EU projekta Open Science s podatcima dostupnim za istraživanje.<sup>13</sup> Baza podataka LTER (engl. *Long Term Ecology Research, Northern Adriatic Sea*) rezultat je 50 godina praćenja fizikalno-kemijskih i bioloških parametara tijekom 1965. – 2015. Podatci su prikupljeni sa 79 istraživačkih lokacija u sjevernom Jadranu u Italiji, Sloveniji i Hrvatskoj.<sup>14–16</sup> Za analizu je primijenjena baza

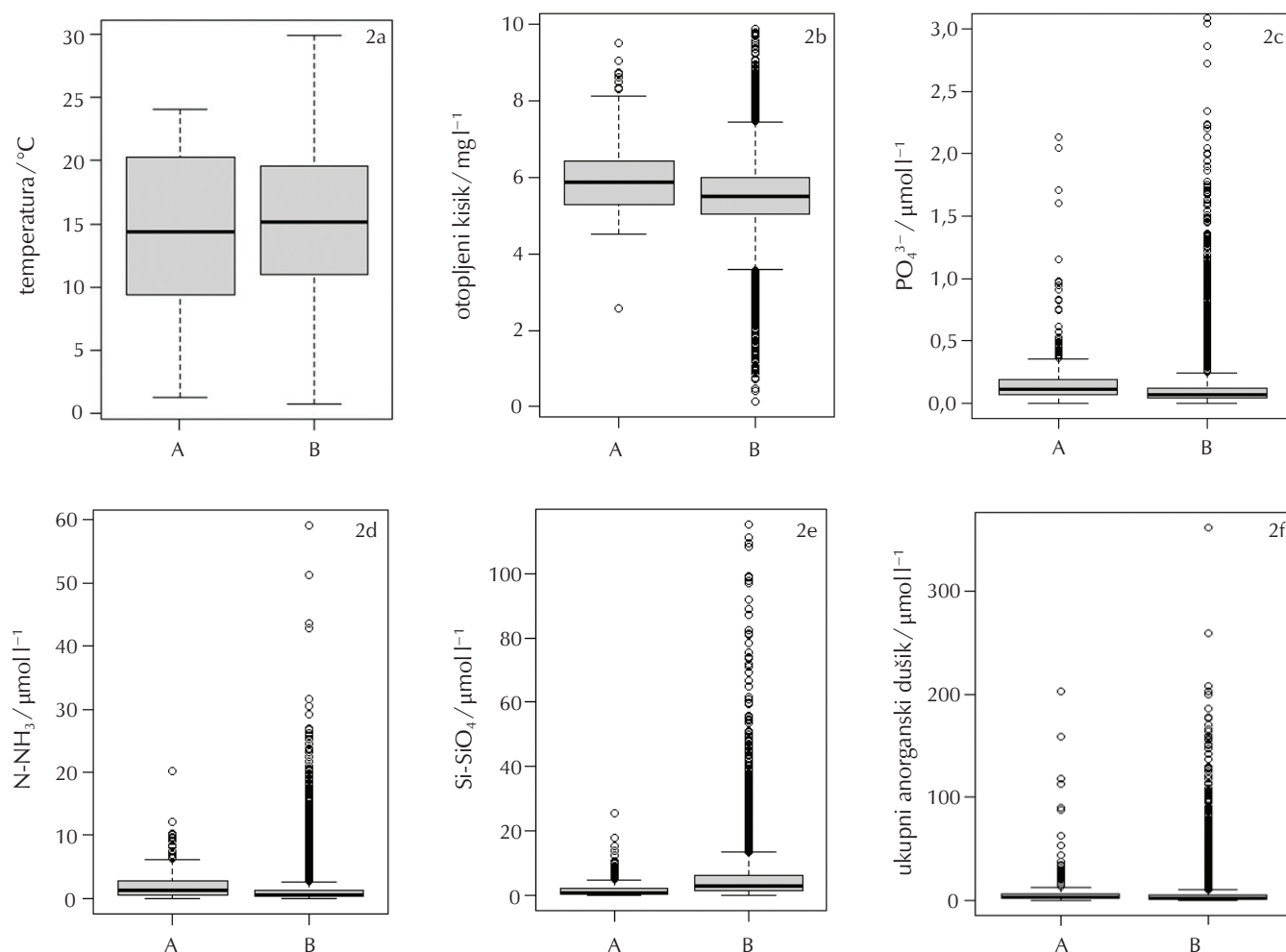
s 108 687 uzoraka sa 16 fizikalnih i kemijskih (abiotičkih) i sedam bioloških parametara (fitoplanktoni i zooplanktoni). Podatci su anotirani s datumom uzorkovanja, geografskim položajem mjernog mjesta, dubinom mora i dubinom mjesta uzimanja uzorka. Praćeni parametri su: prozornost mora, temperatura, salinitet, gustoća, pH, otopljeni kisik, relativno zasićenje kisikom,  $\text{NH}_3$ ,  $\text{NO}_2^-$ ,  $\text{NO}_3^-$ , ukupni anorganski dušik,  $\text{PO}_4^-$ , silicij, alkalnost, klorofil A, feopigmenti (*Pheopigment*), dijatomeji (*Diatom*), dinoflagelati (*Dinoflagellate*), kokolitofori (*Coccolithoophores*), ukupni fitoplanktoni (*Phytoplanktons*) i ukupni mesozooplanktoni (*Mesozooplanktons*). Na slici 1 shematski su prikazana mjerna mjesta, a detaljna karta dostupna je u radovima *Acri i sur.* i *Minelli i sur.*<sup>14,15</sup> Većina mjernih mjesta smještena je u širem području lagune Venecije. Za svaki podatak iz baza poznata je informacija o mjernim instrumentima, nazivima stacionarnih točaka i nazivima istraživačkih brodova. Podatci su pohranjeni u .xls i .csv formatima. Analiza podataka provedena je računalnom podrškom R za statistiku i podatkovnu znanstvenu analizu (engl. *data science*).<sup>17</sup>



Slika 1 – Shematski prikaz mjernih mjesta. Položaji u Hrvatskoj označeni su crveno, a mjerna mjesta u Sloveniji i Italiji označena su crno.<sup>14,15</sup>

Fig. 1 – Schematic depiction of the measurement sites. The red dots mark the sites in Croatia Adriatic Sea, while the black are in Slovenia and Italy.<sup>14,15</sup>

Za komparativnu analizu podataka iz hrvatskog (A) i sjevernog Jadrana (Italija i Slovenija, B) podatci su na slici 2 prikazani kao "box plots". Podatci obuhvaćaju sva mjerna mjesta kroz cjelokupno razdoblje od 1965. do 2015. godine. Medijani temperature mora hrvatskog i sjevernog dijela Jadrana su 15,5 i 15,16 °C (slika 2a). Razlike srednjih vrijednosti su zanemarive, ali značajna je razlika u maksimalnim zabilježenim temperaturama od 25 i 30 °C. Medijani koncentracije otopljenog kisika su 5,87 i 5,57 mg  $\text{DO}_2 \text{ l}^{-1}$ , slika 2b. Medijani koncentracije otopljenog kisika je nešto veći u hrvatskom dijelu Jadrana, ali značajna je razlika u velikoj učestalosti ekstremno niskih i visokih koncentracija otopljenog kisika u cjelokupnom sjevernom Jadranu.



Slika 2 – Komparativni prikaz podataka za hrvatski (A) i sjeverni Jadran (Italija i Slovenija, B) tijekom razdoblja od 1965. do 2015. godine. Podatci su: temperatura (2a), otopljeni kisik (2b),  $\text{PO}_4^{3-}$  (2c),  $\text{NH}_3$  (2d),  $\text{Si-SiO}_4$  (2e) i ukupni anorganski dušik (2f).

Fig. 2 – Comparative depiction of the data in Croatia Adriatic Sea (A), and Slovenia and Italy (B) measured during the period of 1965 to 2015. The data are temperature (2a), dissolved oxygen (2b),  $\text{PO}_4^{3-}$  (2c),  $\text{NH}_3$  (2d),  $\text{Si-SiO}_4$  (2e), and total inorganic nitrogen (2f).

Medijani koncentracija  $\text{PO}_4^{3-}$  su 0,11 i 0,47  $\mu\text{mol l}^{-1}$ , slika 2c; medijani koncentracije amonijaka  $\text{NH}_3$  su 1,29 i 0,56  $\mu\text{mol l}^{-1}$ , slika 2d; koncentracije otopljenog silicija su 1,95 i 2,96  $\mu\text{mol l}^{-1}$ , slika 2e; i ukupnog otopljenog anorganskog dušika 3,0 i 2,87  $\mu\text{mol l}^{-1}$ , slika 2f. Za sve značajke može se utvrditi mala razlika u medijanima i/ili srednjim vrijednostima, ali je značajna razlika u učestalosti ekstremnih niskih i visokih vrijednosti u usporedbi s hrvatskim dijelom Jadrana.

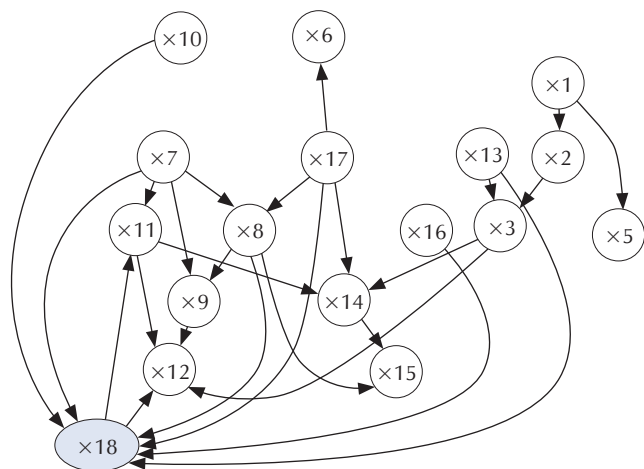
### 3. Modeliranje

Matematička analiza i modeliranje biosfere mora iznimno je složen, ali jedan od najvažnijih ciljeva za razumijevanje, predviđanje i donošenje dugoročnih politika održivosti okoliša. Biosfera mora je otvoreni, stohastički i nelinearni sustav s velikim brojem varijabli stanja od kojih mnoge

nisu poznate i/ili mjerene. Interakcija abiotičkih i bioloških varijabli je nelinearna, a razdiobe gustoće vjerojatnosti najčešće nisu Gaussove i/ili nisu poznate. Međusobni utjecaj dinamičkih veličina stanja je posljedica brojnih povratnih sprega kojima se održava život u biosferi. Matematički dinamički modeli najčešće su izraženi statističkim odnosima (korelacijama) između vremenskih nizova podataka i često uz primjenu modela neuronskih mreža i algoritama za duboko učenje (engl. *deep learning*). Ovaj rad ima cilj odrediti kauzalne odnose između abiotičkih i bioloških parametara za koje je pretpostavljeno da su vremenski neovisni.

Zakonitosti kauzalnosti između varijabli biosfere i fizikalnih veličina posljedice su kemijskih i biokemijskih procesa i ne mijenjaju se pod utjecajem okoline. Strukturni kauzalni model (SCM) matematički je model odnosa kauzalnih zavisnosti i prikazuje se kao usmjereni aciklički graf (DAG). On je osnova Bayesova kauzalnog modela (engl. *Causal Bayes Network*, CBN). Određivanje hipoteze same struktu-





Slika 3 – Bayesova mreža strukturalnog kauzalnog modela (SCM) ekosfere sjevernog Jadrana. Čvorovi mreže su: ×1 = geografska dužina, ×2 = geografska širina, ×3 = dubina dna, ×4 = datum, ×5 = dubina uzimanja uzorka, ×6 = transparentnost, ×7 = temperatura, ×8 = salinitet, ×9 = gustoća, ×10 = pH, ×11 = otopljeni kisik, ×12 = ravnotežna koncentracija otopljenog kisika %, ×13 = koncentracija  $\text{NH}_3$ , ×14 = koncentracija  $\text{NO}_3^-$ , ×15 = koncentracija ukupnog anorganskog dušika, ×16 = koncentracija  $\text{PO}_4^{3-}$ , ×17 = koncentracija Si, ×18 = koncentracija klorofila A.

Fig. 3 – Bayes network of the structural causal model (SCM) of the ecosystem in northern Adriatic Sea. The network nodes are: ×1 = latitude, ×2 = longitude, ×3 = sea depth, ×4 = date of measurement, ×5 = depth of sampling, ×6 = turbidity, ×7 = temperature, ×8 = salinity, ×9 = density, ×10 = pH, ×11 = dissolved oxygen, ×12 = % oxygen equilibria; ×13 =  $\text{NH}_3$  concentration, ×14 =  $\text{NO}_3^-$  concentration, ×15 = concentration of total inorganic nitrogen, ×16 =  $\text{PO}_4^{3-}$  concentration, ×17 = Si concentration, ×18 = chlorophyll A concentration.

re modela moguće je statističkim testiranjem uvjetne nezavisnosti ili optimiranjem Bayesova informacijskog kriterija (BIC). Zbog pretpostavke o nelinearnoj zavisnosti i nepoznatoj razdiobi vjerojatnosti, primijenjen je neparametarski Hilbert-Schmidt uvjetni test nezavisnosti (HSC1).<sup>18-20</sup> Na slici 3 dan je prikaz dobivenog SCM modela s abiotičkim varijablama ( $X$ ) i biološkom varijablom klorofil ( $Y$ ). Osnova evaluacije razdiobe vjerojatnosti varijabli DAG mreže je Markovljevo svojstvo kojim je razdioba vjerojatnosti varijable (čvora mreže) zavisna samo o razdiobama roditeljskih (engl. *parental*) varijabli (čvorova), jedn. (1). Na primjer, zajednička razdioba vjerojatnosti za koncentraciju klorofila,  $Y$  biološki parametar je slučajna varijabla  $X_{18}$ , određena je razdiobama vjerojatnosti "parental" varijabli od  $X_1$  do  $X_N$ , prema izrazu (1):

$$P(Y, X_1, X_2, \dots, X_N) = P(Y | \text{par}(Y)) \prod_{i=1}^{i=N} P(X_i | \text{par}(X_i)) \quad (1)$$

Model prikazan na slici 3 predviđa da je koncentracija klorofila  $X_{18}$  kauzalno određena "roditeljskim" varijablama: temperatura ( $X_7$ ), salinitet ( $X_8$ ), pH ( $X_{10}$ ),  $\text{NH}_3$  ( $X_{13}$ ) i  $\text{PO}_4^-$  ( $X_{16}$ ). Prema modelu, nezavisne egzogene varijable kojima okolina utječe na biosferu su: temperatura ( $X_7$ ), koncentracija silicija ( $X_{17}$ ),  $\text{PO}_4^-$  ( $X_{16}$ ) i  $\text{NH}_3$  ( $X_{13}$ ). Iako su to

veličine stanja sustava mjerene u biosferi, izravna su posljedica temperature atmosfere i ulaznih tokova otpadnih tvari iz okoliša (gradovi, rijeke, turizam itd.) Za određivanje funkcionalne zavisnosti biološke varijable o pojedinoj abiotičkoj varijabli potrebno je iz modela ukloniti utjecaj interferirajućih (engl. *confounding*) varijabli. Analizom mogućih puteva od kauzalne varijable ( $X$ ) i posljedice ( $Y$ ) primjenjuje se postupak razdvajanja (engl. *d-separation*), da bi se postigla nepristrana procjena.<sup>12</sup> Razdvajanjem dobiveni graf definira podskup  $Z$  varijabli  $X$  kojima je spriječena neposredna interferencija, ali i utjecaj povratnim kontra kauzalnim putem. Kauzalni nepristrani učinak  $Y$  po volji determiniranom varijablom  $X$  izražava se kao funkcija  $Y(\text{do}(x))$ , odnosno kao marginalna razdioba:

$$P(Y | \text{do}(X = x)) = \int P(Y(x, Z = z)) dP(z) \quad (2)$$

Za numeričko određivanje funkcije (2) moguće je primijeniti podatke i model strojnog učenja (neuronska mreža i/ili stabla odlučivanja) i rezultate prikazati grafički kao parcijalne zavisnosti (engl. *partial dependence plot*):

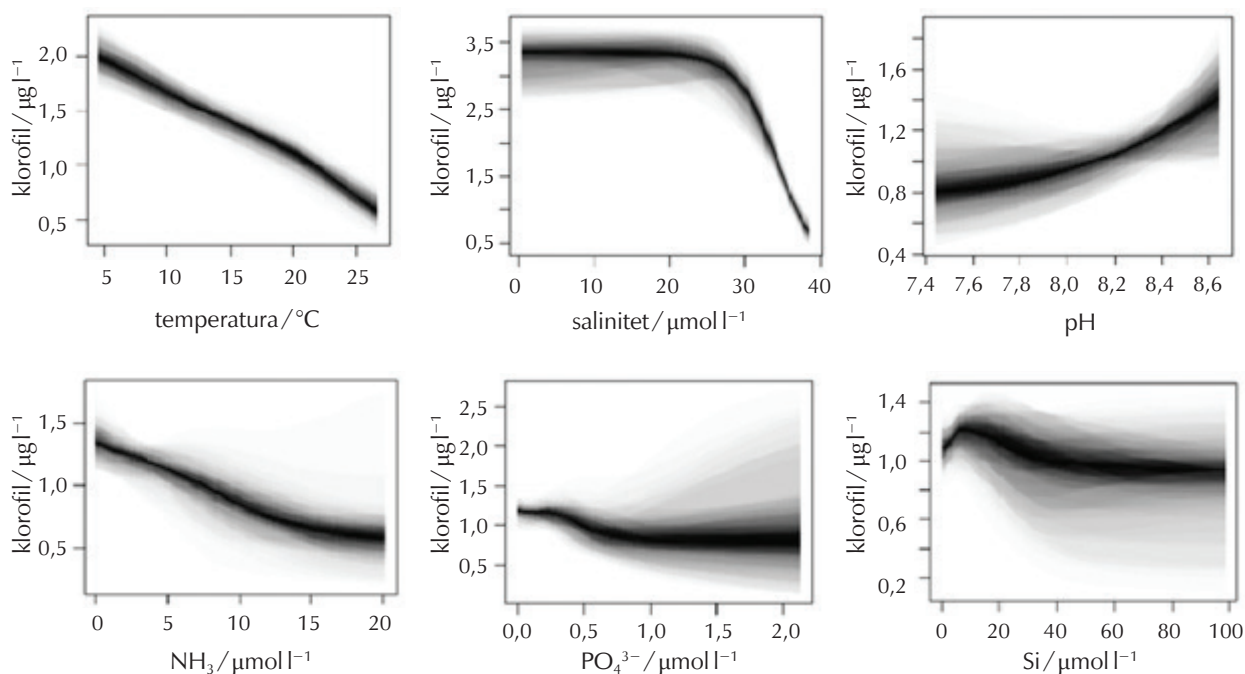
$$Y(x) = \frac{1}{N} \sum_{k=1}^N \text{BNN}(x, Z_k) \quad (3)$$

Na slikama 4.1 – 4.6 prikazane su kauzalne zavisnosti koncentracije klorofila o neposrednim ("roditeljskim") varijablama. Primijenjeni su modeli s Bayesovom neuronskom mrežom s jednim unutarnjim slojem neurona.<sup>22,23</sup> Model Bayesove mreže (slika 3) pokazuje da su koncentracija klorofila i temperatura neposredni kauzalni uzroci koji određuju koncentraciju otopljenog kisika. Kauzalnost je modelirana kao Bayesova NN mreža (jedn. (3)), a predikcije modela prikazane su na slici 5.

Kao pokazatelji biološkog stanja sjevernog Jadrana, osim praćenja koncentracije klorofila, određivane su količine fitoplanktona i zooplanktona. S ciljem predviđanja utjecaja abiotičkih varijabli na biološko stanje mora primijenjeni su modeli sa stablima odlučivanja. Pojedina stabla odlučivanja u procesu učenja optimirana su aproksimacijama maksimalnih gradijenata. Za statističku evaluaciju pouzdanosti prediktivne točnosti modela eksperimentalni podatci podijeljeni su slučajnim izborom u pet podskupina od kojih svaka podskupina služi kao novi podatci za procjenu pogreške. Rezultati su dani u tablici 1. Srednje relativne pogreške su u granicama od 7 do 19 %. Najmanja pogreška 6,5 % je za predikciju koncentracije klorofila.

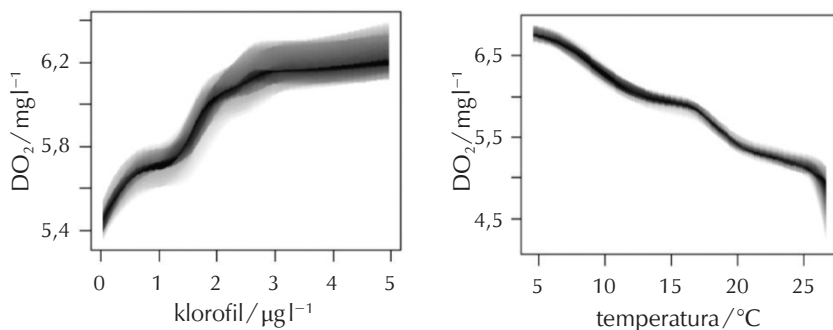
## 4. Rasprava

Matematičko modeliranje biosfere mora konceptualno je i sadržajno iznimno složen zadatak. Sa sustavskog gledišta biosfera mora je stohastički, dinamički i otvoren sustav. Brojne fizikalnokemijske i biološke veličine stanja su u međusobnoj interakciji čije zakonitosti su samo djelomično poznate. Za složenost sustava potreban je velik broj podataka i dugoročno praćenje, da bi se primjenom modela umjetne inteligencije razvili prediktivni modeli. Za upotrebu modela za razumijevanje i donošenje potencijalnih odluka o mjerama tehnoloških intervencija potrebno je iz



Slika 4 – Funkcionalne zavisnosti klorofila o neposrednim kauzalnim varijablama određene Bayesovim neuronskim mrežama s d-razdvojenim modelom s usmjerenim acikličkim grafom (DAG)

Fig. 4 – Functional dependencies of chlorophyll concentration on the parental (direct) variables evaluated by Bayes neural networks with d-separated DAG network



Slika 5 – Kauzalna funkcionalna zavisnost koncentracije otopljenog kisika  $\text{DO}_2$  o koncentraciji klorofila i temperaturi

Fig. 5 – Causal functional dependency of dissolved oxygen concentration  $\text{DO}_2$  on chlorophyll concentration and temperature

Tablica 1 – Ključne varijable (var. 1 – 5) i srednje relativne pogreške e% predikcije bioloških deskriptora u hrvatskom Jadranu primjenom modela stabala odlučivanja

Table 1 – Key variables (var. 1-5) and mean relative errors for prediction of biological descriptors in Croatian Adriatic Sea evaluated by random forest of decision trees

Bio-deskriptor	var. 1	var. 2	var. 3	var. 4	var. 5	e%
klorofil	kisik %	$\text{DO}_2$	salinitet	prozirnost	pH	6,5
feopigmenti	$\text{DO}_2$	silicij	geo. širina	dubina	salinitet	17,4
diatomeje	pH	$\text{NO}_3$	gustoća	ukupni anorganski N	$\text{NH}_3$	18,8
dinoflagelat	gustoća	geo. dužina	dubina	pH	$\text{DO}_2$	17,4
kokolitofore	salinitet	dubina dna	temperatura	gustoća	$\text{NH}_3$	12,1

velikog broja podataka "otkriti" kauzalne relacije osnovnih mehanizama interakcije abiotičke i biološke sfere. To je složen problem koji zahtijeva strukturno povezivanje signifikantnih međusobnih interakcija. Primijenjen je kriterij Hilbert-Schmidt nezavisnosti i algoritam permutacija, da se iz podataka otkriju signifikantne međuzavisnosti abiotičkih i bioloških značajki.

Ovdje je primijenjen Bayesov model mreže za usredotočeni i vremenski invarijantni kauzalni model interakcije fizikalno-kemijskih parametara i koncentracije klorofila. Primjenom algoritma d-razdvajanja blokirane su interferencije i kontrakuzalni tokovi utjecaja preostalih faktora. Za koncentraciju klorofila utvrđeni su sljedeći neposredni kauzalni čimbenici: temperatura, salinitet, pH,  $\text{NH}_3$ , i  $\text{PO}_4^{3-}$ . Najveći kauzalni učinak imaju temperatura  $0,07 \mu\text{g l}^{-1} \text{K}^{-1}$  i  $\text{NH}_3$   $0,05 \mu\text{g mmol}^{-1}$ . Funkcije kauzalnih zavisnosti modelirane su kao marginalne razdiobe primjenom Bayesovih neuronskih mreža s jednim unutarnjim slojem. Model kauzalne zavisnosti koncentracije otopljenog kisika pokazuje negativan učinak temperature  $0,05 \text{ mg DO}_2/\text{°C}$  i dominantni pozitivni efekt koncentracije klorofila  $0,2 \text{ mg DO}_2/\mu\text{g}$  klorofila.

Za abiotičku predikciju količine fitoplanktona i zooplanktona u hrvatskom Jadranu primijenjeni su modeli sa stabilima odlučivanja. Modeli predikcije validirani su s peterostrukom slučajnom podjelom podataka i statističkom analizom predikcije s podskupinom podataka izostavljenom u procesu učenja modela. Prosječne relativne pogreške su u granicama od 7 do 18 %.

## 5. Zaključak

Ključni rezultati primjene kauzalne analize velike baze podataka iz projekta "LTER Northern Adriatic Sea" pokazali su funkcionalne zavisnosti koncentracije otopljenog kisika o koncentraciji klorofila i temperaturi te zavisnosti koncentracije klorofila A o abiotičkim parametrima. Temperatura ima najveći neposredni kauzalni učinak na koncentraciju klorofila ( $-0,07 \mu\text{g}$  klorofila  $\text{A}/\text{°C}$ ). Utvrđena je neposredna pozitivna kauzalnost koncentracije klorofila na koncentraciju otopljenog kisika ( $0,2 \text{ mg DO}_2/\mu\text{g}$  klorofila A). Primjena modela umjetne inteligencije (strojnog učenja) sa stabilima odlučivanja omogućuje predikcije pojedinih bioloških parametara s relativnim pogreškama od 6 do 19 % i procjenu ključnih abiotičkih parametara.

Istaknuti su glavni zaključci da je za matematičko modeliranje složenih ekoloških sustava nužno primijeniti pristup modeliranja strukturne kauzalnosti, da bi se spriječili brojni utjecaji interferencije između abiotičkih i bioloških parametara. Modeli s Bayesovim mrežama omogućuju predikciju učinaka potencijalnih mjere zaštite i tehnoloških intervencija sa svrhom očuvanja okoliša i održivosti proizvodnje.

Moguća primjena agnostičkih modela umjetne inteligencije s velikim brojem prediktora bez poznavanja njihove interakcije može imati za posljedicu dobro slaganje modela s podacima upotrijebljenim za razvoj modela (učenje), ali

uz potpuno pogrešne predikcije kad se upotrijebe u svrhu upravljanja procesa.<sup>24</sup>

Vjerodostojnost strukturnih kauzalnih modela (SCM) može se bitno unaprijediti integracijom deduktivnog i induktivnog kauzalnog zaključivanja primjenom "a priori informacije" temeljnim načelima kemijskog inženjersva s "posteriori" informacijama o eksperimentalnim podacima empirijskim modelima (na primjer, prijenosa tvari i energije, kemijske kinetike).

Napredak algoritama umjetne opće inteligencije, UOI (engl. *Artificial General Intelligence*, AGI) od najveće je važnosti za kemijske inženjere, ali i ostale struke za razvoj održivog industrijskog razvoja i regulacije učinaka globalnih klimatskih promjena.

## Literatura References

1. D. L. Chandler, *Why the Mediterranean is a climate change hotspot*, MIT Press, Cambridge, SAD, 2020.
2. J. Cos, F. Doblas-Reyes, M. Jury, R. Marcos, P.-A. Bretonnière, M. Samsó, *The Mediterranean climate change hotspot in the CMIP5 and CMIP6 projections*, *Earth. Syst. Dynam.* **13** (1) (2022) 321–340, doi: <https://doi.org/10.5194/esd-13-321-2022>.
3. V. Noon, C. Rivière, P. Strosser, E. Manea, E. Gissi, A. Barbanti, *Assessing and valuing ecosystem services in the Northern Adriatic*, EU Commission Final report, 2021., Aug.
4. H. K. Solvang, S. Subbey, *An improved methodology for quantifying causality in complex ecological systems*, *PLOS ONE* **14** (6) (2019) e0217195, doi: <https://doi.org/10.1371/journal.pone.0208078>.
5. European Commission, 2021., *Assessing and valuing ecosystem services in the Northern Adriatic*, URL: [https://cinea.ec.europa.eu/publications/study-integrating-ecosystem-based-approach-maritime-spatial-planning-lit-review\\_en](https://cinea.ec.europa.eu/publications/study-integrating-ecosystem-based-approach-maritime-spatial-planning-lit-review_en).
6. RH Ministarstvo gospodarstva i održivog razvoja, *Akcijski program strategije upravljanja morskim okolišem i obalnim područjem*, 2022, URL: [https://mingor.gov.hr/UserDocsImages/Uprava\\_vodnoga\\_gospodarstva\\_i\\_zast\\_mora/Strategija\\_upravljanja\\_morem/Akcijski%20program%20Sustav%20pra%C4%87enja%202021\\_2026.pdf](https://mingor.gov.hr/UserDocsImages/Uprava_vodnoga_gospodarstva_i_zast_mora/Strategija_upravljanja_morem/Akcijski%20program%20Sustav%20pra%C4%87enja%202021_2026.pdf).
7. M. Goodwin, K. T. Halvorsen, L. Jiao, K. M. Knausgård, A. H. Martin, M. Moyano, R. A. Oomen, J. H. Rasmussen, T. K. Sørдалen, S. H. Thorbjørnsen, *Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook*, *ICES J. Marine Sci.* **79** (2) (2022) 319–336, doi: <https://doi.org/10.1093/icesjms/fsab255>.
8. S. Christan, E. Herve, N. Lecomte, *Applications for deep learning in ecology*, *Methods Ecol. Evol.* **10** (10) (2019) 1632–1644, doi: <https://doi.org/10.1111/2041-210X.13256>.
9. T. Bolton, L. Zanna, *Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization*, *JAMES* **11** (1) (2019) 376–399, doi: <https://doi.org/10.1029/2018MS001472>.
10. T. de Wolff, H. Carrillo, L. Martí, N. Sanchez-Pi, I. Chile, *Assessing Physics Informed Neural Networks in Ocean Modeling and Climate Change Applications*, 2021, url: <https://hal.inria.fr/hal-03262684>.

11. S. Franceschini, L. Tancioni, M. Lorenzoni, F. Mattei, M. Scardi, An ecologically constrained procedure for sensitivity analysis of Artificial Neural Networks and other empirical models, *PLOS ONE* **14** (1) (2019) e0211445, doi: <https://doi.org/10.1371/journal.pone.0211445>.
12. J. Pearl, D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Penguin Random House, London, UK, 2018.
13. EU Open Marine Science, URL: <https://mes.openaire.eu/>.
14. F. Aciri, M. Bastianini, F. Bernardi Aubry, E. Camatti, A. Boldrin, C. Bergami, D. Cassin, A. De Lazzari, S. Finotto, A. Minelli, A. Oggioni, M. Pansera, A. Sarretta, G. Socal, A. Pugnetti, A long-term (1965–2015) ecological marine database from the LTER-Italy Northern Adriatic Sea site: plankton and oceanographic observations, *Earth. Syst. Sci. Data* **12** (2020) 215–230, doi: <https://doi.org/10.5194/essd-12-215-2020>.
15. A. Minelli, A. Sarretta, A. Oggioni, C. Bergami, M. Bastianini, F. Bernardi Aubry, E. Camatti, A. Pugnetti, Opening Marine Long-Term Ecological Science: Lesson Learned From the LTER-Italy Site Northern Adriatic Sea, *Front. Mar. Sci.* **8** (6) (2021) 659522, doi: <https://doi.org/10.3389/fmars.2021.659522>.
16. LTER database Dynamic Ecological Information Management System, URL: <https://deims.org>.
17. R: A language and environment for statistical computing. R Foundation for Statistical Computing, R Core Team, Beč, Austrija, 2020, URL: <https://www.R-project.org/>.
18. R. Herbrich, A. Smola, O. Bousquet, B. Schölkopf, Kernel Methods for Measuring Independence, *JMLR* **6** (2005) 2075–2129, url: <https://www.jmlr.org/papers/volume6/gretton05a/gretton05a.pdf>.
19. P. Verbyla, N. I. Bertille Desgranges, L. Wernisch, Kernel PC Algorithm for Causal Structure Detection, 2017, URL: <https://CRAN.R-project.org/package=kpcalg>.
20. R. Nagarajan, M. Scutari, S. Lèbre, *Bayesian Networks in R*, Springer, New York, 2013, doi: <https://doi.org/10.1007/978-1-4614-6446-4>.
21. M. Scutari, M. Denis, J. B. Bayesian, *Bayesian Networks with Examples in R*, 2. izd., Chapman & Hall/CRC, New York, 2021, doi: <https://doi.org/10.1201/9780429347436>.
22. Q. Zhao, T. Hastie, Causal Interpretations of Black-Box Models, *J. Bus. Econ. Stat.* **39** (1) (2021) 272–281, doi: <https://doi.org/10.1080/07350015.2019.1624293>.
23. URL: <https://cran.r-project.org/web/packages/BoomSpikeSlab/index.html>.
24. Ž. Kurtanjek, Kad zaključivanje matematičkim modelom može biti pogrešno: Primjer protočnog kemijskog reaktora PKR, *Kem. Ind.* **70** (11-12) (2021) 739–741.



## SUMMARY

### The Causal Ecological Model Based on EU Project Data “LTER Northern Adriatic Sea”

Želimir Kurtanjek

The aim of this work was to show possibilities of applied artificial intelligence methodologies and structural causal modelling (“Structural Causal Model”, SCM) with the object of gaining a scientific level contribution to the determination of functional causal dependencies in complex ecological systems. In this work, applied was SCM for the determination of dependencies of chlorophyll concentration on physical and chemical parameters in the northern Adriatic Sea during the period 1965 to 2015. The experimental data are the outcome of the long-term and extensive investigation as a part of the EU project “LTER Northern Adriatic Sea”, and are freely available within the EU Open Science policy. The data are a “Big Data” base with 108 687 samples and 43 descriptors. Proposed is a mathematical model with Bayes network (BN) as a directed acyclic graph (DAG). The model structure was determined by the Hamilton-Schmidt conditional independence test with a significance level of  $\alpha = 0.05$ . The SCM model shows that the direct causal variables for chlorophyll concentration are: temperature, salinity, pH, and concentrations of nitrogen, phosphor, and silica. The BN model was adjusted according to d-separation with the objective to block confounding and contra-causal back door interference. The functions of causal dependencies were determined as the marginal distributions with Bayes network models with a single interior layer for interpolation. The most important causal effect was due to temperature ( $-0.07 \mu\text{g chlorophyll A}/^\circ\text{C}$ ). The model predicted reversed positive causality between chlorophyll concentration and dissolved oxygen ( $0.2 \text{ mg DO}_2/\mu\text{g chlorophyll A}$ ). Also evaluated was nonparametric comparative analysis of chlorophyll and abiotic parameters between Croatian and northern Adriatic Sea (Slovenia and Italy). The comparison was based on median metrics to avoid the pronounced influence of outliers due to hydrodynamic effects. The median concentration of dissolved oxygen in Croatian Adriatic was  $5.8 \text{ mg O}_2/\text{l}$ , while in Slovenian and Italian  $5.5 \text{ mg O}_2/\text{l}$ , and the median temperature was  $T = 14.6 \text{ }^\circ\text{C}$  compared to  $T = 15.1 \text{ }^\circ\text{C}$ . There is a significant difference in the abundance of dinoflagellates in Croatia  $3 \text{ cell/l}$ , while in Slovenia and Italian  $5 \text{ cells/l}$ . The difference is more pronounced by the number and values of “hot spots” outliers. The difference between chlorophyll concentrations is not significant ( $0.65$  and  $0.90 \mu\text{g l}^{-1}$ ); however, the difference in the distribution of the outliers is significant with more frequent and bigger outliers in Italian and Slovenian Adriatic. Also observed was a significant difference in  $\text{SiO}_4$  distribution, with higher concentrations in the western Adriatic. The random forest RF decision tree models are applied for the development of the predictive models of biological parameters based on abiotic data. The RF models are validated by 5-fold cross-validation. The models have out-of-box mean relative errors of 6.5 % for chlorophyll, photopigment 17.4 %; diatoms 18.8 %; dinoflagellate 17.4 %; and 12.1 % for coccolithophores. For each predictive model determined are the first five most important predictors accounting for 95 % of importance.

#### Keywords

SCM causality, LTER marine biosphere, Adriatic Sea

University of Zagreb, Faculty of Food  
Technology and Biotechnology, Pierrotijeva 6,  
10 000 Zagreb, Croatia

Original scientific paper  
Received March 31, 2022  
Accepted June 27, 2022