

PREDICTION OF NITRATE CONCENTRATION IN GROUNDWATER USING MLR AND AGRICULTURAL INDICATORS

Lidija Stamenković

The Academy of Applied Technical and Preschool Studies, Serbia,
lidija.stamenkovic@akademijanis.edu.rs

Tijana Milanović

The Academy of Applied Technical and Preschool Studies, Serbia, tijana.milanovic@akademijanis.edu.rs

Gordana Bogdanović

The Academy of Applied Technical and Preschool Studies, Serbia,
gordana.bogdanovic@akademijanis.edu.rs

Abstract: Environmental pollution results in the degradation of all segments of nature. Groundwater, as one of the significant sources of drinking water at the global level, is increasingly exposed to various types of pollutants that are mainly the result of anthropogenic action. Nitrate are one of the pollutants that can persist for a long time in groundwater and their concentrations can reach high levels. Namely, the fact that groundwater is used for various purposes such as: irrigation, food production, drinking, special attention is focused on monitoring the concentration of nitrate in groundwater. In this sense, EU legislation also prescribed a standard for drinking water in terms of nitrate concentration in order to avoid possible threats to human health. The main sources of nitrate in groundwater are agricultural activities, which primarily involve the use of chemical fertilizer, animal farms, waste water from industries that produce chemical fertilizer. In this paper, the prediction of average annual concentrations of nitrate in groundwater at the national level was performed. Bearing in mind the mentioned sources of nitrate in groundwater, a total of seven agricultural indicators were selected in this paper, which were assumed to contribute to higher nitrate concentrations, namely: the area used for agriculture, meat production on farms, the balance of indicators - which represents the total potential environmental risk caused by excess or deficit of N and P in agricultural soils, area used for organic agricultural production, consumption of chemical fertilizer, sale of pesticides and percentage of population associated with at least secondary wastewater treatment. The prediction was made using multiple linear regression, where agricultural indicators were used as input parameters, i.e. independent variables, a total of seven, while the dependent variable was the average annual concentration of nitrate in groundwater at the national level. In this paper, five EU countries were selected for nitrate prediction: Belgium, Bulgaria, the Czech Republic, Germany and France. For the development of the MLR (Multiple linear regression) model, available data from 2011 to 2015 were used for all eight variables. Before developing the model, a correlation analysis was first performed to determine whether the selected independent variables were correlated with the dependent variable. The results of the correlation analysis showed that there is a significant correlation between the independent and dependent variables. The developed MLR model showed good prediction results with a value of coefficient of determination R^2 - 0.96. The results of the one year prediction by the MLR model showed satisfactory results, the biggest deviations between the measured and model predicted values are in the case of Germany and France. Based on the obtained results, it can be concluded that the MLR model can be applied as one of the alternatives in the assessment of nitrate concentration in groundwater.

Keywords: nitrate in groundwater, agricultural indicators, groundwater pollution, MLR

PREDVIĐANJE KONCENTRACIJE NITRATA U PODZEMNIM VODAMA PRIMENOM MLR I POLJOPRIVREDNIH INDIKATORA

Lidija Stamenković

Akademija tehničko-vaspitačkih strukovnih studija, Srbija, lidija.stamenkovic@akademijanis.edu.rs

Tijana Milanović

Akademija tehničko-vaspitačkih strukovnih studija, Srbija, tijana.milanovic@akademijanis.edu.rs

Gordana Bogdanović

Akademija tehničko-vaspitačkih strukovnih studija, Srbija, gordana.bogdanovic@akademijanis.edu.rs

Rezime: Zagađenje životne sredine za posledicu ima degradaciju svih segmenata prirode. Podzemne vode, kao jedan od značajnih izvora pijaće vode na globalnom nivou, sve više su izložene različitim vrstama polutanata koji su uglavnom posledica antropogenog delovanja. Jedan od polutanata koji mogu dugo da opstaju u podzemnim vodama

i da njihove koncentracije dostignu visok nivo jesu nitrati. Imajući u vidu činjenicu da se podzemne vode koriste za različite namene kao što su: navodnjavanje, proizvodnja hrane, za piće, posebna pažnja je usmerena na praćenje koncentracije nitrata u podzemnim vodama. U tom smislu je i zakonska regulativa EU propisala standard za vodu za piće u pogledu koncentracije nitrata kako bi se izbegle moguće pretnje po ljudsko zdravlje. Kao glavni izvori nitrata u podzemnim vodama mogu se navesti poljoprivredne aktivnosti koje pre svega podrazumevaju upotrebu veštačkih đubriva, farme životinja, otpadne vode iz industrija koje se bave proizvodnjom veštačkih đubriva. U ovom radu vršeno je predviđanje srednjih godišnjih koncentracija nitrata u podzemnim vodama na nacionalnom nivou. Imajući u vidu pomenute izvore nitrata u podzemnim vodama, u ovom radu je odabrano ukupno sedam poljoprivrednih indikatora za koje se pretpostavljalo da doprinose povišenim koncentracijama nitrata i to: površina koja se koristi za poljoprivrednu proizvodnju, proizvodnja mesa na farmama, bilans indikatora-koji predstavlja ukupnu potencijalnu opasnost po životnu sredinu uzrokovanog viškom ili deficitom N i P u poljoprivrednim zemljištima, površina koja se koristi za organsku poljoprivrednu proizvodnju, potrošnja veštačkih đubriva, prodaja pesticida i procenat stanovništva povezan sa najmanje sekundarnim tretmanom otpadnih voda. Predviđanje je vršeno primenom višestruke linearne regresije, pri čemu su kao ulazni parametri tj.nezavisne promenljive korišćeni poljoprivredni indikatori, ukupno sedam, dok je zavisna promenljiva srednja godišnja koncentracija nitrata u podzemnim vodama na nacionalnom nivou. U ovom radu za predikciju nitrata odabrano je pet država EU i to: Belgija, Bugarska, Češka, Nemačka i Francuska. Za razvoj modela MLR (Multiple linear regression) korišćeni su dostupni podaci od 2011. do 2015. godine za svih osam promenljivih. Pre razvoja modela najpre je urađena korelaciona analiza da bi se utvrdilo da li su odabrane nezavisne promenljive u korelaciji sa zavisnom promenljivom. Rezultati korelacione analize su pokazali da postoji značajna korelacija između nezavisnih i zavisne promenljive. Razvijeni MLR model je pokazao veoma dobre rezultate predviđanja sa vrednošću koeficijenta determinacije R^2 - 0.96. Rezultati jednogodišnjeg predviđanja MLR modelom pokazali su zadovoljavajuće rezultate, najveća odstupanja između izmerenih i modelom predviđenih vrednosti su u slučaju Nemačke i Francuske. Na osnovu dobijenih rezultata može se zaključiti da se MLR model može primeniti kao jedna od alternativa u proceni koncentracije nitrata u podzemnim vodama.

Ključne reči: nitrati u podzemnim vodama, poljoprivredni indikatori, zagađenje podzemnih voda, MLR

1. UVOD

Podzemne vode, kao jedan od najznačajnijih vodenih resursa, od velikog su značaja za društvo. Ekološki pritisci kojima su izložene, usled antropogenog delovanja i klimatskih promena, predstavljaju značajan problem savremenog društva (He et al., 2022; Snousy et al., 2022). Primena hemijskih sredstava u poljoprivrednoj proizvodnji uslovlili su značajno zagađenje podzemnih voda na globalnom novou. Ubrzani razvoj industrije i stalni rast poljoprivredne proizvodnje uslovio je povećanje koncentracije nitrata u vodenim sistemima (Richa et al., 2022). U tom smislu nitrati predstavljaju jedan od najčešćih zagađivača podzemnih voda, što predstavlja značajan ekološki problem. Posledice visokih koncentracija nitrata su pogoršanje kvaliteta podzemnih voda i efekti koje te koncentracije nitrata imaju na ljudsko zdravlje (Karlović et al., 2022; Musacchio et al., 2021). Povećane koncentracije nitrata u vodi uzročnik su brojnih zdravstvenih problema poput methemoglobinemije kod dece, disfunkcija štitne žlezde kod dece i trudnica i različitih tipova karcinoma kod odraslih (Ahada & Suthar, 2018; Ceballos et al., 2021). Suočeni sa ovim problemom, brojne organizacije na globalnom nivou donele su set regulatornih mera u vidu Direktiva, kojima je ograničava koncentracija nitrata u vodenim sistemima. Imajući u vidu da je glavni uzrok zagađenja podzemnih voda nitratima sektor poljoprivrede Evropska unija je 1991. usvojila nitratnu Direktivu (Council Directive 91/676/EEC of 12 December 1991 Concerning the Protection of Waters against Pollution Caused by Nitrates from Agricultural Sources, 1991). Cilj Direktive, koja promoviše dobru poljoprivrednu praksu, je zaštita i redukcija daljeg zagađivanja podzemnih i pitkih voda nitratima koje je uzrokovano poljoprivrednim aktivnostima. Kako se Direktiva odnosi na zemlje članice Evropske unije, od njih se traži da pronađu načine i osmisle programe radi sprečavanja ove vrste zagađenja u osetljivim područjima. Pored toga, a imajući u vidu posledice koje nitrati imaju na ljudsko zdravlje Evropska unija i Svetska zdravstvena organizacija su definisale graničnu vrednost za nitrate (NO_3^-) u vodi za piće koja iznosi 50mg/L (Council Directive 98/83/EC of 3 November 1998 on the Quality of Water Intended for Human Consumption, 1998).

Prema postojećoj metodologiji monitoring podzemnih voda vrši se pomoću državnih mernih stanica u kojima se vrši uzorkovanje, fizičko-hemijska i hemijska analiza uzoraka voda prema unapred utvrđenim metodama. Nitrati u podzemnim vodama mereni su miligramima po litru (NO_3 mg/L). Podaci se obrađuju i daju srednje godišnje vrednosti za svaku državu EU koja je uključena u sistem monitoringa (EUROSTAT, 2022). Pored toga, kao alternativa postojećim metodologijama, postoje brojna istraživanja koja primenjuju različite tehnike mašinskog učenja radi određivanja koncentracije nitrata u podzemnim vodama za različite regione (Gholami & Booi, 2022; He

et al., 2022; Knoll et al., 2019; Wagh et al., 2018). Rezultati pomenutih studija su pokazali da modeli pokazuju veoma dobre performanse u predviđanju koncentracije nitrata u vodi.

Predmet ovog rada je bio da se razvije MLR (Multiple linear regression) model za predviđanje srednjih godišnjih koncentracija nitrata u podzemnim vodama. U svrhu razvoja MLR modela korišćeni su poljoprivredni indikatori za koje se pretpostavilo da imaju najveći uticaj na kontaminaciju podzemnih voda nitratima. Osnovni cilj ovog istraživanja je bio da se utvrdi da li MLR model zasnovan na poljoprivrednim indikatorima kao nezavisnim promenljivama može da se koristi za predviđanje srednjih godišnjih koncentracija nitrata u podzemnim vodama.

2. MATERIJALI I METODE

Izvor podataka

U ovom radu korišćeni su dostupni podaci kancelarije Evropske unije zadužene za statistiku na nivou cele EU (*Home - Eurostat*). Za razvoj modela korišćeni su podaci za pet država EU za period od 2011. do 2015. godine i to Belgije, Bugarske, Češke, Nemačke i Francuske. U radu je modelovana srednja godišnja koncentracija nitrata u podzemnim vodama izražena u mg/L. Na osnovu literaturnih podataka o sektorima koji najviše doprinose kontaminaciji podzemnih voda nitratima, u ovom radu odabrano je sedam nezavisnih promenljivih: površina koja se koristi za poljoprivrednu (PKP), proizvodnja mesa na farmama (PM), bilans indikatora-koji predstavlja ukupnu potencijalnu opasnost po životnu sredinu uzrokovano viškom ili deficitom N i P u poljoprivrednim zemljištima (BI), površina koja se koristi za organsku poljoprivrednu proizvodnju (PO), potrošnja veštačkih đubriva (PVD), prodaja pesticida (PP) i procenat stanovništva povezan sa najmanje sekundarnim tretmanom otpadnih voda (POV). Deskriptivna statistika svih promenljivih data je u Tabeli 1.

Tabela 1. Statistička analiza svih promenljivih za razvoj MLR modela

Indikator	Jedinica	Sr. vr.	St. Dev.	Min. vr.	Maks. vr.
PKP	1000 ha	11124.03	10630.859	1330.88	29146.02
PM	1000 t	1820.6704	2037.6422	48.22	5598
BI	kg po ha	70.388	42.698744	11.9	143.2
PO	%	5.8224	4.2200289	0.5	13.68
PVD	t	932999.96	863407.26	143317	2332390
PP	kg	25105405	26881286	5663374	74873617
POV	%	77.9208	13.17319	53.63	95.77
Nitrati	mg/L	23.7812	4.6389477	18.15	31.26

Višestruka linearna regresija (MLR)

MLR predstavlja skup statističkih tehnika koje su našle značajnu primenu za predviđanje u svim oblastima, a poslednjih godina i u oblasti zaštite životne sredine (Stamenkovic et al., 2015). MLR tehnika zasnovana je na pronalaženju linearne kombinacije koju čine skup nezavisnih promenljivih na osnovu koje se procenjuje nepoznata regresijska funkcija zavisne promenljive (Giacomino et al., 2011). Ova linearna zavisnost nezavisnih i zavisne promenljive prikazuje se putem matematičke jednačine 1:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \beta_a X_a + \varepsilon_i \quad (1)$$

Gde: Y predstavlja zavisnu promenljivu, X predstavlja nezavisne promenljive, β_0 predstavlja konstantu, $\beta_1, \beta_2, \dots, \beta_a$ predstavlja koeficijente nezavisne promenljive i ε predstavlja residual.

Postoje različite vrste linearne regresije. U zavisnosti od načina na koji se uvode nezavisne promenljive u jednačinu, postoje tri osnovne vrste i to: standardna MLR gde se sve nezavisne promenljive unose u jednačinu istovremeno, postepena MLR gde se nezavisne promenljive u jednačinu unose na osnovu primene određenih statističkih kriterijuma i hijerarhijska MLR kod koje se nezavisne promenljive u jednačinu uvode jedna po jedna, ili po skupovima promenljivih (Stamenković, 2016). U ovom radu je korišćena standardna MLR kod koje su sve nezavisne promenljive unete u jednačinu istovremeno. Za razvoj MLR modela u ovom radu korišćen je statistički softverski paket IBM SPSS 19 Statistics.

3. REZULTATI I DISKUSIJA

Pre samog razvoja MLR modela za predviđanje koncentracije nitrata u podzemnim vodama, urađena je korelaciona analiza kako bi se utvrdilo da li postoji i kolika je korelacija između promenljivih uključenih u razvoj modela. Rezultati korelacione analize prikazani su u Tabeli 2.

Tabela 2. Koerelaciona analiza

	PKP	PM	BI	PO	PVĐ	PP	POV	Nitrati
PKP	1							
PM	0.5445324	1						
BI	-0.4832377	0.0920329	1					
PO	-0.2447577	-0.0638833	0.3577447	1				
PVĐ	0.9719445	0.7079183	-0.3632021	-0.153308	1			
PP	0.9691323	0.670091	-0.2835628	-0.1506007	0.9814249	1		
POV	0.3781252	0.7964786	0.4595728	0.4357177	0.5479134	0.5704311	1	
Nitrati	-0.5223519	0.0042142	0.3178811	-0.535895	-0.4796231	-0.4855928	-0.263865	1

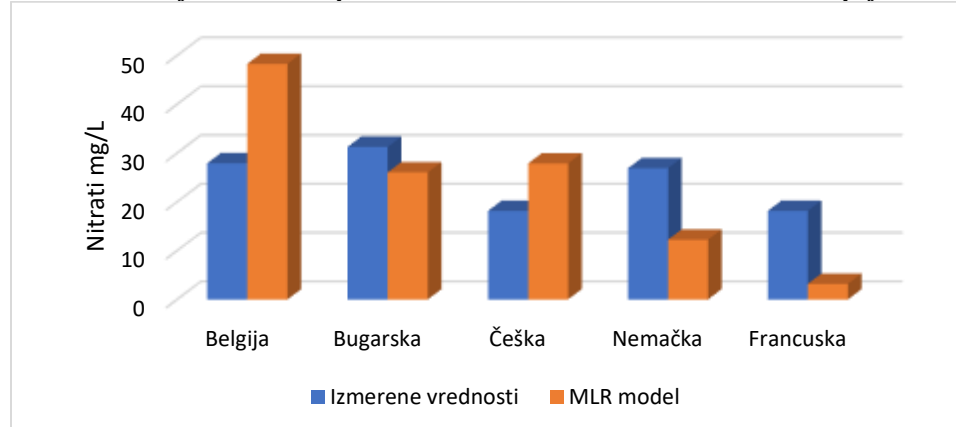
Na osnovu dobijenih rezultata korelacione analize može se videti da su skoro svih sedam nezavisnih promenljivih u dobroj korelaciji sa zavisnom promenljivom osim PM i POV. Takođe se može uočiti da su ova dva indikatora u visokim korelacijama sa ostalim indikatorima što može uticati na rezultat MLR modela. Naime, iako MLR daje dobre rezultate predviđanja prema literaturnim podacima, poteškoće u rešavanju postavljenog problema mogu se javiti ukoliko su ulazne promenljive u značajnoj korelaciji (Abdul-Wahab et al., 2005).

Nakon sprovedene korelacione analize pristupilo se razvoju MLR modela. Dobijena vrednost R^2 kreiranog modela je 0.96 što ukazuje na dobre rezultate kreiranog modela. Dobijeni MLR koeficijenti predstavljeni su jednačinom 2:

$$\text{Nitrati} = 33.09195 - 0.3745\text{PKP} + 1.178733\text{PM} + 0.024485\text{BI} - 0.70563\text{PO} - 0.00208\text{PVĐ} + 8.22\text{E} - 05\text{PP} - 0.06839\text{POV} \quad (2)$$

Rezultati jednogodišnjeg predviđanja koncentracije nitrata u podzemnim vodama za odabrane države u ovom radu prikazani su na Slici 1. Kao što se može videti na Slici 1. rezultati predviđanja MLR modela su zadovoljavajući. U slučaju Bugarske i Češke model pokazuje dobro slaganje sa izmerenim vrednostima nitrata dok su u slučaju Belgije, Nemačke i Francuske nešto lošiji rezultati predviđanja. Problem odstupanja između izmerenih i predviđenih vrednosti za pomenute države može se pripisati kvalitetu ulaznih podataka korišćenih prilikom razvoja MLR modela. Naime, podaci za neke od odabranih nezavisnih promenljivih su kod pomenutih država procenjeni tj. nisu dobijeni primenom standardnih metoda što može usloviti ovakva odstupanja između izmerenih i modelovanih vrednosti. Pored toga, kao što je korelaciona analiza pokazala, postoji značajna korelacija između nezavisnih promenljivih što je uticalo na performanse kreiranog modela.

Slika 1. Poređenje izmerenih i predviđenih vrednosti nitrata za svaku državu pojedinačno



4. ZAKLJUČAK

Nitrati u podzemnim vodama predstavljaju jedan od značajnih problema sa kojima se proteklih godina društvo suočava. Imajući u vidu porast poljoprivrednih aktivnosti, koje se smatraju jednim od bitnijih izvora zagađenja ovim polutantom, kontaminacija podzemnih voda nitratima zaokuplja sve veću pažnju istraživača koji se bave upravo ovom temom. U tom smislu se i ovaj rad bavi upravo ovim problemom. U radu je predstavljen razvoj modela zasnovan na višestrukoj linearnoj regresiji radi predviđanja srednjih godišnjih koncentracija nitrata u podzemnim vodama. Za razvoj modela korišćeni su dostupni podaci za pet EU država od 2011 do 2015 godine. Kao nezavisne promenljive za razvoj modela korišćeni su indikatori poljoprivrede za koje se smatralo da doprinose povećanoj koncentraciji nitrata. Dobijeni rezultati kreiranog MLR modela $R^2=0.96$ ukazuju na dobre performanse u predviđanju. Dobijeni rezultati su takođe pokazali da se poljoprivredni indikatori koji se redovno prate na nivou EU, mogu uspešno primeniti za predviđanje nitrata u podzemnim vodama. Kako bi se postigli bolji rezultati u predviđanju, buduća istraživanja bi mogla ići u pravcu primene nekih novih tehnika za odabir najznačajnijih ulaznih promenljivih kako bi se izbegao problem korelacije nezavisnih promenljivih.

LITERATURA

- Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, 20(10), 1263–1271. <https://doi.org/10.1016/j.envsoft.2004.09.001>
- Ahada, C. P. S., & Suthar, S. (2018). Groundwater nitrate contamination and associated human health risk assessment in southern districts of Punjab, India. *Environmental Science and Pollution Research*, 25(25), 25336–25347. <https://doi.org/10.1007/s11356-018-2581-2>
- Ceballos, E., Dubny, S., Othax, N., Zabala, M. E., & Peluso, F. (2021). Assessment of Human Health Risk of Chromium and Nitrate Pollution in Groundwater and Soil of the Matanza-Riachuelo River Basin, Argentina. *Exposure and Health*, 13(3), 323–336. <https://doi.org/10.1007/s12403-021-00386-9>
- Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources, CONSIL, 375 OJ L (1991). <http://data.europa.eu/eli/dir/1991/676/oj/eng>
- Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption, CONSIL, 330 OJ L (1998). <http://data.europa.eu/eli/dir/1998/83/oj/eng>
- EUROSTAT. (2022). *Nitrate in groundwater* (source: EEA) (sdg_06_40). https://ec.europa.eu/eurostat/cache/metadata/en/sdg_06_40_esmsip2.htm
- Gholami, V., & Booiij, M. J. (2022). Use of machine learning and geographical information system to predict nitrate concentration in an unconfined aquifer in Iran. *Journal of Cleaner Production*, 360, 131847. <https://doi.org/10.1016/j.jclepro.2022.131847>
- Giacomino, A., Abollino, O., Malandrino, M., & Mentasti, E. (2011). The role of chemometrics in single and sequential extraction assays: A Review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques. *Analytica Chimica Acta*, 688(2), 122–139. <https://doi.org/10.1016/j.aca.2010.12.028>
- He, S., Wu, J., Wang, D., & He, X. (2022). Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. *Chemosphere*, 290, 133388. <https://doi.org/10.1016/j.chemosphere.2021.133388>
- Home—Eurostat. Retrieved September 16, 2022, from <https://ec.europa.eu/eurostat>
- Karlović, I., Posavec, K., Larva, O., & Marković, T. (2022). Numerical groundwater flow and nitrate transport assessment in alluvial aquifer of Varaždin region, NW Croatia. *Journal of Hydrology: Regional Studies*, 41, 101084. <https://doi.org/10.1016/j.ejrh.2022.101084>
- Knoll, L., Breuer, L., & Bach, M. (2019). Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Science of The Total Environment*, 668, 1317–1327. <https://doi.org/10.1016/j.scitotenv.2019.03.045>
- Musacchio, A., Mas-Pla, J., Soana, E., Re, V., & Sacchi, E. (2021). Governance and groundwater modelling: Hints to boost the implementation of the EU Nitrate Directive. The Lombardy Plain case, N Italy. *The Science of the Total Environment*, 782, 146800. <https://doi.org/10.1016/j.scitotenv.2021.146800>
- Richa, A., Touil, S., & Fizir, M. (2022). Recent advances in the source identification and remediation techniques of nitrate contaminated groundwater: A review. *Journal of Environmental Management*, 316, 115265. <https://doi.org/10.1016/j.jenvman.2022.115265>
- Snousy, M. G., Wu, J., Su, F., Abdelhalim, A., & Ismail, E. (2022). Groundwater Quality and Its Regulating Geochemical Processes in Assiut Province, Egypt. *Exposure and Health*, 14(2), 305–323. <https://doi.org/10.1007/s12403-021-00445-1>

- Stamenkovic, L., Antanasijevic, D., Ristic, M., Peric-Grujic, A., & Pocajt, V. (2015). Modeling of methane emissions using artificial neural network approach. *Journal of the Serbian Chemical Society*, 80(3), 421–433. <https://doi.org/10.2298/JSC020414110S>
- Stamenković, L. J. (2016). Predviđanje emisije gasovitih zagađujućih materija na nacionalnom nivou primenom modela zasnovanih na veštačkim neuronskim mrežama. *Универзитет у Београду*. <https://nardus.mpn.gov.rs/handle/123456789/7650>
- Wagh, V., Panaskar, D., Muley, A., Mukate, S., & Gaikwad, S. (2018). Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. *Groundwater for Sustainable Development*, 7, 436–445. <https://doi.org/10.1016/j.gsd.2017.12.012>