

PREDICTION OF PM2.5 EMISSIONS USING ANN

Lidija Stamenković

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

Ljiljana Đorđević

The Academy of Applied Technical and Preschool Studies, Serbia,
ljiljana.djordjevic@akademijanis.edu.rs

Gordana Bogdanović

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

Tijana Milanović

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

Abstract: Rapid economic and industrial development and the emission of various pollutants into the atmosphere result in increasing pressure on the environment. Particulate matter, PM2.5 represent one of the most significant atmospheric pollutants and one of the most significant indicators of air quality. In this sense, the monitoring of this pollutant is extremely important, especially if we take into account the negative consequences on human health caused by long-term exposure to this pollutant. The subject of this paper is the development of a model based on artificial neural networks (ANN) for the prediction of PM2.5 emissions at the national level. For the development of the model, a standard three-layer neural network with one input layer, one hidden layer and one output layer was used in this work. For the development of the ANN model, available data for 23 countries of Europe for the period from 2010 to 2019 were used. One of the most important segments in the development of the ANN model is the selection of the most important input parameters, i.e. those input parameters that have the greatest influence on the output. As industry and traffic have a significant effect on the emissions of PM2.5, in this paper 7 inputs were selected that were considered to contribute to the emissions of this pollutant. Industrial, economic and traffic indicators were used as input data for the development of the model: Energy productivity, Final energy consumption-total, Final energy consumption - transport sector, Final energy consumption in households, Modal split of passenger transport, Primary energy consumption and Gross domestic product. To evaluate the performance of the created model, the statistical performance indicator coefficient of determination R^2 was used. The obtained results of the created ANN model show a very good agreement between the measured and the ANN model predicted PM2.5 emissions values, with the value of the coefficient of determination $R=0.869$. During the development of the ANN model, the significance of individual inputs was also assessed. It turned out that Final energy consumption in households and Gross domestic product have the greatest influence on PM2.5 emissions. Other inputs have a slightly smaller influence, but certainly significant. Bearing in mind the importance of PM2.5 as an atmospheric pollutant, the existence of alternative models, in addition to the existing ones, for estimating the emissions of this pollutant is of great importance. Based on the results of the model, it can be said that ANN can serve as an alternative model for estimating PM2.5 emissions.

Keywords: PM2.5, emissions prediction, air pollution, MLP, ANN

PREDVIĐANJE EMISIJE PM2.5 PRIMENOM VNM

Lidija Stamenković

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

Ljiljana Đorđević

Akademija tehničko-vaspitačkih strukovnih studija, Srbija, ljiljana.djordjevic@akademijanis.edu.rs

Gordana Bogdanović

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

Tijana Milanović

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

Rezime: Brz ekonomski i industrijski razvoj i emisija zagađujućih materija u atmosferu rezultiraju sve većim pritiskom na životnu sredinu. Suspendovane čestice PM2.5 predstavljaju jedan od najznačajnijih zagađivača

atmosfera i jedan od najznačajnijih pokazatelja kvaliteta vazduha. U tom smislu, monitoring ove zagađujuće materije je izuzetno važan, posebno ako se imaju u vidu negativne posledice po zdravlje ljudi izazvane dugotrajnom izlaganju ovom zagađivaču. Predmet ovog rada je razvoj modela zasnovanog na veštačkim neuronskim mrežama (VNM) za predviđanje emisije PM_{2.5} na nacionalnom nivou. Za razvoj modela u radu je korišćena standardna troslojna neuronska mreža sa jednim ulaznim slojem, jednim skrivenim slojem i jednim izlaznim slojem neurona. Za razvoj ANN modela korišćeni su dostupni podaci za 23 zemlje Evrope za period od 2010. do 2019. godine. Jedan od najvažnijih segmenata u razvoju ANN modela je izbor najvažnijih ulaznih parametara, odnosno onih ulaznih parametara koji imaju najveći uticaj na izlaz. Kako industrija i saobraćaj imaju značajan uticaj na emisiju PM_{2.5}, u ovom radu je odabrano 7 ulaznih parametara za koje se smatra da doprinose emisiji ovog zagađivača. Kao ulazni podaci za razvoj modela korišćeni su industrijski, ekonomski i saobraćajni indikatori: Energetska produktivnost, Finalna potrošnja energije-ukupna, Finalna potrošnja energije - sektor saobraćaja, Finalna potrošnja energije u domaćinstvima, Udeo putničkog saobraćaja u ukupnom transportnom učinku, Potrošnja primarne energije i Bruto domaći proizvod. Za procenu performansi kreiranog modela korišćen je statistički indikator performansi koeficijent determinacije R². Dobijeni rezultati kreiranog VNM modela pokazuju veoma dobro slaganje izmerenih i VNM modelom predviđenih vrednosti emisije PM 2.5, sa vrednošću koeficijenta determinacije R²-0.869. Tokom razvoja ANN modela procenjen je i značaj pojedinačnih inputa. Pokazalo se da finalna potrošnja energije u domaćinstvima i bruto domaći proizvod imaju najveći uticaj na emisiju PM_{2.5}. Ostali inputi imaju nešto manji uticaj, ali svakako značajan. Imajući u vidu značaj PM_{2.5} kao zagađivača atmosfere, postojanje alternativnih modela, pored postojećih, za procenu emisije ovog zagađivača je od velikog značaja. Na osnovu rezultata modela, može se reći da ANN može poslužiti kao alternativni model za procenu emisije PM_{2.5}.

ključne reči: PM_{2.5}, predviđanje emisije, zagađenje vazduha, MLP, VNM

1. UVOD

Ubrzani industrijski i ekonomski razvoj i primena ekološki nepovoljnih tehnoloških postupaka proizvodnje u velikoj meri su ispoljili negativne efekte na životnu sredinu. Ti negativni efekti odražavaju se na sve segmente životne sredine, a posebno na kvalitet vazduha i njegovo zagađenje (F. Wu et al., 2023; Yang et al., 2022). Jedna od glavnih komponenta zagađivača vazduha su i suspendovane čestice čiji je prečnik manji od 2.5μm, koje sadrže toksične i štetne supstance i koje mogu dugo ostati u vazduhu. Prisustvo PM_{2.5} u vazduhu može imati dugoročne posledice po ljudsko zdravlje, budući da dugotrajno izlaganje visokim koncentracijama PM_{2.5} može dovesti do pojave respiratornih i kardiovaskularnih bolesti, što svakako predstavlja ozbiljnu pretnju životu i zdravlju ljudi (Xia et al., 2022; Yang et al., 2023). Visoke koncentracije PM_{2.5} imaju višestruki negativni efekat na životnu sredinu. Što je sadržaj PM_{2.5} u atmosferi veći to i ozbiljnost zagađenja vazduha raste, te se u tom smislu može reći da se koncentracija PM_{2.5} smatra ključnim indikatorom kvaliteta vazduha (Dong et al., 2022). Kao najznačajniji antropogeni izvori ovih polutanata smatraju se energetska sektor, stambeni sektor i drumski saobraćaj. Zakonska regulativa kako na globalnom, tako i na nacionalnom nivou, usmerava i obavezuje razvojne politike država potpisnica međunarodnih sporazuma o redukciji emisije zagađujućih materija u vazduhu, da vrše redovan monitoring emisije i koncentracije svih zagađujućih materija u vazduhu kao i PM_{2.5}. U tom smislu se podnose godišnji izveštaji o emisijama PM_{2.5} na nacionalnom nivou. Neizvesnost u proceni emisije dodatno komplikuje zagađenje izazvano suspendovanim česticama (Su et al., 2016). Postojanje alternativnih modela za procenu emisije PM_{2.5} na nacionalnom nivou, pored već postojećih, je od krucijalnog značaja kako bi se dobila tačna predviđanja emisije PM_{2.5} koja mogu biti od pomoći donosiocima odluka u planiranju razvoja u raznim delatnostima.

Kao jedan od mogućih alternativnih modela za predviđanje emisije PM_{2.5} u ovom radu se predlaže model zasnovan na veštačkim neuronskim mrežama (VNM). Veštačke neuronske mreže, pristup zasnovan na mašinskom učenju, poslednjih godina sve više je zastupljen za predviđanje različitih indikatora kvaliteta životne sredine. Modeli zasnovani na VNM su pokazali veoma dobre performanse u predviđanju emisije zagađujućih materija u vazduhu (Antanasijević et al., 2018; Stamenković et al., 2016, 2017). Pored toga, literaturni podaci pokazuju da su modeli VNM veoma uspešno primenjeni za predviđanje koncentracije PM_{2.5} za različite regione i različite vremenske periode (Bera et al., 2021; Kim et al., 2023; Masood & Ahmad, 2020; Park et al., 2023; Peng et al., 2022; Sobri et al., 2021; S. Wu & Li, 2022).

2. MATERIJALI I METODE

Izvor podataka

U ovom radu korišćeni su dostupni podaci kancelarije Evropske unije zadužene za statistiku na nivou cele EU i OECD (*Database - Eurostat; Environment - OECD Data*). Za razvoj modela korišćeni su podaci za 23 države Evrope za period od 2010. do 2019. godine. U radu je modelovana godišnja emisija PM_{2.5} u vazduhu izražena u kg po stanovniku (kg pc). Na osnovu literaturnih podataka o sektorima koji najviše doprinose emisiji PM_{2.5} u vazduhu,

u ovom radu odabrano je sedam nezavisnih promenljivih: Energetska produktivnost (EP), Finalna potrošnja energije-ukupna (FEP), Finalna potrošnja energije - sektor saobraćaja (FPT), Finalna potrošnja energije u domaćinstvima (FPD), Udeo putničkog saobraćaja u ukupnom transportnom učinku (UT), Potrošnja primarne energije (PEP) i Bruto domaći proizvod (BDP). Deskriptivna statistika svih promenljivih data je u Tabeli 1.

Tabela 1. Statistička analiza svih promenljivih za razvoj VNM modela

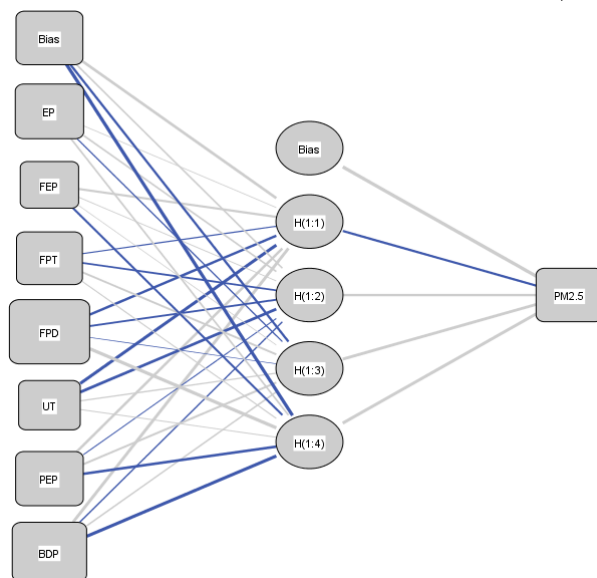
Indikator	EP	FEP	FPT	FPD	UT	PEP	BDP	PM2.5
Jedinica	euro kgoe	toe pc	toe pc	kgoe	%	toe pc	euro pc	kg pc
Sr.vr.	7.208	0.772	0.653	622.10	81.955	3.849	27445.739	3.587
St.dev.	3.250	0.817	0.168	225.90	6.265	3.167	14810.968	2.010
Minimum	1.760	0.241	0.201	248.00	59.200	1.340	8000.000	0.876
Maximum	19.640	4.436	1.061	1433.00	91.900	18.690	69560.000	11.274

Razvoj ANN modela

Veštačke neuronske mreže predstavljaju matematičke modele kod kojih se na pojednostavljen način oponaša način funkcionisanja biološkog nervnog sistema. Obrada informacija u VNM vrši se uz pomoć paralelno distribuirane arhitekture sastavljene od nekoliko jednostavnih procesora koji su međusobno povezani. Povezivanjem više nezavisnih jednostavnih procesora (veštačkih neurona) u slojevima formira se VNM. Strukturu VNM čine tri glavna sloja: ulazni sloj, skriveni sloj i izlazni sloj neurona. U izgradnji modela zasnovanom na VNM proces obuke, ili treninga mreže je od izuzetnog značaja. U ovom procesu mreži se prezentuju ulazni i izlazni podaci, pri čemu mreža, uočavajući veze među podacima teži da dobije odgovarajuću izlaznu vrednost, odnosno vrednost koja je približna očekivanoj vrednosti. To se postiže promenom težinskih koeficijenata kroz algoritam obuke mreže do trenutka dok mreža ne nauči da funkcioniše na odgovarajući način. Na osnovu literaturnih podataka proces razvoja modela zasnovanog na VNM podeljen je u nekoliko osnovnih koraka: prikupljanje podataka, izbor ulaznih promenljivih (prediktora), podela podataka, izbor parametara mreže, trening, ili obuka mreže i validacija modela (Cabaneros et al., 2019).

Nakon odabira ulaznih promenljivih, sledeći korak u razvoju modela VNM je bio podela podataka za razvoj mreže, pri čemu je 68,6% prezentovanih podataka korišćeno za obuku mreže, dok je 31.4% podataka korišćeno za testiranje mreže u fazi obuke. Izbor parametra neuronske mreže: arhitektura, broj skrivenih slojeva neurona, aktivacione funkcije je bio sledeći korak u razvoju modela. U ovom radu korišćena je standardna troslojna neuronska mreža (multilayer perceptron-MLP) sa jednim skrivenim slojem neurona. Broj neurona u skrivenom sloju je 4. Aktivaciona funkcija skrivenog sloja je *Hyperbolic tangent* i *Identity* je aktivaciona funkcija izlaznog sloja. Struktura modela veštačke neuronske mreže prikazana je na Slici 1.

Slika 1. Struktura modela veštačke neuronske mreže (VNM)



Procena performansi kreiranog modela može se vršiti primenom različitih statističkih indikatora performansi. U ovom radu je za procenu performansi kreiranog VNM modela korišćen statistički indikator koeficijent determinacije R^2 . Vrednost ovog indikatora se određuje prema Formuli 1. Vrednosti koeficijenta determinacije se kreću od 0 do 1, i što je vrednost bliža 1 to model daje bolje rezultate predviđanja.

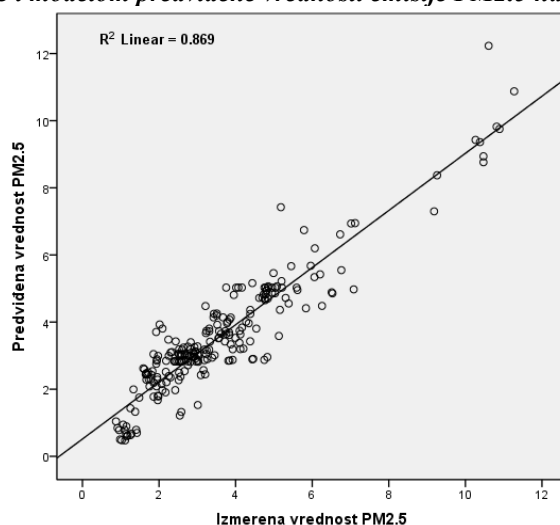
$$R^2 = \frac{[\sum(C_p - \bar{C}_p)(C_o - \bar{C}_o)]}{\sum(C_o - \bar{C}_o)^2 \sum(C_p - \bar{C}_p)^2} \quad (1)$$

Gde: C_p predstavlja vrednost parametra predviđenu modelom, dok C_o predstavlja izmerenu (aktuelnu) vrednost posmatrane promenljive.

3. REZULTATI I DISKUSIJA

Na Slici 2 prikazani su rezultati VNM modela na trening podacima. Kao što se može videti model je dao veoma dobre rezultate predviđanja emisije PM2.5 sa vrednošću koeficijenta determinacije 0.869.

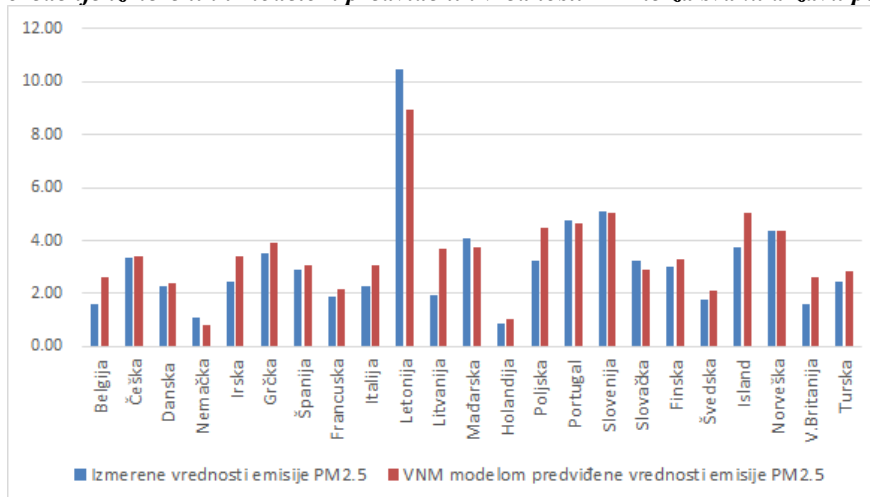
Slika 2. Aktuelne i modelom predviđene vrednosti emisije PM2.5 na trening podacima



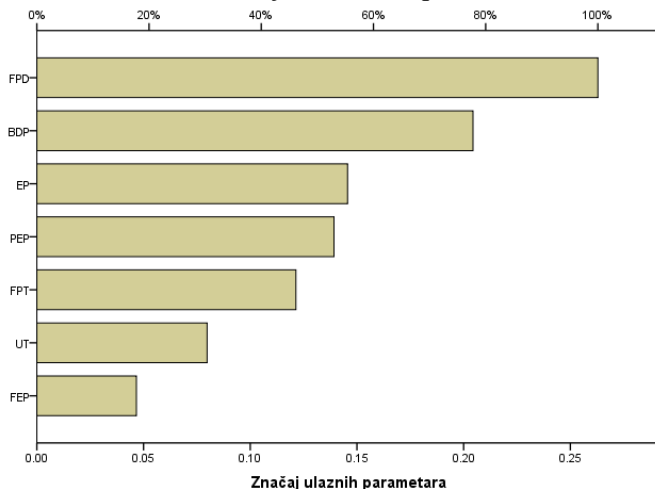
Za validaciju VNM modela mreži su prezentovani potpuno novi podaci. Mogućnosti jednogodišnjeg predviđanja emisije PM2.5 za svaku državu uključenu u razvoj modela prikazan je na Slici 3. Kao što se može videti, VNM model je dao veoma dobra slaganja između izmerenih i modelom predviđenih vrednosti za svaku državu. Nešto veća odstupanja između izmerenih i VNM modelom predviđenih vrednosti mogu se uočiti u slučaju Letonije, Litvanije, Poljske i Islanda. Nešto lošija predviđanja u slučaju ovih država mogu se objasniti time da vrednosti pojedinih ulaznih parametara u trening podacima predstavljaju procenjene a ne stvarne vrednosti.

Tokom razvoja VNM modela izvršena je i analiza značajnosti ulaznih parametara za model. Rezultati značajnosti ulaznih parametara prikazana je na Slici 4. Kao što se može videti najveći uticaj na emisiju PM2.5 imaju parametri FPD i BDP, preko 70%. Parametri EP i PEP imaju nešto manji, ali značajan uticaj, preko 50%, dok ostali parametri imaju značajnost ispod 50%.

Slika 3. Poređenje izmerenih i modelom predviđenih vrednosti PM2.5 za svaku državu pojedinačno



Slika 4. Značajnost ulaznih parametara



4. ZAKLJUČAK

U ovom radu primenjen je model zasnovan na veštačkim neuronskim mrežama za predviđanje godišnje emisije PM2.5 u vazduhu. Kao ulazni parametri za razvoj modela korišćeni su ekonomski, industrijski i indikatori saobraćaja. Za razvoj modela odabrano je sedam ulaznih parametara za period od 2010. do 2019. godine za 23 države Evrope. Dobijeni rezultati su pokazali da model zasnovan na veštačkim neuronskim mrežama pokazuje veoma dobre rezultate predviđanja sa vrednošću koeficijenta determinacije 0.869 na trening datasetu. Rezultati predviđanja modela na podacima za validaciju pokazuje takođe veoma dobro slaganje između izmerenih i modelom predviđenih vrednosti. Dobijeni rezultati ukazuju na zaključak da su inicijalno odabrane ulazne promenljive adekvatne za razvoj modela i da posmatrani indikatori imaju značajan uticaj na emisiju PM2.5. Takođe, na osnovu rezultata se može zaključiti da VNM model može biti uspešno primenjen za predikciju emisije PM2.5 na nacionalnom nivou. Pravci daljeg istraživanja biće usmereni ka poboljšanju performansi kreiranog VNM modela kroz primenu neke od tehnika za odabir najznačajnijih ulaznih parametara.

LITERATURA

Antanasijević, D., Pocajt, V., Perić-Grujić, A., & Ristić, M. (2018). Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutant emissions. *Atmospheric Pollution Research*, 9(2), 388–397. <https://doi.org/10.1016/j.apr.2017.10.011>

- Bera, B., Bhattacharjee, S., Sengupta, N., & Saha, S. (2021). PM_{2.5} concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models. *Environmental Challenges*, 4, 100155. <https://doi.org/10.1016/j.envc.2021.100155>
- Cabaneros, S. M., Calautit, J. K., & Hughes, B. R. (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software*, 119, 285–304. <https://doi.org/10.1016/j.envsoft.2019.06.014>
- Database—Eurostat. Retrieved 18 November 2022, from <https://ec.europa.eu/eurostat/data/database>
- Dong, L., Hua, P., Gui, D., & Zhang, J. (2022). Extraction of multi-scale features enhances the deep learning-based daily PM_{2.5} forecasting in cities. *Chemosphere*, 308, 136252. <https://doi.org/10.1016/j.chemosphere.2022.136252>
- Environment—OECD Data. Retrieved 26 June 2023, from The OECD website: <http://data.oecd.org/environment.htm>
- Kim, Y., Park, S.-B., Lee, S., & Park, Y.-K. (2023). Comparison of PM_{2.5} prediction performance of the three deep learning models: A case study of Seoul, Daejeon, and Busan. *Journal of Industrial and Engineering Chemistry*, 120, 159–169. <https://doi.org/10.1016/j.jiec.2022.12.022>
- Masood, A., & Ahmad, K. (2020). A model for particulate matter (PM_{2.5}) prediction for Delhi based on machine learning approaches. *Procedia Computer Science*, 167, 2101–2110. <https://doi.org/10.1016/j.procs.2020.03.258>
- Park, S.-Y., Woo, S.-H., & Lim, C. (2023). Predicting PM₁₀ and PM_{2.5} concentration in container ports: A deep learning approach. *Transportation Research Part D: Transport and Environment*, 115, 103601. <https://doi.org/10.1016/j.trd.2022.103601>
- Peng, J., Han, H., Yi, Y., Huang, H., & Xie, L. (2022). Machine learning and deep learning modeling and simulation for predicting PM_{2.5} concentrations. *Chemosphere*, 308, 136353. <https://doi.org/10.1016/j.chemosphere.2022.136353>
- Sobri, N. M., Yaacob, W. F. W., Ismail, N. A., Malik, M. A. A., Rahman, R. A., Baser, N. A., & Sukhairi, S. A. M. (2021). Predicting Particulate Matter (PM_{2.5}) in Malaysia using Multiple Linear Regression and Artificial Neural Network. *Journal of Physics: Conference Series*, 2084(1), 012010. <https://doi.org/10.1088/1742-6596/2084/1/012010>
- Stamenković, L. J., Antanasijević, D. Z., Ristić, M. Đ., Perić-Grujić, A. A., & Pocajt, V. V. (2016). Estimation of NMVOC emissions using artificial neural networks and economical and sustainability indicators as inputs. *Environmental Science and Pollution Research*, 23(11), 10753–10762. <https://doi.org/10.1007/s11356-016-6279-z>
- Stamenković, L. J., Antanasijević, D. Z., Ristić, M. Đ., Perić-Grujić, A. A., & Pocajt, V. V. (2017). Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model. *Air Quality, Atmosphere & Health*, 10(1), 15–23. <https://doi.org/10.1007/s11869-016-0403-6>
- Su, X., Gough, W., & Shen, Q. (2016, June 20). *Correlation of PM_{2.5} and meteorological variables in Ontario cities: Statistical downscaling method coupled with artificial neural network*. 215–226. Crete, Greece. <https://doi.org/10.2495/AIR160201>
- Wu, F., Min, P., Jin, Y., Zhang, K., Liu, H., Zhao, J., & Li, D. (2023). A novel hybrid model for hourly PM_{2.5} prediction considering air pollution factors, meteorological parameters and GNSS-ZTD. *Environmental Modelling & Software*, 167, 105780. <https://doi.org/10.1016/j.envsoft.2023.105780>
- Wu, S., & Li, H. (2022). Prediction of PM_{2.5} concentration in urban agglomeration of China by hybrid network model. *Journal of Cleaner Production*, 374, 133968. <https://doi.org/10.1016/j.jclepro.2022.133968>
- Xia, X., Yao, L., Lu, J., Liu, Y., Jing, W., & Li, Y. (2022). Observed causative impact of fine particulate matter on acute upper respiratory disease: A comparative study in two typical cities in China. *Environmental Science and Pollution Research International*, 29(8), 11185–11195. <https://doi.org/10.1007/s11356-021-16450-5>
- Yang, H., Liu, Z., & Li, G. (2022). A new hybrid optimization prediction model for PM_{2.5} concentration considering other air pollutants and meteorological conditions. *Chemosphere*, 307, 135798. <https://doi.org/10.1016/j.chemosphere.2022.135798>
- Yang, H., Wang, W., & Li, G. (2023). Prediction method of PM_{2.5} concentration based on decomposition and integration. *Measurement*, 216, 112954. <https://doi.org/10.1016/j.measurement.2023.112954>