E-ISSN: 2347-5129
P-ISSN: 2394-0506
(ICV-Poland) Impact Value: 76.37
(GIF) Impact Factor: 0.549
IJFAS 2023; 11(6): 122-131

www.fisheriesjournal.com
Received: 01-09-2023
Accepted: 05-10-2023

Ekene Jude Nwankwo
Department of Civil Engineering, University of Nigeria, Nsukka, Nigeria

Corresponding Author:
Ekene Jude Nwankwo
Department of Civil Engineering, University of Nigeria, Nsukka, Nigeria

Optimal regression model for the prediction of inactivation rate of *E. coli* in solar disinfection (SODIS)

Ekene Jude Nwankwo

DOI: https://doi.org/10.22271/fish.2023.v11.i6b.2932

Abstract
Solar disinfection (SODIS) is a simple and cost-effective method for disinfecting drinking water with questionable microbial quality by exposing it to sunlight in transparent containers. The effectiveness of SODIS relies on various parameters such as UV intensity (*I*), water temperature (*T*), and turbidity (*Tu*). This study aimed to identify the optimal regression model for predicting the inactivation rate constant of *E. coli* among 28 possible regression equations using *I*, *T*, and *Tu* as predictors. The 28 regression equations were derived from seven combinations of predictor variables (*I*, *T*, & *Tu*; *I* & *T*; *I* & *Tu*; *T* & *Tu*; *I*; *T*; and *Tu*; and *Tu*) utilizing four trends (linear, logarithmic, exponential, and power). The proposed models were calibrated using data collected from 33 SODIS experiments conducted over a five-month period from April to August 2021. Based on rankings from the Taylor diagram, the regression equation combining the linear trend and *I* & *Tu* as predictors demonstrated the best predictive performance. Residual analysis indicated that square root transformation was necessary to improve normality and homogeneity of residuals. Notably, turbidity within the range of 1 – 30 NTU, previously considered nonsignificant, became significant after the square root transformation. This study underscores the importance of an exhaustive approach that considers all possible combinations of predictor variables and trends, allowing the data to reveal patterns and correlations without premature restrictions, thereby ensuring no potentially valuable insights are overlooked.

Keywords: SODIS, temperature, UV, turbidity, *E. coli*, Taylor diagram

1. Introduction
Solar Disinfection (SODIS) is an economical and straightforward household water treatment (HWT) method that eliminates water-borne pathogens by exposing water in transparent containers to a day or two of strong sunlight [1]. The effectiveness of Solar Water Disinfection (SODIS) hinges upon a complex interplay of various factors, namely Ultraviolet (UV) intensity, water temperature, water turbidity, dissolved oxygen (DO), and dissolved organic matter, each exerting significant influence on the process outcome [2-5]. Among these factors, UV radiation intensity stands out as a pivotal determinant, governing the degree of microbial inactivation. UV radiation operates through a dual mechanism: directly damaging microbial DNA and RNA, thereby impeding replication and transcription processes [6, 7], and interacting with photosensitizers in water to produce reactive oxygen species (ROS) [8, 9]. These ROS, comprising singlet oxygen, superoxide, hydrogen peroxide, and hydroxyl radicals, are highly reactive, and can attack cellular components like DNA, proteins, and lipids, culminating in cell death [10-13]. The role of DO is equally crucial, as it facilitates ROS production. Strategies such as vigorous agitation of partially filled containers before topping them off have been proposed to enhance DO content and optimize SODIS performance [14]. Solar heat increases water temperature, which causes cell death by denaturing proteins, disrupting molecular structures, and inhibiting DNA repair mechanisms [15, 16].
Turbidity, stemming from suspended particles, compromise SODIS efficacy by absorbing and scattering UV radiation, diminishing its penetration depth and microbial inactivation effectiveness [3]. Research underscores the detrimental impact of turbidity on SODIS efficacy, with higher turbidity levels correlating with decreased pathogen removal rates and fostering conditions conducive to post-irradiation regrowth of pathogens [15, 17, 18]. Moreover, dissolved organic matter, including humic acids, can impede disinfection efficacy by acting as internal

UV filters, absorbing UVA radiation, although the precise mechanism remains unclear and may vary based on organic material type and concentration [4].

It is important to note that the functional relationship between SODIS effectiveness and the mentioned process parameters remains underexplored, despite their significant impact. Few studies have delved into this relationship, with second-order polynomial regression analysis being a prevalent tool. Gómez-Couso *et al.* [18] used multifactorial regression to examine the joint effect of radiation intensity, water turbidity, and exposure time on *C. parvum* oocysts, highlighting significant interactions between radiation intensity and exposure time. Mansoor Ahammed *et al.* [19] employed response surface regression, uncovering a significant turbidity-dissolved oxygen interaction in optimizing bacterial removal. Recently, Samoili *et al.* [16] introduced an innovative approach combining static and dynamic multivariate linear regression models to characterize the *E. coli* inactivation curve, demonstrating robustness across varied climates. However, Nwankwo *et al.* [20] faced collinearity issues in their regression model, hindering precise estimates due to high UV and temperature correlation in the study area. Separate regression equations for UV and temperature were the only alternative to deal with this issue, albeit with reduced prediction accuracy. It is also noteworthy that while second-order polynomial regression analysis has been commonly used, other trends such as linear, logarithmic, exponential, and power trends have received little attention. Therefore, the objective of the present study was to compare various trends and find optimal regression model for the prediction of pathogen inactivation rate in SODIS using UV intensity, water temperature, and water turbidity as predictor variables.

2. Materials and Method

2.1 Experimental setup

The regression data were collected through 33 SODIS experiments conducted over 5 months between April and August 2021. The experimental setup involved a 1.5 L Coca-Cola polyethylene terephthalate (PET) bottle equipped with a mercury-in-glass thermometer and positioned on an absorptive surface (black styrofoam material), as illustrated in Figure 1. Coca-Cola PET bottles have become the preferred plastic container due to their thermal stability, durability, and resistance to scratches and abrasion during prolonged use. Additionally, Coca-Cola PET is widely available worldwide. The absorptive support was expected to expedite water temperature increase by trapping and transmitting absorbed heat directly to the water. To prevent leakage without glue, the thermometer was tightly fitted through a catheter hub before insertion into a hole made in the plastic bottle's cork. Temperature was monitored at 30-minute intervals to obtain the daily maximum temperature. Thermometer accuracy was ensured by regular cross-checking with a reference digital thermometer at the National Centre for Energy Research and Development (NCERD), confirming that the experimental thermometer measured within +0.2 °C of the reference digital thermometer.

Fig 1: Experimental setup

2.2 Test water preparation, cultivation and enumeration of *E. coli*

E. coli was selected as the model pathogen for this study due to its role as an indicator of fecal contamination and its frequent examination in SODIS research, as it exhibits greater resistance to sunlight's germicidal effects compared to many other bacteria [2]. The cultivation and enumeration procedure for *E. coli* were detailed in Ubomba-Jaswa *et al.* [21]. Test water for all experiments was drawn from a 200 L plastic water storage drum collected in one batch from a borehole source to maintain consistent physicochemical properties, which were monitored monthly. Before each experiment, the test water was sterilized and then contaminated from a previously cultivated *E. coli* stock, achieving approximately 10^6 CFU/mL through appropriate dilution. Bottles were filled two-thirds full, vigorously shaken to enhance oxygen absorption, and then topped up to full volume, ensuring dissolved oxygen (DO) values of 6.1±0.21 mg/L. The test water was confirmed free of chlorine. Turbid water was prepared by adding kaolin (China clay) to the test water until the desired turbidity level was reached. Turbidity levels of test water for experiments were randomly varied between 1 and 30 NTU. Initial samples were collected from the reactors before solar exposure, with subsequent samples taken at 30-minute intervals or longer, depending on the received radiation since the last sampling, using a sterilized hypodermic syringe.

Ultraviolet radiation was measured with a digital UV light meter (General Tools UV513AB Digital UVA/UVB Meter, 280–400 nm), which displays radiant energy per unit area in mW/cm² or µW/cm² on its LCD screen (see Figure 1). Readings were captured using the Open Camera 1.48.3 app for Android phones, capable of taking shots repeatedly at preset time intervals. During experiments, the camera was set up to capture the LCD screen of the UV meter every 60 seconds from 10 a.m. to 4 p.m. These readings were utilized to estimate the daily UV profile, assessing the daily maximum of 5-hour averages of UV intensity between 10 a.m. and 4 p.m.

The inactivation rate constant of *E. coli* was evaluated using the classic first-order kinetics based on Chick's law [22]. Chick's law is given as follows: $N_t = N_o \exp.(-k_e t)$, where N_t is the *E. coli* concentration at time t; N_o is the initial concentration of *E. coli*; $exp.$ is the exponential function; and k_e is inactivation rate constant of *E. coli*. The value of k_e was determined by regressing $\log_e N_t/N_o$ with respect to t. All statistical analysis, including the computation of regression parameters and test of assumptions, were executed with Real Statistics Using Excel (version: Rel 8.9.1, released 2 October 2023). Regression theory, as well as the procedure for model development and test of regression assumptions can be found in Introduction to Linear Regression Analysis by Montgomery *et al*. [23]

2.3 Procedure for the selection of optimal regression equations

The response variable is the inactivation rate constant of *E. coli* (k_e); whereas, the predictor variables are 5-h average of around-noon UV intensity (I), maximum water temperature (T), and water turbidity (Tu). To find the optimal regression model, all the 7 possible combinations of the predictor variables ($I, T \& Tu$; $I \& T$; $I \& Tu$; $T \& Tu$; I; T: and Tu) were regressed against k_e using the four proposed trends (Linear, Logarithmic, Exponential, and power trends), which resulted in a total of 28 model equations (7 possible combinations multiplied by 4 trends equals 28), as shown in Table 2.

The flowing steps were applied to narrow down to the optimal model equation:

1. Fit the full model (The model with all of the predictor variable under consideration).
2. Perform a thorough analysis of each model and trend, including a full residual analysis and test of assumptions. Also investigate the possibility of collinearity.
3. Determine if there is need for transformation or otherwise based on the analysis in (2) above. After transformation, repeat the analysis in (2) above to evaluate improvement and determine model adequacy.
4. Use Student's t-test on individual predictor variables to edit the model, so that model equations with a non-significant coefficient are not considered further. In cases where only the intercept term (β_o) is not significant, regress the model again without the intercept term.
5. Finally, rank the remaining model equations using Taylor diagram, so that the best model can be selected based on correlation, *RMSE*, and closeness of the fitted and the observed standard deviation.

3. Results and Discussion

3.1 Model building process

Table 2 show the results for the test of regression assumptions before the square root transformation was applied. While the normality test was performed using Shapiro-Wilk method, heteroscedasticity (Stability of variance) and autocorrelation were tested using Breusch-Pagan and Durbin-Watson statistics, respectively. All tests were performed at 95% confidence interval ($\alpha = 0.05$). The result revealed that the regression data required transformation and improvement. It can be seen that none of the proposed regression equations passed all the three assumption tests. To improve the regression data, square root transformation was applied to the response variable k_e. Square root transformation is usually employed where the residual plot forms an out-ward opening funnel [23]. A plot of studentized residual versus predicted response for the full linear model is shown in Figure 1, and it can be seen that the plot is shaped like an outward-opening funnel. Tables 2 and 3 show the result of assumption tests for the original and the transformed data. There is notable improvement in the number of models that meets the regression criterion after the square root transformation of the response variable was applied. In general, the normality rate of the residuals increased from 14% before transformation to 54% after transformation. Likewise, the percentage of models with stable variance (homogenous variance) increased from initial 61% to 86% after transformation. No changes were observed in the autocorrelation rate after transformation.

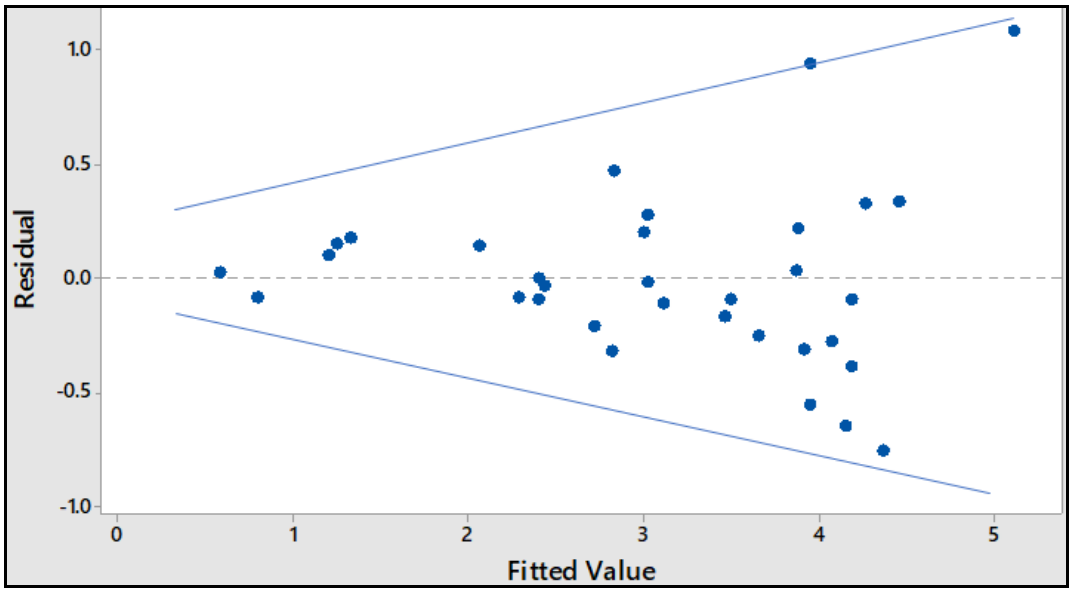

Fig 2: Residual Plot for full linear model ($I, T \& Tu$)

The 28 models generated from the transformed data underwent assumption tests. Models failing any assumption test were excluded from further consideration. In total, 11 regression models (Models with serial number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, and 18 in Table 3) out of the 28 models satisfied all three regression assumptions. All the equations with exponential and power trends were eliminated at this stage, remaining only equations with linear and logarithmic trends. Not that regression equations with "unclear" Durbin-Watson autocorrelation result were also eliminated at this stage. Subsequently, the regression equations of these 11 models underwent student's t-test editing so that models with a non-significant coefficient could be eliminated. The intercept term was removed for models with non-significant intercept terms and the regressions were re-run with no intercept term. Six models (Models with serial number 2, 4, 5, 6, 9 and 12) retained after the t-test editing. To identify the optimal regression model, the performance of these 6 models was compared using Taylor's diagram as well as Box and Whisker plot. The Taylor diagram shown in Figure 3 evaluates model performance by comparing correlations, root mean square errors (RMSE), and standard deviations. The position of the marker representing each model in Taylor diagram is determined based on these metrics. Models are then assigned ranks based on the relative distance of its marker from the "Reference" point. The marker representing Mode1 4 in Table 4 ($\sqrt{k_s} = -1.142 + 0.066T - 0.005Tu$) is the closest to the "Reference" and therefore is considered the best model. Unexpectedly the model that emerged the best does not contain UV intensity (I) as a predictor. UV intensity is known to be the most important process parameter in SODIS. One plausible explanation for this anomaly is the high correlation between UV intensity and water temperature. This means that the temperature variable may have captured the variability of pathogen die-off rate as much as UV intensity, and captured the variability more when combined with turbidity. Moreover, to large extent, especially in the Tropics, variability in solar radiation can be captured by temperature [24, 25]. However, the model that ranked second ($\sqrt{k_s} = -7.264 + 0.515(\ln I) + 1.849(\ln T)$) contained both UV intensity and water temperature as predictors.

Box-and-whisker plots (see Figure 4) were used to compare the distributional characteristics of observed and predicted responses. In these plots, the upper and lower lines of the box represent the 25th and 75th percentiles, respectively. The height of the box signifies the interquartile range, while the line and asterisk at the middle indicate the median and mean, respectively. Whiskers extend above and below the box, with any point outside them considered a potential outlier. It can be observed that, except for the two outliers, the distributional characteristics of the observed response are reasonably preserved in the six regression models.

3.2 Effect of UV intensity and water temperature

The results show that UV intensity and water temperature have strong influence on die-off rate of *E. coli*. The two-process parameter can be used to explain more than 99% variability in *E. coli* die-off rate. Other studies, including those by Brockliss *et al.* [26] and Samoili *et al.* [16], have similarly reported a high correlation when the cumulative die-off rate of *E. coli* was regressed using irradiance and water temperature as predictors. Despite the higher *R-square* values recorded by the regression equations that contain I and T as predictors, Taylor diagram ranked them low due to the higher *RMSE* and larger difference between the standard deviation of the fitted and the observed responses. It is important to note that the VIF value of these regression equations suggest high correlation between UV intensity and water temperature. Where one or more predictor variables are highly correlated, the *R-squared* value can become artificially inflated and misleading, suggesting a better fit than actually exists [27]. This is particularly relevant in tropical regions where sunny periods are typically accompanied by high temperatures, as highlighted by Nwankwo and Agunwamba [24]. Weather conditions such as 'sunny and cold' are rare occurrences in the study area. Therefore, future regression models should address the potential interference between water temperature and UV intensity by either eliminating one of the variables or exploring alternative methods to handle multicollinearity effectively.

The rankings of the single variable models (Models ranking 3, 4, and 6 in Table 1) revealed that maximum water temperature provided a superior fit compared to UV intensity. This indicates that, in this region, maximum water temperature serves as a more reliable predictor of SODIS efficiency than average UV intensity. A prior study noted that daily maximum water temperature could contain some dosimetric information [24]. For instance, complete inactivation was observed in all experiments where water temperature exceeded 45 °C. In the Tropics, maximum water temperature typically occurs between 1 and 3 p.m. [25]. Furthermore, temperature data is more cost-effective to acquire and may require only a simple mercury-in-glass thermometer, costing less than a dollar, compared to the expensive 'high-tech' devices required for UV light intensity measurement.

3.3 Effect of water turbidity on SODIS efficiency

Water turbidity is a critical consideration in applying SODIS, as previous studies consistently demonstrate a notable decrease in its efficacy as turbidity levels rise [15, 18, 28]. However, these studies examined water with turbidity levels well beyond those tested in our study. Over time, a turbidity threshold of 30 NTU has emerged as a criterion for selecting SODIS water, with Luzi *et al.* [1] proposing a simple test for determining if water turbidity is below this threshold. It's recommended to pretreat water to reduce turbidity below 30 NTU as part of the SODIS protocol. Our results show that the regression coefficient for the turbidity variable is significant, indicating substantial evidence that water turbidity within the range of 1–30 NTU significantly affects *E. coli* inactivation rates. Amirsoleimani & Brion [28] also found a statistically significant difference (t-test: $p < 0.001$) between the *E. coli* inactivation rates of 0 and 30 NTU SODIS water, highlighting the protective nature of turbidity in this range, which can help a significant number of *E. coli* evade inactivation by UV radiation. These results contrast with those of Nwankwo *et al.* [20], who found the turbidity variable to be nonsignificant in regression analysis, both alone and when combined with other predictor variables. However, the study by Nwankwo *et al.* [20] did not assess regression assumptions and therefore applied no transformation to mitigate data ill-conditioning.

It's important to note that in turbid water, while the direct effect of UVB on the major cellular components of bacterial pathogens at the inner layer of SODIS water is obstructed by

suspended particles, the absorbed and scattered portions of UVA in water remain effective. These can still generate reactive oxygen species (ROS) capable of migrating to inner water layers and killing pathogens [1]. This likely explains why complete disinfection can still be achieved in turbid water, even when most cells do not directly receive sunlight. Another factor contributing to disinfection in turbid water is water temperature, which tends to be higher due to the capacity of suspended particles to trap radiant heat emitted by the sun as infrared (IR). Amirsoleimani & Brion [28] observed a statistically significant difference in temperatures between SODIS water with turbidity levels of 0 and 30 NTU under the same conditions, with SODIS containers containing 30 NTU water being notably warmer. Surprisingly, Studies have shown improved efficiency in SODIS when treating moderately turbid water (38 NTU) compared to water with lower turbidity levels (<5 NTU) [29]. Complete inactivation can be achieved in SODIS-treated water with higher turbidity levels, provided that temperatures of 55 °C and above are reached [30].

While higher turbidity levels (>30 NTU) may offer a temperature advantage, they are not desirable qualities in SODIS water. SODIS-treated water exhibits greater durability, with no bacteria regrowth, when clear water is used provided the necessary UV dose for complete disinfection is reached [31, 32]. Furthermore, highly turbid water can significantly diminish the aesthetics and user acceptance of SODIS-treated water.

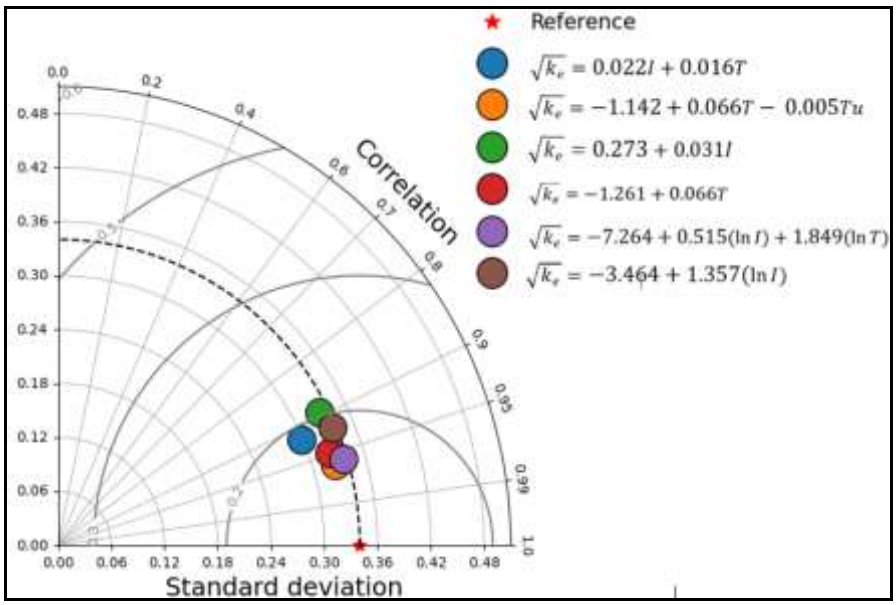

Fig 3: Taylor diagram, comparing the predictive performance of the proposed models

Table 1: Rankings of significant regression equations based on Taylor diagram

Models retained after t-test	Ranking based Taylor diagram
$\sqrt{k_e} = 0.022I + 0.016T$	5
$\sqrt{k_e} = -1.142 + 0.066T - 0.005Tu$	1
$\sqrt{k_e} = 0.273 + 0.031I$	6
$\sqrt{k_e} = -1.261 + 0.066T$	3
$\sqrt{k_e} = -7.264 + 0.515(\ln I) + 1.849(\ln T)$	2
$\sqrt{k_e} = -3.464 + 1.357(\ln I)$	4

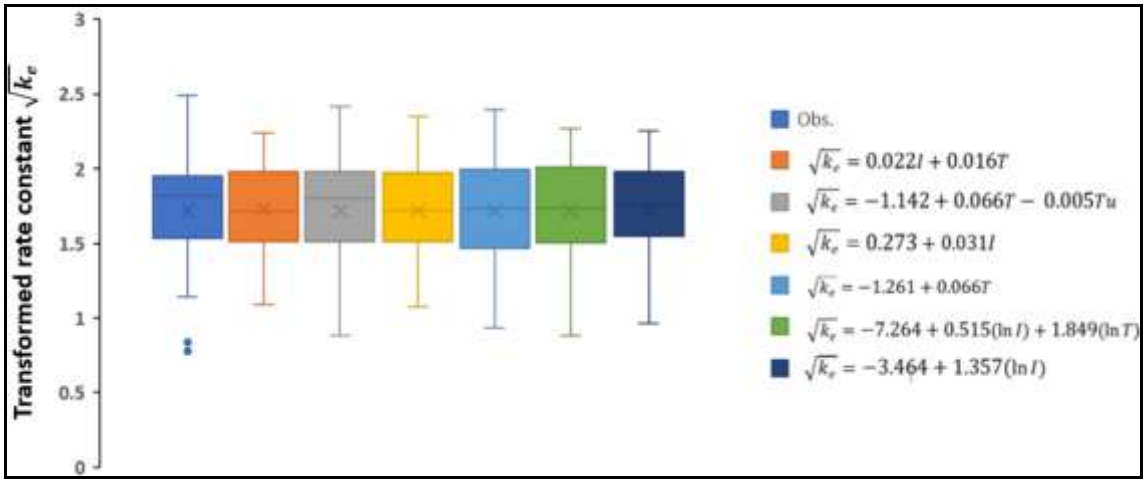

Fig 4: Box and Whisker's plot, comparing the distributional characteristics the proposed models

Table 2: Test of regression assumptions before transformation of response variable (k_e)

SN	Trend	Proposed Model	Shapiro-Wilk (SW) Normality Test			Durbin-Watson (DW) Autocorrelation Test				Breusch-Pagan (BP) Heteroscedasticity Test for Stability of Variance				
			SW-stat	p-value	Normal?	DW-stat	DW-lower	DW-upper	Auto correlated?	F stat	df1	df2	p-value	Stable variance?
1	Linear	$k_e = \beta_o + \beta_1(I) + \beta_2(T_m) + \beta_3(Tu)$	0.944	0.087	Yes	2.283	1.258	1.651	No	3.97	3	29	0.017	No
2	Linear	$k_e = \beta_o + \beta_1(I) + \beta_2(T_m)$	0.935	0.05	No	2.215	1.321	1.577	No	4.974	2	30	0.014	No
3	Linear	$k_e = \beta_o + \beta_1(I) + \beta_3(Tu)$	0.924	0.024	No	2.269	1.321	1.577	No	2.589	2	30	0.092	Yes
4	Linear	$k_e = \beta_o + \beta_2(T_m) + \beta_3(Tu)$	0.931	0.037	No	2.425	1.321	1.577	Unclear	6.503	2	30	0.005	No
5	Linear	$k_e = \beta_o + \beta_1(I)$	0.923	0.022	No	2.259	1.383	1.508	No	5.23	1	31	0.029	No
6	Linear	$k_e = \beta_o + \beta_2(T_m)$	0.921	0.02	No	2.368	1.383	1.508	No	11.036	1	31	0.002	No
7	Linear	$k_e = \beta_o + \beta_3(Tu)$	0.982	0.844	Yes	0.775	1.383	1.508	Yes	0.024	1	31	0.879	Yes
8	Logarithmic	$k_e = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.908	0.009	No	2.289	1.258	1.651	No	2.248	3	29	0.104	Yes
9	Logarithmic	$k_e = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m)$	0.907	0.008	No	2.288	1.321	1.577	No	3.522	2	30	0.042	No
10	Logarithmic	$k_e = \beta_o + \beta_1 \log_e(I) + \beta_3 \log_e(Tu)$	0.876	0.001	No	2.343	1.321	1.577	No	1.689	2	30	0.202	Yes
11	Logarithmic	$k_e = \beta_o + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.902	0.006	No	2.336	1.321	1.577	No	4.496	2	30	0.02	No
12	Logarithmic	$k_e = \beta_o + \beta_1 \log_e(I)$	0.888	0.003	No	2.363	1.383	1.508	No	3.448	1	31	0.073	Yes
13	Logarithmic	$k_e = \beta_o + \beta_2 \log_e(T_m)$	0.897	0.004	No	2.345	1.383	1.508	No	9.156	1	31	0.005	No
14	Logarithmic	$k_e = \beta_o + \beta_3 \log_e(Tu)$	0.984	0.907	Yes	0.759	1.383	1.508	Yes	0.38	1	31	0.542	Yes
15	Exponential	$k_e = \beta_o \exp\beta_1(I) \times \exp\beta_2(T_m) \times \exp\beta_3(Tu)$	0.91	0.01	No	1.473	1.258	1.651	Unclear	1.671	3	29	0.195	Yes
16	Exponential	$k_e = \beta_o \exp\beta_1(I) \times \exp\beta_2(T_m)$	0.912	0.011	No	1.272	1.321	1.577	Yes	3.948	2	30	0.03	No
17	Exponential	$k_e = \beta_o \exp\beta_1(I) \times \exp\beta_3(Tu)$	0.896	0.004	No	1.563	1.321	1.577	Unclear	3.027	2	30	0.063	Yes
18	Exponential	$k_e = \beta_o \exp\beta_2(T_m) \times \exp\beta_3(Tu)$	0.895	0.004	No	1.65	1.321	1.577	No	0.951	2	30	0.398	Yes
19	Exponential	$k_e = \beta_o \exp\beta_1(I)$	0.879	0.002	No	1.49	1.383	1.508	Unclear	3.409	1	31	0.074	Yes
20	Exponential	$k_e = \beta_o \exp\beta_2(T_m)$	0.91	0.01	No	1.504	1.383	1.508	Unclear	3.909	1	31	0.057	Yes
21	Exponential	$k_e = \beta_o \exp\beta_3(Tu)$	0.888	0.003	No	0.619	1.383	1.508	Yes	2.305	1	31	0.139	Yes
22	Power	$k_e = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2} \times (T_m)^{\beta_3}$	0.936	0.052	Yes	1.43	1.258	1.651	Unclear	2.483	3	29	0.081	Yes
23	Power	$k_e = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2}$	0.934	0.044	No	1.398	1.321	1.577	Unclear	4.918	2	30	0.014	No
24	Power	$k_e = \beta_o(I)^{\beta_1} \times (Tu)^{\beta_3}$	0.933	0.042	No	1.757	1.321	1.577	No	2.989	2	30	0.065	Yes
25	Power	$k_e = \beta_o(T_m)^{\beta_2} \times (Tu)^{\beta_3}$	0.916	0.014	No	1.581	1.321	1.577	No	1.224	2	30	0.308	Yes
26	Power	$k_e = \beta_o(I)^{\beta_1}$	0.932	0.04	No	1.753	1.383	1.508	No	4.947	1	31	0.034	No
27	Power	$k_e = \beta_o(T_m)^{\beta_2}$	0.93	0.034	No	1.575	1.383	1.508	No	3.66	1	31	0.065	Yes
28	Power	$k_e = \beta_o(Tu)^{\beta_3}$	0.892	0.003	No	0.603	1.383	1.508	Yes	1.784	1	31	0.191	Yes

k_e – inactivation rate constant of *E. coli*; I – 5-hour average of around-noon UV intensity; T – maximum water temperature; Tu – water turbidity. β_o, β_1, β_2 and β_3 are regression coefficients.

Table 3: Test of regression assumptions using transformed response variable ($\sqrt{k_e}$)

SN	Trend	Proposed Model	Shapiro-Wilk (SW) Normality Test			Durbin-Watson (DW) Autocorrelation Test				Breusch-Pagan (BP) Heteroscedasticity Test for Stability of Variance				
			SW-stat	p-value	Normal?	DW-stat	DW-lower	DW-upper	Auto correlated?	F stat	df1	df2	p-value	Stable variance?
1*	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I) + \beta_2(T_m) + \beta_3(Tu)$	0.988	0.97	Yes	NA	NA	NA	NA	0.627	3	29	0.603	Yes
2*	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I) + \beta_2(T_m)$	0.971	0.494	Yes	NA	NA	NA	NA	0.262	2	30	0.772	Yes
3*	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I) + \beta_3(Tu)$	0.951	0.139	Yes	1.914	1.321	1.577	No	0.974	2	30	0.389	Yes
4*	Linear	$\sqrt{k_e} = \beta_o + \beta_2(T_m) + \beta_3(Tu)$	0.971	0.519	Yes	2.191	1.321	1.577	No	1.785	2	30	0.185	Yes
5*	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I)$	0.947	0.109	Yes	1.876	1.383	1.508	No	0.304	1	31	0.585	Yes
6*	Linear	$\sqrt{k_e} = \beta_o + \beta_2(T_m)$	0.978	0.73	Yes	2.047	1.383	1.508	No	2.039	1	31	0.163	Yes
7	Linear	$\sqrt{k_e} = \beta_o + \beta_3(Tu)$	0.959	0.247	Yes	0.646	1.383	1.508	Yes	0.935	1	31	0.341	Yes
8*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.977	0.686	Yes	2.045	1.258	1.651	No	1.444	3	29	0.25	Yes
9*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m)$	0.974	0.587	Yes	2.038	1.321	1.577	No	2.062	2	30	0.145	Yes
10*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_3 \log_e(Tu)$	0.975	0.632	Yes	2.183	1.321	1.577	No	0.811	2	30	0.454	Yes
11*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.981	0.818	Yes	2.163	1.321	1.577	No	2.78	2	30	0.078	Yes
12*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I)$	0.98	0.78	Yes	2.209	1.383	1.508	No	1.18	1	31	0.286	Yes
13	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_2 \log_e(T_m)$	0.973	0.555	Yes	2.171	1.383	1.508	No	4.707	1	31	0.038	No
14	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_3 \log_e(Tu)$	0.964	0.328	Yes	0.631	1.383	1.508	Yes	1.23	1	31	0.276	Yes
15	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_2(T_m) \times \exp \beta_3(Tu)$	0.91	0.01	No	1.473	1.258	1.651	Unclear	1.671	3	29	0.195	Yes
16	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_2(T_m)$	0.912	0.011	No	1.272	1.321	1.577	Yes	3.948	2	30	0.03	No
17	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_3(Tu)$	0.896	0.004	No	1.563	1.321	1.577	Unclear	3.027	2	30	0.063	Yes
18*	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_2(T_m) \times \exp \beta_3(Tu)$	0.895	0.004	No	1.65	1.321	1.577	No	0.951	2	30	0.398	Yes
19	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I)$	0.879	0.002	No	1.49	1.383	1.508	Unclear	3.409	1	31	0.074	Yes
20	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_2(T_m)$	0.91	0.01	No	1.504	1.383	1.508	Unclear	3.909	1	31	0.057	Yes
21	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_3(Tu)$	0.888	0.003	No	0.619	1.383	1.508	Yes	2.305	1	31	0.139	Yes
22	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2} \times (T_m)^{\beta_3}$	0.936	0.052	Yes	1.43	1.258	1.651	Unclear	2.483	3	29	0.081	Yes
23	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2}$	0.934	0.044	No	1.398	1.321	1.577	Unclear	4.918	2	30	0.014	No
24	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (Tu)^{\beta_3}$	0.933	0.042	No	1.757	1.321	1.577	No	2.989	2	30	0.065	Yes
25	Power	$\sqrt{k_e} = \beta_o(T_m)^{\beta_2} \times (Tu)^{\beta_3}$	0.916	0.014	No	1.581	1.321	1.577	No	1.224	2	30	0.308	Yes
26	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1}$	0.932	0.04	No	1.753	1.383	1.508	No	4.947	1	31	0.034	No
27	Power	$\sqrt{k_e} = \beta_o(T_m)^{\beta_2}$	0.93	0.034	No	1.575	1.383	1.508	No	3.66	1	31	0.065	Yes
28	Power	$\sqrt{k_e} = \beta_o(Tu)^{\beta_3}$	0.892	0.003	No	0.603	1.383	1.508	Yes	1.784	1	31	0.191	Yes

k_e – inactivation rate constant of *E. coli*; I – 5-hour average of around-noon UV intensity; T – maximum water temperature; Tu – water turbidity. β_o, β_1, β_2 and β_3 are regression coefficients; NA – Not applicable; asterisk (*) – passed the three regression assumptions

Table 4: Regression results using transformed response variable ($\sqrt{k_e}$)

SN	Trend	Proposed Model	R_a^2	RMSE	Std.	β_o	p	β_1	p	VIF	β_2	p	VIF	β_3	p	VIF
1	Linear	$\sqrt{k_e} = \beta_1(I) + \beta_2(T_m) + \beta_3(Tu)$	0.995	0.121	0.328	NA	NA	0.019	<0.001	6.749	0.02	<0.001	6.572	-0.004	0.144	1.083
2*	Linear	$\sqrt{k_e} = \beta_1(I) + \beta_2(T_m)$	0.995	0.126	0.325	NA	NA	0.022	<0.001	6.277	0.016	<0.001	6.277			
3	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I) + \beta_3(Tu)$	0.841	0.143	0.348	0.299	0.038	0.031	<0.001	1.034				-0.001	0.744	1.034
4*	Linear	$\sqrt{k_e} = \beta_o + \beta_2(T_m) + \beta_3(Tu)$	0.911	0.107	0.361	-1.142	<0.001				0.066	<0.001	1.007	-0.005	0.043	1.007
5*	Linear	$\sqrt{k_e} = \beta_o + \beta_1(I)$	0.846	0.144	0.348	0.273	0.021	0.031	<0.001	NA						
6*	Linear	$\sqrt{k_e} = \beta_o + \beta_2(T_m)$	0.901	0.115	0.359	-1.261	<0.001				0.066	<0.001				
7	Linear	$\sqrt{k_e} = \beta_o + \beta_3(Tu)$	0.005	0.365	0.072	1.872	<0.001							-0.009	0.289	
			R_a^2	RMSE	Std.	β_o	p	β_1	p	VIF	β_2	p	VIF	β_3	p	VIF
8	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.921	0.099	0.364	-7.277	<0.001	0.5	0.033	9.827	1.872	<0.001	9.62	-0.008	0.782	1.071
9*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_2 \log_e(T_m)$	0.924	0.099	0.363	-7.264	<0.001	0.515	0.023	9.3	1.849	<0.001	9.3			
10	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I) + \beta_3 \log_e(Tu)$	0.88	0.124	0.355	-3.523	<0.001	1.363	<0.001	1.036				0.013	0.705	1.036
11	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_2 \log_e(T_m) + \beta_3 \log_e(Tu)$	0.911	0.107	0.361	-9.011	<0.001				2.843	<0.001	1.014	-0.022	0.448	1.014
12*	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_1 \log_e(I)$	0.884	0.125	0.355	-3.464	<0.001	1.357	<0.001	NA						
13	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_2 \log_e(T_m)$	0.912	0.108	0.361	-9.125	<0.001				2.857	<0.001	NA			
14	Logarithmic	$\sqrt{k_e} = \beta_o + \beta_3 \log_e(Tu)$	-0.009	0.367	0.057	1.944	<0.001							-0.083	0.4	NA
			R_a^2	RMSE	Std.	$In(\beta_o)$	p	β_1	p	VIF	β_2	p	VIF	β_3	p	VIF
15	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_2(T_m) \times \exp \beta_3(Tu)$	0.884	0.082	0.243	-1.262	<0.001	0.003	0.333	6.749	0.037	<0.001	6.572	-0.004	0.075	1.083
16	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_2(T_m)$	0.875	0.087	0.241	-1.277	<0.001	0.005	0.152	6.277	0.034	<0.001	6.277			
17	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I) \times \exp \beta_3(Tu)$	0.788	0.113	0.23	-0.406	0.001	0.02	<0.001	1.034				-0.001	0.565	1.034
18	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_2(T_m) \times \exp \beta_3(Tu)$	0.884	0.083	0.243	-1.387	<0.001				0.044	<0.001	1.007	-0.004	0.037	1.007
19	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_1(I)$	0.793	0.113	0.23	-0.441	<0.001	0.02	<0.001	NA						
20	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_2(T_m)$	0.87	0.09	0.24	-1.482	<0.001				0.045	<0.001	NA			
21	Exponential	$\sqrt{k_e} = \beta_o \exp \beta_3(Tu)$	0.013	0.247	0.054	0.628	<0.001							-0.006	0.243	NA
			R_a^2	RMSE	Std.	$In(\beta_o)$	p	β_1	p	VIF	β_2	p	VIF	β_3	p	VIF
22	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2} \times (T_m)^{\beta_3}$	0.907	0.073	0.246	-5.762	<0.001	0.271	0.112	9.827	1.392	<0.001	9.62	-0.017	0.432	1.071
23	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (T_m)^{\beta_2}$	0.908	0.148	0.491	-11.472	<0.001	0.604	0.069	9.3	2.685	<0.001	9.3			
24	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1} \times (Tu)^{\beta_3}$	0.858	0.092	0.239	-2.972	<0.001	0.913	<0.001	1.036				-0.001	0.967	1.036
25	Power	$\sqrt{k_e} = \beta_o(T_m)^{\beta_2} \times (Tu)^{\beta_3}$	0.902	0.077	0.245	-6.703	<0.001				1.919	<0.001	1.014	-0.024	0.249	1.014
26	Power	$\sqrt{k_e} = \beta_o(I)^{\beta_1}$	0.863	0.092	0.239	-2.976	<0.001	0.914	<0.001	NA						
27	Power	$\sqrt{k_e} = \beta_o(T_m)^{\beta_2}$	0.901	0.078	0.244	-6.828	<0.001				1.934	<0.001	NA			
28	Power	$\sqrt{k_e} = \beta_o(Tu)^{\beta_3}$	-0.001	0.249	0.045	0.69	0.001							-0.065	0.33	NA

k_e – inactivation rate constant of *E. coli*; I – 5-hour average of around-noon UV intensity; T – maximum water temperature; Tu – water turbidity. β_o, β_1, β_2 and β_3 are regression coefficients; NA – Not applicable; asterisk (*) – all regression coefficients are significant ($\alpha = 0.05$)

4. Conclusions

Regression models offer a powerful tool for understanding the intricate relationship between SODIS treatment efficiency and various treatment conditions. Despite the importance attached to UV intensity in SODIS, temperature might wield a more substantial influence on SODIS efficiency than previously recognized. Given the high correlation between UV intensity and water temperature, caution is warranted when combining them in a single least-square regression equation, as this may artificially inflate the R-square value, suggesting a better fit than reality. Thus, it's imperative to assess and address the degree of dependence between these parameters before their incorporation into regression models for predicting SODIS efficiency. To enhance regression assumptions of normality and homogeneity of residuals, particularly when the residual plot resembles an outward opening funnel, employing square root transformation on the response variable proves effective. This transformation also aids in reinforcing the statistical significance of regression coefficients. Notably, turbidity within the range of 1 – 30 NTU, previously deemed nonsignificant, gains significance after square root transformation. Adopting an exhaustive approach by considering all possible combinations of predictor variables facilitates a thorough exploration of potential relationships between predictors and the response variable. Such approach would allow the data to speak for itself without imposing restrictions prematurely and ensures that no potentially valuable patterns or correlations are overlooked.

5. Acknowledgement

I would like to express my gratitude to Mrs. Victoria Onyibo Eze and Engr. Emeka Aroh from the Sanitary Engineering Laboratory at the University of Nigeria, Nsukka, for their valuable technical assistance and guidance throughout the preparation and execution of the experiments.

6. References

1. Luzi S, Tobler M, Suter F, Meierhofer R. SODIS Manual: Guidance on Solar Water Disinfection. Swiss Federal Institute of Aquatic Science and Technology (SANDEC) Dübendorf, Switzerland; 2016. [Online]. Available: https://www.sodis.ch/.
2. McGuigan KG, Conroy RM, Mosler HJ, du Preez M, Ubomba-Jaswa E, Fernandez-Ibañez P. Solar water disinfection (SODIS): A review from bench-top to roof-top. J Hazard Mater. 2012;235–236:29–46. doi: 10.1016/j.jhazmat.2012.07.053.
3. Sommer B, *et al*. SODIS - An emerging water treatment process. J Water Supply Res Technol - AQUA. 1997;46(3):127–137.
4. Wilson SA, Andrews SA. Impact of a natural coagulant pretreatment for colour removal on solar water disinfection (SODIS). J Water Sanit Hyg Dev. 2011;1(1):57–67. doi: 10.2166/washdev.2011.027.
5. Reed RH. Solar inactivation of faecal bacteria in water: The critical role of oxygen. Lett Appl Microbiol. 1997;24(4):276–280. doi: 10.1046/j.1472-765X.1997.00130.x.
6. Jagger J. Solar-UV actions on living cells. 1985.
7. Mullenders LHF. Solar UV damage to cellular DNA: from mechanisms to biological effects. Photochem Photobiol Sci. 2018;17(12):1842–1852. doi: 10.1039/c8pp00182k.
8. Castro-Alférez M, Polo-López MI, Marugán J, Fernández-Ibáñez P. Mechanistic model of the Escherichia coli inactivation by solar disinfection based on the photo-generation of internal ROS and the photo-inactivation of enzymes: CAT and SOD. Chem Eng J. 2017;318:214–223. doi: 10.1016/j.cej.2016.06.093.
9. Rincón A-G, Pulgarin C. Bactericidal action of illuminated TiO2 on pure Escherichia coli and natural bacterial consortia: post-irradiation events in the dark and assessment of the effective disinfection time. Appl Catal B Environ. 2004;49(2):99–112. doi: 10.1016/j.apcatb.2003.11.013.
10. Pigeot-Rémy S, Simonet F, Atlan D, Lazzaroni JC, Guillard C. Bactericidal efficiency and mode of action: A comparative study of photochemistry and photocatalysis. Water Res. 2012;46(10):3208–3218. doi: 10.1016/j.watres.2012.03.019.
11. Sinha RP, Häder DP. UV-induced DNA damage and repair: A review. Photochem Photobiol Sci. 2002;1(4):225–236. doi: 10.1039/b201230h.
12. Berney M, Weilenmann H-U, Egli T. Flow-cytometric study of vital cellular functions in Escherichia coli during solar disinfection (SODIS). Microbiology. 2006;152(6):1719–1729. doi: 10.1099/mic.0.28617-0.
13. Bosshard F, Riedel K, Schneider T, Geiser C, Bucheli M, Egli T. Protein oxidation and aggregation in UVA-irradiated Escherichia coli cells as signs of accelerated cellular senescence. Environ Microbiol. 2010;12(11):2931–2945. doi: 10.1111/j.1462-2920.2010.02268.x.
14. Meierhofer R, Wegelin M. Solar Water Disinfection: A Guide for the Application of SODIS. Dübendorf; 2002. [Online]. Available: www.sodis.ch.
15. McGuigan KG, Joyce TM, Conroy RM, Gillespie JB, Elmore-Meegan M. Solar disinfection of drinking water contained in transparent plastic bottles: Characterizing the bacterial inactivation process. J Appl Microbiol. 1998;84(6):1138–1148. doi: 10.1046/j.1365-2672.1998.00455.x.
16. Samoili S, *et al*. Predicting the bactericidal efficacy of solar disinfection (SODIS): from kinetic modeling of *in vitro* tests towards the in silico forecast of E. coli inactivation. Chem Eng J. 2022;427:130866. doi: 10.1016/j.cej.2021.130866.
17. Kehoe SC, Joyce TM, Ibrahim P, Gillespie JB, Shahar RA, McGuigan KG. Effect of agitation, turbidity, aluminium foil reflectors and container volume on the inactivation efficiency of batch-process solar disinfectors. Water Res. 2001;35(4):1061–1065. doi: 10.1016/S0043-1354(00)00353-5.
18. Gómez-Couso H, Fontán-Saínz M, Sichel C, Fernández-Ibáñez P, Ares-Mazás E. Efficacy of the solar water disinfection method in turbid waters experimentally contaminated with Cryptosporidium parvum oocysts under real field conditions. Trop Med Int Heal. 2009;14(6):620-627.
19. Ahammed MM, Dave S, Nair AT. Effect of water quality parameters on solar water disinfection: a statistical experiment design approach. Desalin Water Treat. 2015;56(2):315-326. doi: 10.1080/19443994.2014.940398.
20. Nwankwo EJ, Agunwamba JC, Igwe SE, Odenigbo C. Regression models for predicting the die-off rate of E. coli in solar water disinfection. J Water Sanit Hyg Dev. 2022;12(8):575–586. doi: 10.2166/washdev.2022.056.
22. Ubomba-Jaswa E, Fernández-Ibáñez P, Navntoft C, Polo-López MI, McGuigan KG. Investigating the microbial inactivation efficiency of a 25 L batch solar disinfection (SODIS) reactor enhanced with a compound parabolic collector (CPC) for household use. J Chem Technol Biotechnol. 2010;85(8):1028–1037. doi: 10.1002/jctb.2398.
23. Chick H. An investigation of the laws of disinfection. Epidemiol Infect. 1908;8(1):92–158.
24. Montgomery DC. Design and Analysis of Experiment. 8th ed. John Wiley & Sons, Inc.; c2013.
25. Nwankwo EJ, Agunwamba JC. Effect of reactor characteristics on the seasonal effectiveness of solar disinfection: A factorial study. Water SA. 2021;47(1):113–122. doi: 10.17159/wsa/2021.v47.i1.9451.
26. Nwankwo EJ, Agunwamba JC, Nnaji CC. Effect of Radiation Intensity, Water Temperature and Support-Base Materials on the Inactivation Efficiency of Solar Water Disinfection (SODIS). Water Resour Manag. 2019;33(13):4539–4551. doi: 10.1007/s11269-019-02407-4.
27. Brockliss S, Luwe K, Ferrero G, Morse T. Assessment of the 20L SODIS bucket household water treatment technology under field conditions in rural Malawi. Int J Hyg Environ Health. 2022;240:113913. doi: 10.1016/j.ijheh.2021.113913.
28. Suárez E, Pérez CM, Rivera R, Martínez MN. Applications of Regression Models in Epidemiology. 1st ed. Hoboken, New Jersey: A John Wiley & Sons, Inc.; 2017.
29. Amirsoleimani A, Brion GM. Solar disinfection of turbid hygiene waters in Lexington, KY, USA. J Water Health. 2021;19(4):642–656. doi: 10.2166/wh.2021.003.
30. Meera V, Ahammed MM. Solar disinfection for household

treatment of roof-harvested rainwater. Water Sci Technol Water Supply. 2008;8(2):153–160. doi: 10.2166/ws.2008.054.
31. Joyce TM, McGuigan KG, Elmore-Meegan M, Conroy RM. Inactivation of fecal bacteria in drinking water by solar heating. Appl Environ Microbiol. 1996;62(2):399–402. doi: 10.1128/aem.62.2.399-402.1996.
32. Vivar M, Fuentes M. Using solar disinfected water: On the bacterial regrowth over 1-week of water usage including direct intake after sun exposure and long-term dark storage. Sol Energy. 2016;131:138–148. doi: 10.1016/j.solener.2016.02.044.
33. Keogh MB, Elmusharaf K, Borde P, McGuigan KG. Evaluation of the natural coagulant Moringa oleifera as a pretreatment for SODIS in contaminated turbid water. Sol Energy. 2017;158:448–454. doi: 10.1016/j.solener.2017.10.010.