EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology 2021, 22(1), 109–116
https://doi.org/10.12912/27197050/132088
ISSN 2719-7050, License CC-BY 4.0

Received: 2020.12.11
Accepted: 2020.12.28
Published: 2021.01.05

Estimation of Water Disinfection by Using Data Mining

Esra'a Bashayreh[1], Ahmad Manasrah[2], Shahnaz Alkhalil[2], Eman Abdelhafez[3*]

[1] Department of Electrical Engineering, Communication and Computer, Al-Zaytoonah University of Jordan, Amman, Jordan

[2] Department of Mechanical Engineering, Al-Zaytoonah University of Jordan, Amman, Jordan

[3] Department of Alternative Energy Technology, Al-Zaytoonah University of Jordan, Amman, Jordan

* Corresponding author's email: eman.abdelhafez@zuj.edu.jo

ABSTRACT

In this study, the Artificial Neural Network (ANN) models and multiple linear regression techniques were used to estimate the relation between the concentration of total coliform, E. coli and Pseudomonas in the wastewater and the input variables. Two techniques were used to achieve this objective. The first is a classical technique with multiple linear regression models, while the second one is data mining with two types of ANN (Multilayer Perceptron (MLP) and Radial Basis Function (RBF). The work was conducted using (SPSS) software. The obtained estimated results were verified against the measured data and it was found that data mining by using the RBF model has good ability to recognize the relation between the input and output variables, while the statistical error analysis showed the accuracy of data mining by using the RBF model is acceptable. On the other hand, the obtained results indicate that MLP and multiple linear regression have the least ability for estimating the concentration of total coliform, E. coli and pseudomonas in wastewater.

Keywords: data mining; water disinfection, regression, artificial neural network.

INTRODUCTION

Freshwater availability in countries is mainly based on precipitation as well as on water flowing from one region to another. However, the amount of available freshwater is less than 0.05%; the UN estimates that over 30 countries in the world lack freshwater resources (Barlow et al., 2017). Even though the average amount of freshwater available per person reaches over 100,000 m^3 per year in few humid and sparsely populated areas, it could be less than 50 m^3 in some parts in the Middle East (World Water Assessment Programme, 2006). In fact, a recent study showed that almost every nation experiences some sort of a vulnerability regarding the freshwater supplies and the most vulnerable is Jordan in the Middle East region (Padowski et al., 2017). Therefore, protecting water sources and improving the quality of drinking water is becoming more important every year especially in remote and rural areas .Multiple water disinfection techniques have been implemented for this purpose, like chlorination and water boiling; in addition, the solar water disinfection (SODIS) technique has been used, which is considered an easy, low cost and environmentally sustainable solution for water purification at a household level (Burhan 2015).

The solar water disinfection (SODIS) technique has gained a lot of attention in the past decade since the method is simple, cost effective, and can be implemented at households (Stubbé et al., 2016). The concept of the technique depends on solar radiation where the ultraviolet rays (UV) produce a synergistic effect that inactivates and kills microbial pathogens in contaminated water (Boyle et al., 2008; Castro-Alférez et al., 2017). Three to five hours of sunlight exposure with solar radiations above 500 W/m^2 is enough to eliminate pathogens (Meierhofer and Wegelin, 2002) given little to no water turbidity and favorable ambient temperatures (Oates et al., 2003).

SODIS has been investigated in previous studies with several modifications based on the conditions of the experiments and the nature of infected water. Exposing infected water to direct sunlight contributed to a significant reduction in the growth of microbes and viruses in general, as shown in the work of (Lawrie et al., 2015; Islam et al., 2015; Polo et al., 2015; Aboushi et al., 2019). Other researchers even tried to enhance the process of SODIS by adding iron oxide (Shekoohiyan et al., 2019), for example or using polymer bags (Gutiérrez-Alfaro et al., 2017). However, there are many factors that might affect the presence and inactivation of microbial pathogens using this method. For example, the effect of water temperature has been investigated in (Sift et al., 2016) and (Vivar et al., 2017). Water turbidity also may impact the inactivation process (Keogh et al., 2017; Dawney et al., 2012) as well as the level of pH in water (Sahel et al., 2017). Even the duration of light exposure may affect the inactivation process in solar disinfection (Giannakis et al., 2015).

Therefore, predicting the presence of microbial pathogens using the data-driven techniques can enhance the disinfection process of water through cutting costs and optimizing the previously stated variables. For instance, a previous study used three methods based on a data-mining technique to predict the levels of chlorine in water in order to optimize the costs of adding chlorine without sacrificing the water quality (Zounemat-Kermani et al., 2018). The results showed that the multi-layer perceptron neural network method (MLPNN) yielded the greatest accuracy compared to other methods. Other studies also investigated the concentration of chlorine in water using artificial neural networks (ANN) and genetic algorithms (Wu et al., 2014; Hernández Cervantes et al., 2015).

However, when it comes to SODIS, the sunlight exposure period plays a major part in the inactivation process of bacteria (Shekoohiyan et al., 2019). Therefore, mathematical models were developed to estimate the time period needed to kill all microscopic organisms in water. For instance, a previous study introduced a fuzzy rule-based logic model that estimates the sunlight exposure time required to remove all fecal coliforms under different turbidity levels (Haider et al., 2017). The results showed agreement between the predicted and measured values of total coliform. Another study proposed a simple equation that provides the estimated amount of lethal UV dose that is needed for solar water disinfection (Figueredo-Fernández et al., 2017).

There is very little research, however, regarding the estimation of residual microbes in water that is treated with SODIS. A previous study presented this methodology to predict the level of Coliforms and E. coli on tomato fruits and lettuce leaves after the sanitizing process, rather than in water (Keeratipibul et al., 2011). In this paper, multiple regression and Artificial Neural Network (ANN) methods were used to predict the concentrations of total coliform, E. coli and Pseudomonas in the wastewater that is treated with SODIS. The results will help us optimize this disinfection technique by identifying the factors and variables that positively or negatively impact the solar disinfection process.

EXPERIMENT SETUP

BOECO Germany Laboratory glass bottles of 500 ml were used as wastewater containers which in its turn were directed to solar radiation. These containers were installed side by side and their measurements were collected every hour. Thermometers were used for monitoring temperatures.

Total coliform, E. coli and Pseudomonas were tested by means of the IDEXX setup, this technique is considered certificated, rapid, easy, and accurate. In addition, a quality and quantity test was performed (Hamdan and Darabee, 2017).

RESULTS AND DISCUSSION

Multiple linear regression

The regression model resulted from SPSS, time (t), water temperature (T), pH and turbidity (Tr) were used as input variables and the concentration of total coliform, E. coli and pseudomonas in the wastewater were used as the output variables. In total, 48 samples were used to obtain the following linear equations:

$$Total\ Coliform = -318.666 * \mathrm{t} - 117.566 * \mathrm{T} - 321.693 * \mathrm{PH} + 3.99 * \mathrm{Tr} + 7584.166 \quad (1)$$

$$E.coli = 9.186 * \mathrm{t} - 62.132 * \mathrm{T} - 417.009 * \mathrm{PH} + 1.965 * \mathrm{Tr} + 5339.108 \quad (2)$$

$$Pseudomonas = 13.030 * \mathrm{t} - 50.642 * \mathrm{T} - 288.149 * \mathrm{PH} - 0.556 * \mathrm{Tr} + 3978.737 \quad (3)$$

Table 1 represents a summary of the results obtained using this model. As it was shown, the value of R (coefficient of determination) depends strongly on the dependent variable for constant values of time, water temperature, pH and turbidity for the prediction of total coliform and E. coli concentration. On the other hand, the value of R depends weakly on the dependent variable for constant values of time, water temperature, pH and turbidity for the prediction of Pseudomonas concentration. Table 2 shows the relation between the time, water temperature, pH and turbidity as predictors (input) with the concentration of total coliform, E. coli and Pseudomonas as dependent variables.

Artificial neural network model

In this work, two types of Artificial Neural Network (ANN) models were used to estimate the concentration of total coliform, E. coli and Pseudomonas in the wastewater, these models are Multilayer Perceptron (MLP) and Radial Basis Function (RBF). The variables (time, water temperature, pH and turbidity) were the input-variables used in training the ANN models, and the concentrations of total coliform, E. coli and Pseudomonas in the wastewater were used as outputs variables. The obtained results were verified against the multiple regression technique.

Two types of ANN models were built and examined by Statistical Package for the Social Sciences (SPSS) software. The experimental data of previously obtained 48 samples was used as the input of ANN model.

Multilayer Perceptron Model

The Multilayer Perceptron Model (MLP) is a procedure compatible with a particular kind of neural network called a multilayer perceptron which is considered flexible. It uses the feed-forward architecture and can have multiple hidden layers. It is one of the most commonly used neural network architectures. Table 3 shows the case processing summary, Table 4 shows the network information and Table 5 shows the model summary.

Radial Basis Function Model

A Radial Basis Function network is a feed-forward; supervised learning network with only one hidden layer, called radial basis

Table 1. Regression model summary

Model	R	R square	Adjusted R square	Std. error of the estimate
Total coliform	.823^a	.677	.646	503.6373
E. coli	.861^a	.741	.717	126.8169
Pseudomonas	.474^a	.225	.151	324.7226
a – Predictors: (Constant), Turbidity, Time, PH, TEMP.				

Table 2. Coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. e:rror	Beta		
Total coliform	(Constant)	7584.166	4430.029		1.712	.094
	Time	-318.666	135.167	-.345	-2.358	.023
	TEMP	-117.566	37.788	-.597	-3.111	.003
	PH	-321.693	472.822	-.105	-.680	.500
	Turbidity	3.990	5.922	.073	.674	.504
E. coli	(Constant)	5339.108	1115.491		4.786	.000
	Time	9.186	34.035	.035	.270	.789
	TEMP	-62.132	9.515	-1.121	-6.530	.000
	PH	-417.009	119.058	-.485	-3.503	.001
	Turbidity	1.965	1.491	.127	1.318	.195
Pseudomonas	(Constant)	3978.737	2856.283		1.393	.171
	Time	13.030	87.150	.034	.150	.882
	TEMP	-50.642	24.364	-.618	-2.079	.044
	PH	-288.149	304.855	-.227	-.945	.350
	Turbidity	-.556	3.818	-.024	-.146	.885

Table 3. Case processing summary

Specification		N	Percent, %
Sample	Training	31	66.0
	Testing	16	34.0
Valid		47	100.0
Excluded		0	
Total		47	

Table 4. Network information

Input layer	Covariates	1	Time
		2	TEMP
		3	PH
		4	Turbidity
	Number of units[a]		4
	Rescaling method for covariates		Standardized
Hidden layer(s)	Number of hidden layers		1
	Number of units in hidden layer 1[a]		2
	Activation function		Hyperbolic tangent
Output layer	Dependent variables	1	Total Coliform
		2	E. coli
		3	Pseudomonas
	Number of units		3
	Rescaling method for scale dependents		Standardized
	Activation function		Identity
	Error function		Sum of Squares
a – Excluding the bias unit.			

Table 5. MLP model summary

Training	Sum of squares error		15.105
	Average overall relative error		.336
	Relative error for scale dependents	Total coliform	.166
		E. Coli	.137
		Pseudomonas	.703
	Stopping rule used		1 consecutive step(s) with no decrease in error[a]
	Training time		0:00:00.06
Testing	Sum of squares error		2.073
	Average overall relative error		.362
	Relative error for scale dependents	Total coliform	.382
		E. coli	.230
		Pseudomonas	.353
a – Error computations are based on the testing sample.			

function layer. The RBF network can do both prediction and classification exactly the same as to what multi-layer perceptron network can do. However, it can be much faster than the MLP, but it is not as flexible in the types of models it can fit. Table 6 shows the case processing summary, Table 7 shows the network information and Table 8 shows the model summary.

Figures 1 to 3 show the comparison between the obtained experimental data and the estimated power, as mentioned previously. Table 9 summarizes the comparison of performance of the used models based on statistical analysis. Lower

Table 6. Case processing summary

Specification		N	Percent, %
Sample	Training	35	74.5
	Testing	12	25.5
Valid		47	100.0
Excluded		0	
Total		47	

Table 7. Network information

Input layer	Covariates	1	Time
		2	TEMP
		3	PH
		4	Turbidity
	Number of units		4
	Rescaling method for covariates		Standardized
Hidden layer	Number of units		10[a]
	Activation function		Softmax
Output layer	Dependent variables	1	Total Coliform
		2	Ecoli
		3	Pseudomonas
	Number of units		3
	Rescaling method for scale dependents		Standardized
	Activation function		Identity
	Error function		Sum of Squares
a – Determined by the testing data criterion: the “best” number of hidden units is the one that yields the smallest error in the testing data.			

Table 8. RBF model summary

Training	Sum of squares error		3.875
	Average overall relative error		.076
	Relative error for scale dependents	Total coliform	.194
		E. coli	.033
		Pseudomonas	.001
	Training time		0:00:00.06
Testing	Sum of squares error		.456[a]
	Average overall relative error		.048
	Relative error for scale dependents	Total coliform	.018
		E. coli	.060
		Pseudomonas	1.132
a – The number of hidden units is determined by the testing data criterion: the “best” number of hidden units is the one that yields the smallest error in the testing data.			

values of MBE indicate higher accuracy of the model, similarly to the higher values of RMSE.

From the table and figures presented above, it can be noticed that data mining by using RBF model; which is one type of ANN, gives more accurate results compared with the other models. Consequently, this model may be used for the estimation of the data with a high accuracy.

CONCLUSIONS

In this study, neural network models and multiple linear regression techniques were successfully used to estimate the relation between the concentration of total coliform, E. coli and Pseudomonas in the wastewater and the input variables. Two techniques were used to achieve

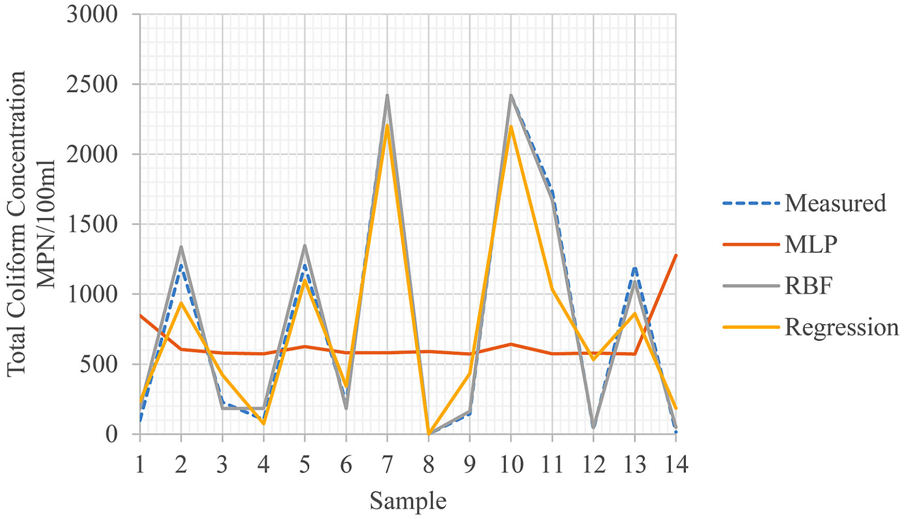

Figure 1. Comparison between the experimental and estimated concentration of total coliform

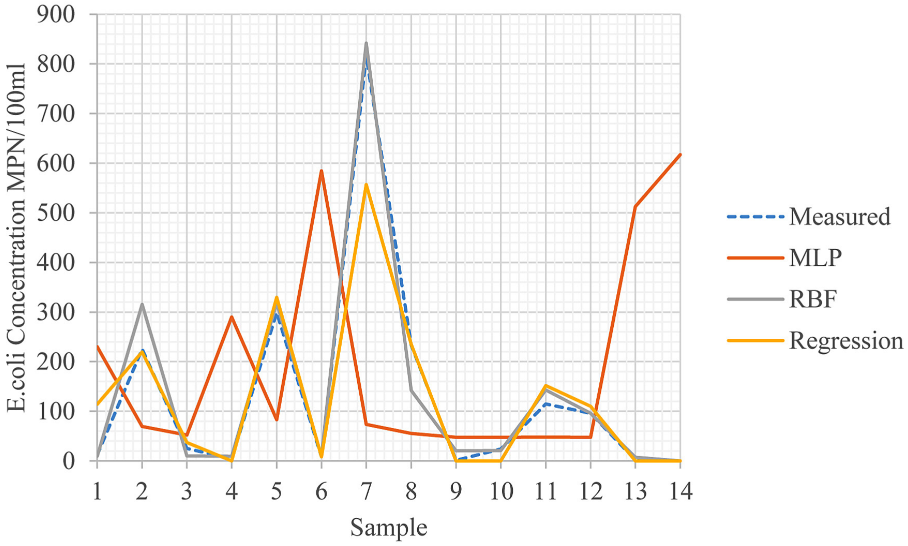

Figure 2. Comparison between the experimental and estimated concentration of E. coli

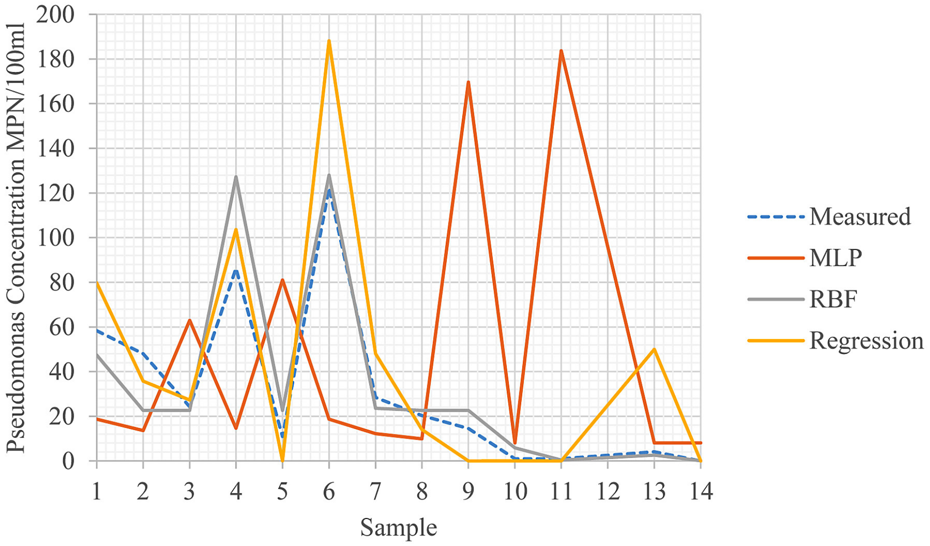

Figure 3. Comparison between the experimental and estimated concentration of Pseudomonas

Table 9. Comparison of performance of the used models based on statistical analysis

Specification	Regression			MLP			RBF		
	R	RMSE	MBE	R	RMSE	MBE	R	RMSE	MBE
Total coliform	0.823	0.325650	976.742605	0.752903	0.416709	1813.351944	0.935967	0.319052	968.608866
E. coli	0.861	0.356578	221.529191	0.881106	0.420189	307.617626	0.983331	0.316701	174.751390
Pseudomonas	0.474	0.286816	127.434216	0.986202	0.116404	20.990208	0.997884	0.305554	144.628897

this objective. The first is a classical technique with multiple linear regression model, while the second one is data mining with two types of ANN (Multilayer Perceptron and Radial Basis Function).

The comparisons between the estimated data and the experimental data showed that data mining by using RBF model has ability to recognize the relation between input and output variables. Moreover, the statistical error analysis showed the accuracy of data mining by using the RBF model.

On the other hand, the obtained results indicate that MLP and multiple linear regression have the least ability for the estimation of the concentration of total coliform, E. coli and Pseudomonas in the wastewater, respectively.

REFERENCES

1. Ahmad A., M. Hamdan, E. Abdelhafez, E. Turk, J. Ibbini & N. Abu Shaban (2019) Water Disinfection by Solar Energy. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects Journal, DOI: 10.1080/15567036.2019.1666182.
2. Barlow, M., and T. Clarke (2017) Blue gold: the battle against corporate theft of the world's water. Routledge.
3. Boyle, M., et al. (2008) Bactericidal effect of solar water disinfection under real sunlight conditions. Appl. Environ. Microbiol. 74, 10, 2997-3001.
4. Burhan, D. (2015) Solar water disinfection considerations: Using ultraviolet light methods to make water safe to drink, IJISET - International Journal of Innovative Science, Engineering & Technology 2, 253–64.
5. Castro-Alférez, M., et al. (2017) Mechanistic modeling of UV and mild-heat synergistic effect on solar water disinfection. Chemical Engineering Journal 316, 111-120.
6. Dawney, B., and J.M. Pearce (2012) Optimizing the solar water disinfection (SODIS) method by decreasing turbidity with NaCl. Journal of Water, Sanitation and Hygiene for Development 2, 2, 87-94.
7. Figueredo-Fernández, M., S. Gutiérrez-Alfaro, A. Acevedo-Merino, and M.A. Manzano (2017) Estimating lethal dose of solar radiation for enterococcus inactivation through radiation reaching the water layer. Application to Solar Water Disinfection (SODIS). Solar Energy 158, 303-310.
8. Giannakis, S., E. Darakas, A.Escalas-Cañellas, and C. Pulgarin (2015) Temperature-dependent change of light dose effects on E. coli inactivation during simulated solar treatment of secondary effluent. Chemical Engineering Science 126, 483-487.
9. Gutiérrez-Alfaro, S., A. Acevedo, M. Figueredo, M. Saladin, and M.A. Manzano (2017) Accelerating the process of solar disinfection (SODIS) by using polymer bags. Journal of Chemical Technology & Biotechnology 92, 2, 298-304.
10. Haider, H. (2017) Exposure Period Assessment for Solar Disinfection (Sodis) under Uncertain Environmental Conditions: A Fuzzy Rule-Based Model. International Journal of Water Resources and Arid Environments, 6.
11. Hernández C.D., J.M. Rodríguez, X.D. Galván, J.O. Medel, and M.R.J. Magaña (2015) Optimal use of chlorine in water distribution networks based on specific locations of booster chlorination: Analyzing conditions in Mexico. Water Science and Technology: Water Supply 16, 2, 493-505.
12. Islam, Md, A.K. Azad, Md Akber, M. Rahman, and I. Sadhu (2015) Effectiveness of solar disinfection (SODIS) in rural coastal Bangladesh. Journal of water and health 13, 4, 1113-1122.
13. Keeratipibul, S., A. Phewpan, and C. Lursinsap (2011) Prediction of coliforms and Escherichia coli on tomato fruits and lettuce leaves after sanitizing by using Artificial Neural Networks. LWT-Food Science and Technology 44, 1, 130-138.
14. Keogh, M.B., K. Elmusharaf, P. Borde, and K.G. McGuigan (2017) Evaluation of the natural coagulant Moringa oleifera as a pretreatment for SODIS in contaminated turbid water. Solar energy, 158, 448-454.
15. Lawrie, K., A. Mills, M. Figueredo-Fernández, S. Gutiérrez-Alfaro, M. Manzano, and M. Saladin (2015) UV dosimetry for solar water disinfection (SODIS) carried out in different plastic bottles and bags. Sensors and Actuators B: Chemical 208, 608-615.

16. Meierhofer, R., and M. Wegelin (2002) Solar water disinfection: a guide for the application of SODIS. Report by SANDEC (Water & Sanitation in Developing Countries) at EAWAG (Swiss Federal Institute for Environmental Science and Technology).
17. Hamdan M. and S. Darabee (2017) Enhancement of solar water disinfection using nanotechnology. International Journal of Thermal & Environmental Engineering 15, 2, 111-116.
18. Oates, P.M., P. Shanahan, and M.F. Polz (2003) Solar disinfection (SODIS): simulation of solar radiation for global assessment and application for point-of-use water treatment in Haiti. Water Research 37, 1, 47-54.
19. Padowski, J.C., S.M. Gorelick, B.H. Thompson, S. Rozelle, and S. Fendorf (2015) Assessment of human–natural system characteristics influencing global freshwater supply vulnerability. Environmental Research Letters 10, 10, 104014.
20. Polo, D., I. García-Fernández, P. Fernández-Ibáñez, and J.L. Romalde (2015) Solar water disinfection (SODIS): Impact on hepatitis A virus and on a human Norovirus surrogate under natural solar conditions. Int. Microbiol 18, 1, 41-49.
21. Sahel, S. Mulaw, N. Belachew, H. Gebretsadik, and G. Gebregziabher (2017) The Effect of Bottle Scratches and Lime Juice on Natural Solar Radiation Disinfection (SODIS) Techniques on Different Bacterial Colonies at ShoaRobit and Surrounding Rural Kebeles. American Journal of Life Sciences 5, 2, 57-64.
22. Shekoohiyan, S., S. Rtimi, G. Moussavi, S. Giannakis, and C. Pulgarin (2019) Enhancing solar disinfection of water in PET bottles by optimized in-situ formation of iron oxide films. From heterogeneous to homogeneous action modes with H2O2 vs. O2–Part 1: Iron salts as oxide precursors. Chemical Engineering Journal 358, 211-224.
23. Sift, M., S. Wagner, and M. Hessling (2016) Investigations on Temperature Effects and Germ Recovery for Solar Water Disinfection (SODIS). International Journal of Applied Sciences and Biotechnology 4, 4, 430-435.
24. Stubbé, S.ML, et al. (2016) Household water treatment and safe storage–effectiveness and economics., Drinking Water Engineering and Science 9. 1, 9-18.
25. Vivar, M., N. Pichel, M. Fuentes, A. López-Vargas (2017) Separating the UV and thermal components during real-time solar disinfection experiments: The effect of temperature. Solar Energy 146, 334-341.
26. Water, a shared responsibility – Report 2 (2006) World Water Assessment Programme (WWAP), New York, US.
27. Wu, W., G.C. Dandy, and H.R. Maier (2014) Optimal control of total chlorine and free ammonia levels in a water transmission pipeline using artificial neural networks and genetic algorithms. Journal of Water Resources Planning and Management 141, 7, 04014085.
28. Zounemat-Kermani, M., A. Ramezani-Charmahineh, J. Adamowski, and O. Kisi (2018) Investigating the management performance of disinfection analysis of water distribution networks using data mining approaches. Environmental monitoring and assessment 190, 7, 397.