

Research Article

A State-Based Sensitivity Analysis for Distinguishing the Global Importance of Predictor Variables in Artificial Neural Networks

Ehsan Ardjmand,¹ David F. Millie,² Iman Ghalehkhondabi,³
William A. Young II,⁴ and Gary R. Weckman^{2,3}

¹Management Department, College of Business, Frostburg State University, Frostburg, MD 21532, USA

²Palm Island Enviro-Informatics LLC, Sarasota, FL 34232, USA

³Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, Athens, OH 45701, USA

⁴Management Information Systems Department, College of Business, Ohio University, Athens, OH 45701, USA

Correspondence should be addressed to Gary R. Weckman; weckmang@ohio.edu

Received 11 April 2016; Accepted 29 June 2016

Academic Editor: Ozgur Kisi

Copyright © 2016 Ehsan Ardjmand et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial neural networks (ANNs) are powerful empirical approaches used to model databases with a high degree of accuracy. Despite their recognition as universal approximators, many practitioners are skeptical about adopting their routine usage due to lack of model transparency. To improve the clarity of model prediction and correct the apparent lack of comprehension, researchers have utilized a variety of methodologies to extract the underlying variable relationships within ANNs, such as sensitivity analysis (SA). The theoretical basis of local SA (that predictors are independent and inputs other than variable of interest remain “fixed” at predefined values) is challenged in global SA, where, in addition to altering the attribute of interest, the remaining predictors are varied concurrently across their respective ranges. Here, a regression-based global methodology, state-based sensitivity analysis (SBSA), is proposed for measuring the importance of predictor variables upon a modeled response within ANNs. SBSA was applied to network models of a synthetic database having a defined structure and exhibiting multicollinearity. SBSA achieved the most accurate portrayal of predictor-response relationships (compared to local SA and Connected Weights Analysis), closely approximating the actual variability of the modeled system. From this, it is anticipated that skepticisms concerning the delineation of predictor influences and their uncertainty domains upon a modeled output within ANNs will be curtailed.

1. Introduction

Variable nonlinearity, correlation, and noise are characteristic, yet inseparable components of real-world databases. Clearly, the managing of such data conditions (both singularly and synergistically) by users, while simultaneously taking into account the large degree of complexity of natural systems, remains a considerable shortcoming of many empirical modeling efforts. A powerful approach for the modeling of nonlinear functions within a defined degree of accuracy is artificial intelligence, of which artificial neural networks (ANNs) are a core technology [1]. ANNs require little (if any) expert knowledge for their application and

do not necessitate known probability distributions for variables. Multilayer feed-forward perceptrons (MLPs), the most commonly utilized ANN topology, can approximate any measurable function of an output (response) variable within a database, regardless of the nonlinear function utilized and/or the data dimensionality and variance [2].

Nonetheless, MLPs exhibit a holistic deficiency in declarative knowledge structure. Predictor-response associations are encoded incomprehensibly as weight and bias values within a multilayered topology, providing little (or no) apparent realization to users regarding network functionality of and/or knowledge extraction for the modeled process (Figure 1). This “black-box” trait remains the main constraint

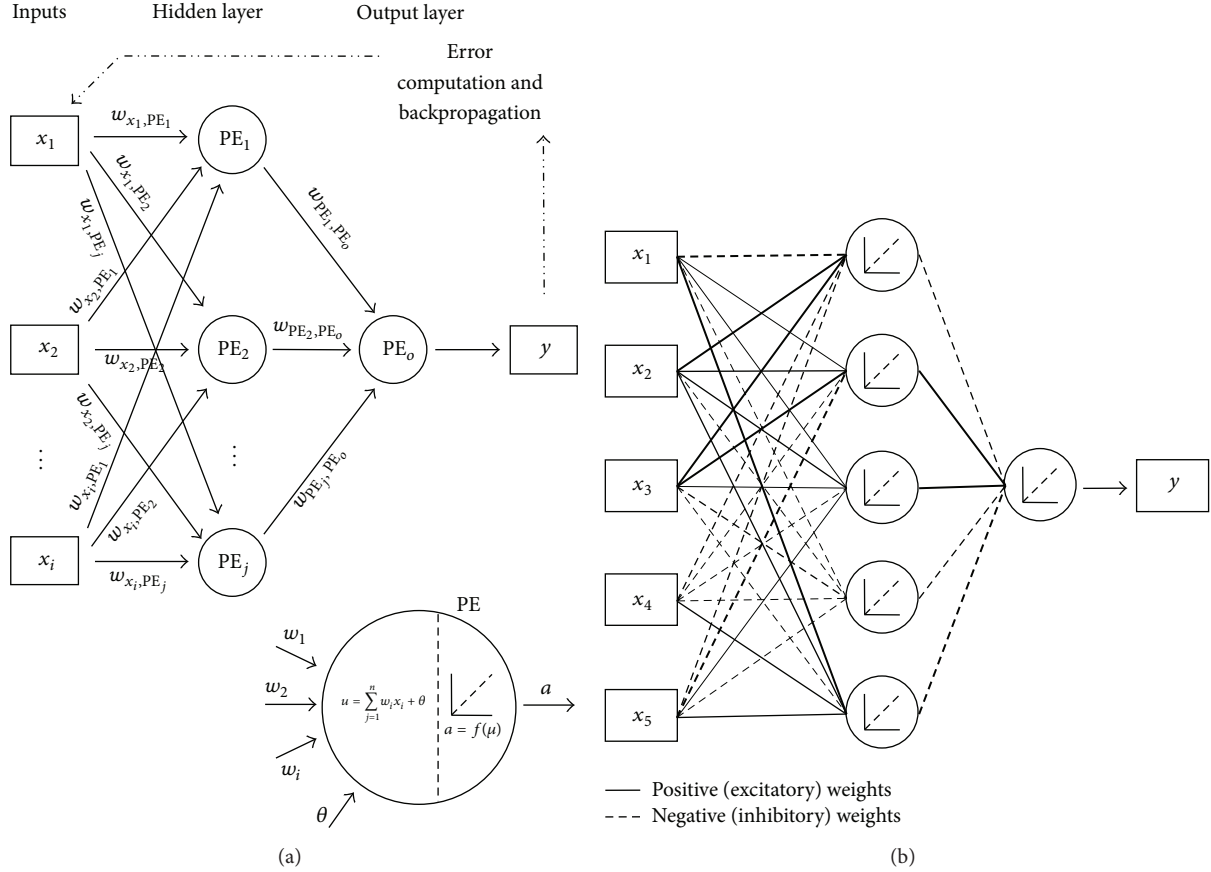


FIGURE 1: Representations of artificial neural networks. (a) A multilayer perceptron depicting interaction/influences among input variables (x_1, \dots, x_i), hidden/output layers with processing elements (PE_1, \dots, PE_j and PE_o), synaptic weights (w), and modeled output (y). (b) A representative network interpretation diagram for a representative simulation within Case Study. Line thickness portrays the relative magnitude of the weight.

to the routine utilization of ANNs within predictive analytics. Overcoming this lack of model transparency (i.e., defining input influences and their uncertainty domains upon the modeled output) has become a focus of recent machine learning research (e.g., [3–6]).

Numerous (symbolic and algorithmic) approaches have been used to identify the confidence of network prediction as well as distinguish uncertainties associated with the “cause and effect” relationships between input and output variables (e.g., [7–11]). One such approach is sensitivity analysis (SA), with methodological perspectives categorized as either qualitative or quantitative. Qualitative SA relies upon domain “expert” opinions, where individual users define, based on (his/her) opinion, the importance of predictor variables. Notably, the relative importance of singular variables cannot be quantified due to a generalized lack of expertise or when a system’s behavior is not understood *a priori*. Conversely, quantitative SAs are classified as either local or global approaches [12]. In local SA, the predictive uncertainty upon the modeled output is quantified by altering a single predictor across a predefined range, while remaining predictors are “fixed” at predefined values. In global SA, the singular

effects of predictors upon the output are evaluated during instances when remaining inputs are varied concurrently across predefined ranges (for reviews, see [13–15]).

However, local SA ignores conditions of model nonlinearity and the lack of independence and multicollinearity among database variables. Model linearity and independence between/among multiple input variables often are not valid assumptions; as a consequence, altering the value of one predictor variable will affect (the values of) other inputs and, in turn, impact the predictive power (and uncertainty) of a response variable. For example, assume that the meteorological and hydrological variables, ambient (air) and water temperatures, respectively, are used as predictors within an ANN attempting to model a water-quality response (e.g., water clarity arising from biotic abundance [16, 17]). Because ambient and water temperatures typically would be highly correlated (yet not perfectly related due to distinct physical properties of water and air), analyzing the network’s predictive uncertainty in regard to ambient temperature alone, while keeping the value of water temperature static, is neither appropriate nor logical. In this simplistic example, the assumption of independence and nonassociation between

these two inputs imposes an unrealistic portrayal of model capability/uncertainty, a situation that becomes more perplexing when, in addition to correlative processes, database input/output relationships are nonlinear.

Here, a global state-based sensitivity analysis (SBSA) is proposed. For SBSA, input vectors associated with a predictor variable initially were assigned to a bin (or “state”) along a distribution range where every “state” was equivalent to a predetermined variance of that distribution. Quantitative uncertainties for individual predictors upon the modeled response then were estimated as a function of dithered alterations across its data range. Importantly, the values of remaining predictors were allowed to vary concurrently across respective data ranges and in correspondence to those of the predictor of interest.

2. Local Sensitivity Analysis

In local SA, the effect of altering the value of a singular predictor upon the response variable(s) is evaluated, while values of other inputs remain unchanging (often fixed at values of the 1st quartile, median, or mean [18]). Typically, inputs not of interest are held at their respective mean values; as such, this methodology often is termed “sensitivity about-the-mean analysis” [14]. Local SA commonly is used within statistical/mechanistic modeling practices due to the ease in its computation and the simplicity of its comprehension [14, 19]. Yet, despite the straightforwardness of its calculation and interpretation, local SA is appropriate only during instances when inputs are independent. In actuality, predictor variables typically display dependability upon and correspondence with each other (fully or in part). As such, “fixing” values of multiple predictors while examining the effect of a single “changing” variable is neither logical nor representative of real-world systems.

To illustrate the functionality of this approach, consider a feed-forward backpropagation MLP comprised of inputs ($x_{1,...,I}$), multiple processing elements ($PE_{j,...,J}$) in one hidden layer (HL), and one output (y). The sensitivity, Sen_i , of output y with respect to input x_i is defined as ([20]; refer to Figure 1(a) for notation)

$$\begin{aligned} Sen_i &= \frac{\partial y}{\partial x_i}, \\ Sen_i &= y' \sum_{j=1}^J w_{PE_j, PE_o} \left| \frac{\partial PE_j}{\partial x_i} \right|, \\ Sen_i &= y' \sum_{j=1}^J w_{x_i, PE_j} PE'_j w_{PE_j, PE_o}, \end{aligned} \quad (1)$$

where PE'_j and y'_j are the derivative values of the HL and output activation function and w_{x_i, PE_j} and w_{PE_j, PE_o} are the weights between the input-hidden and hidden-output layers, respectively.

3. Global Sensitivity Analysis

In global SA, the variation of a singular input variable is considered while the values of other predictors also are allowed to vary [12]. In this manner, the comprehensive effects of predictor variables upon the modeled output(s) are assessed [21], lending to immediate advantages of global approaches (over local SA), most notably, the inclusion of predictor influences in terms of range and shape for input probability density functions and the multidimensional “averaging” of dependencies (or correlations) among multiple inputs [14]. By assessing output uncertainties over multidimensional variable space, global SA is not limited by the conditions of variable linearity or additivity within model estimations, thereby ensuring that such attributes are accounted for within derivation of predictive uncertainty functions [21, 22]. Global SA affords validity to the delineation of importance for multiple predictors to a modeled output and, by yielding a range of model outcomes attributable to concurrent perturbation by a variable “suite,” provides knowledge about the direct (first-order) effect of an individual predictor upon the response variable(s) as well as interaction (higher order) effects [23, 24].

Diverse global SA methodologies exist (dependent upon the objective of the analysis) but can be categorized into screening and regression-based and variance-based approaches [21, 23, 25, 26]. Screening assesses the holistic importance of (select) groups of input variables upon model prediction uncertainty, often via cluster analysis in multidimensional scatter plots, but more advanced techniques have been proposed [27]. Because screening provides minimal, if any, quantitative information pertaining to predictor importance, it typically is utilized as an initial step by users to first sort and then remove variables having little importance to model prediction [28]. Regression approaches traditionally involve fitting predictor (independent) variables to a response (dependent) variable via parametric multiple regression in a complete or stepwise manner [21]. By utilizing standardized regression coefficients within the regression model (note that standardizing the regression coefficients parameterizes the variance of independent variables to values of one, i.e., a correlation coefficient), users can assess the number and direction of standard deviations that a dependent variable will be altered, per standard deviation changes of predictor variables. Population resampling via Monte Carlo simulation followed by regression modeling affords delineation of inconsistencies in predictor effects, while circumventing difficulties with conditions of model nonlinearity and nonadditivity [29]. Variance-based approaches evaluate predictor uncertainty upon the response variable as probability distributions, from which users partition (and ultimately quantify) the response variance attributable to an individual (or groups of) predictor(s). Variance decomposition-based methodologies commonly utilized in science and engineering applications include the analysis of variance for linear models, the Fourier amplitude sensitivity test, and Sobol [30] indices for nonlinear models [31–33]. Nevertheless, the computational “burden” typically associated with population resampling and variance-based algorithms limits their routine usage.

4. Complementary Analyses to Sensitivity Analysis

Techniques complimenting SA in identifying the interactions and/or predictive uncertainties within (trained) ANNs include Neural Interpretation Diagrams (NIDs) and Connected Weights Analysis (CWA). NIDs provide a qualitative depiction of contrasting weight strength and direction among neurons/nodes within a network's topology whereas CWA incorporates a quantitative contribution of synaptic weights among input-HL-output layers, with predictors having greater values considered more influential in describing the behavior of the modeled system. Yet, excitatory/inhibitory influences of predictor variables are not denoted within calculations, and the extent of predictive uncertainties for input-output data ranges cannot be inferred from final values. Accordingly, explanatory interpretation of calculations by users is limited to the holistic magnitude of the impact of predictors upon the modeled response.

4.1. Neural Interpretation Diagram (NID). Originally introduced by S. L. Özemesi and U. Özemesi [34], a NID portrays the magnitude and direction of connection (synaptic) weights between/among network neurons and nodes. Neurons/nodes are arranged in cascading layers (input-hidden-output) with synaptic weight direction/values represented as lines within network depictions (Figure 1(b)). Interaction between input neurons is identified during instances when more than one “significant” connection enters a node. Justifiably, user interpretation of a NID illustrating a complex network (i.e., having numerous inputs exhibiting multicollinearity and/or multiple predictive significance with numerous PEs and HLs) can be difficult, if not impossible (e.g., [16, 35–37]).

4.2. Connected Weights Analysis (CWA). CWA quantifies NID portrayals by utilizing the final synaptic weight values to identify the relative share of prediction associated with model inputs [38]. For this, interacting weight quotients among neurons/nodes are summed, with the predictors having the greatest relative value (in terms of connecting weights to output) deemed the most influential in describing the holistic behavior of the modeled outcome. As an example, users can consider the aforementioned MLP (Figure 1(a)), a single modeled output (y) arising from five predictor variables ($x_{i,...,I}$) via a single HL having five PEs ($PE_{j,...,J}$). The absolute contribution of any one input variable ($C_{x_{i,...,I}}$) is calculated from the final trained weight values (w) as (refer to Figure 1(a) for notation)

$$C_{x_{i,...,I}} = \sum_{j=1}^J \left(\left| w_{x_{i,...,I}, PE_{j,...,J}} \cdot w_{PE_{j,...,J}, PE_o} \right| \right), \quad (2)$$

where $w_{x_{i,...,I}, PE_{j,...,J}}$ and $w_{PE_{j,...,J}, PE_o}$ are the products of the input-HL and HL-output connections, respectively. The relative contribution for a single input (e.g., $RC_{x_{i,...,I}}$) then is

determined from the summation of all input-HL-output contributions (SC_x) as

$$SC_x = \sum_{x=i}^I (C_{x_{i,...,I}}), \quad (3)$$

$$RC_{x_{i,...,I}} = \frac{C_{x_{i,...,I}}}{SC_x}.$$

5. Derivation of State-Based Sensitivity Analysis

In natural systems, a variable's numerical value lies within a predetermined interval, hereafter a variable “state,” expressed from the mean (μ) and standard deviation (σ) of its sample population (Figure 2). Because variable “states” are a function of distinct variances within individual (distinct) sample populations, the number of states and associated numerical ranges among multiple distributions will differ. Given this, the correspondences among variables within distinct states of their sample ranges also will differ. Local SA, where the value of predictors (besides that of the variable of interest) is fixed at their mean value, disregards this principle. For SBSA, the sensitivity of an output (y) corresponds to an input variable (x_i) within its sth state. For x_i , sensitivity (Sen_i^s) becomes

$$Sen_i^s = \frac{\partial y}{\partial x_i^s} \quad (4)$$

or

$$Sen_i^s = y' \sum_{j=1}^J w_{oj} \frac{\partial y_j}{\partial x_i^s} \quad (5)$$

with values for other inputs remaining fixed (to mean values) within respective states corresponding to x_i . By partitioning the (changing) range of values for inputs into s intervals, x_i^s becomes the value of x_i belonging to the sth interval.

As an example, users can consider a hypothetical database (Table 1) having five input variables ($x_{1,...,5}$) and one output variable (y). The data range for x_1 , with mean (μ_{x_1}) and standard deviation (σ_{x_1}) values for variable x_1 of -0.2281 and 0.9144 , respectively, partitioned into twelve “states” (but any number could be utilized), becomes

- State 1: $[\mu_{x_1} - 3\sigma_{x_1}, \mu_{x_1} - 2.5\sigma_{x_1}] = [-2.9713, -2.5141]$,
- State 2: $[\mu_{x_1} - 2.5\sigma_{x_1}, \mu_{x_1} - 2\sigma_{x_1}] = [-2.5141, -2.0569]$,
- State 3: $[\mu_{x_1} - 2\sigma_{x_1}, \mu_{x_1} - 1.5\sigma_{x_1}] = [-2.0569, -1.5997]$,
- State 4: $[\mu_{x_1} - 1.5\sigma_{x_1}, \mu_{x_1} - \sigma_{x_1}] = [-1.5997, -1.1425]$,
- State 5: $[\mu_{x_1} - \sigma_{x_1}, \mu_{x_1} - 0.5\sigma_{x_1}] = [-1.1425, -0.6853]$,
- State 6: $[\mu_{x_1} - 0.5\sigma_{x_1}, \mu_{x_1}] = [-0.6853, -0.2281]$,
- State 7: $[\mu_{x_1}, \mu_{x_1} + 0.5\sigma_{x_1}] = [-0.2281, 0.2291]$,
- State 8: $[\mu_{x_1} + 0.5\sigma_{x_1}, \mu_{x_1} + \sigma_{x_1}] = [0.2291, 0.6863]$,
- State 9: $[\mu_{x_1} + \sigma_{x_1}, \mu_{x_1} + 1.5\sigma_{x_1}] = [0.6863, 1.1435]$,
- State 10: $[\mu_{x_1} + 1.5\sigma_{x_1}, \mu_{x_1} + 2\sigma_{x_1}] = [1.1435, 1.6007]$,
- State 11: $[\mu_{x_1} + 2\sigma_{x_1}, \mu_{x_1} + 2.5\sigma_{x_1}] = [1.6007, 2.0579]$,
- State 12: $[\mu_{x_1} + 2.5\sigma_{x_1}, \mu_{x_1} + 3\sigma_{x_1}] = [2.0579, 2.5151]$,

TABLE 1: Hypothetical database for an artificial neural network. Note that data are not normalized, with vectors sorted based on descending values of x_1 .

x_1	x_2	x_3	x_4	x_5	y
-1.80	-0.41	-0.71	1.01	0.36	-2.03
-1.77	-0.78	0.76	-1.79	0.02	-1.14
-1.53	-0.64	0.67	-0.24	0.69	-1.43
-1.50	0.62	1.36	-0.02	1.35	-0.99
-0.71	-1.90	-2.05	-1.23	-0.22	-1.66
-0.70	-2.37	-1.19	-1.08	-0.58	-1.78
-0.50	-0.81	0.56	-0.64	-1.42	-0.10
-0.38	-0.04	-0.24	-0.40	-1.54	-0.18
-0.31	0.46	-0.36	1.14	0.84	-0.51
-0.16	0.84	-0.58	-0.06	0.60	-0.56
-0.13	-0.36	-0.21	-0.35	1.19	-0.36

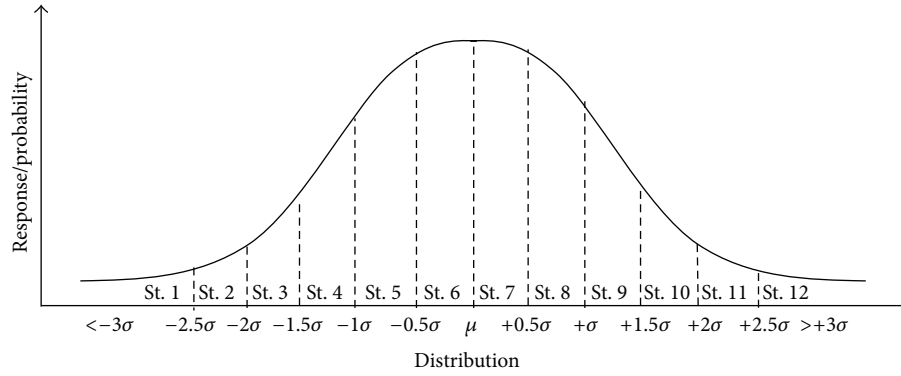


FIGURE 2: Subjective data bins (or “states”) arising from standard deviation intervals for a variable (x) displaying a normal (Gaussian) distribution.

with similar calculations completed for other variables. The relationships among multiple “states” for variable x_1 with corresponding states of other variables then were tabulated (Table 2). Calculations for the sensitivity of an output in relation to alterations of x_1 initiated from the center of the most negative state for x_1 (i.e., state three), with the value of x_1 increasing and values for other variables adjusted to their corresponding “state” mean value (Table 3).

Briefly, when values of x_1 occur within a particular state (e.g., state eight; Table 2), values for other input variables may occur within multiple, distinct states of their respective distribution (e.g., values of x_2 occur within state six and single occurrences in each of states seven, nine, and eleven, with other variables following suit). The value for x_2 would be computed as the weighted mean value for the differing states (i.e., states six, seven, nine, and eleven, with the value for state six tallied three times). Because x_1 was considered to be distributed normally and the associated variable states were based on the sample population mean and standard deviation, no values for x_1 occurred within states one, two, ten, and twelve (see Tables 2 and 3). In this manner, values for input variables, x_2 to x_5 , other than the variable of interest (x_1) within SBSA calculations will vary considerably

from those utilized in local SA, thereby resulting in distinct modeled output values (Table 4).

6. Case Study

An artificial database having a defined structure, consisting of 10,000 data vectors (or “exemplars”) with five predictor variables (x_1, x_2, x_3, x_4, x_5) and one dependent variable (y), was obtained from Olden et al. (University of Washington, Seattle, WA, USA; refer to [8]). The correlation structure for dependent and independent variables was fixed ($r_{y \cdot x_1} = 0.8$, $r_{y \cdot x_2} = 0.6$, $r_{y \cdot x_3} = 0.4$, $r_{y \cdot x_4} = 0.2$, and $r_{y \cdot x_5} = 0.0$), with a correspondence of 0.2 between any two predictors. In this manner, the predictive significance of independent variables within subsequent models progressively decreased from x_1 , the most important predictor, to x_5 , the least important. Monte Carlo simulation then was conducted to provide 50 distinct data subsets (of 150 exemplars each) from the original database.

6.1. Network Training and Testing. MLPs, consisting of predictor variables ($x_{1,\dots,i}$), multiple $PES_{(1,\dots,i)}$ in one HL layer,

and an output variable (y) arising from PE_o and utilizing a backpropagation learning algorithm, were originated using NeuroSolutions v6.31 software (NeuroDimension, Inc., Gainesville, Florida, USA) as (refer to Figure 1(a) for notation)

$$\begin{aligned}
 y = & f \left\{ w_{PE_1, PE_o} \left[f \left(w_{x_1, PE_1} \cdot x_1 + w_{x_2, PE_1} \right. \right. \right. \\
 & \cdot x_2 \cdots w_{x_i, PE_1} \cdot x_i + \varepsilon_1 \left. \left. \left. \right) \right] \right\} + f \left\{ w_{PE_2, PE_o} \left[f \left(w_{x_1, PE_2} \right. \right. \right. \\
 & \cdot x_1 + w_{x_2, PE_2} \cdot x_2 \cdots w_{x_i, PE_2} \cdot x_i + \varepsilon_2 \left. \left. \left. \right) \right] \right\} \\
 & + f \left\{ w_{PE_j, PE_o} \left[f \left(w_{x_1, PE_j} \cdot x_1 + w_{x_2, PE_j} \right. \right. \right. \\
 & \cdot x_2 \cdots w_{x_i, PE_j} \cdot x_i + \varepsilon_j \left. \left. \left. \right) \right] \right\}. \quad (6)
 \end{aligned}$$

This network architecture was identical to that used by Olden et al. [8] in order to make study comparisons possible. Transfer functions within PEs were linear (note that a network with linear transfer functions is capable of modeling nonlinear interactions among dependent variables). Exemplars within each of the data subsets were assigned randomly for model training, cross-validation, and testing (60, 15, and 25% of subset exemplars, resp.). ANNs were trained and cross-validated prior to being applied to test exemplars. The presentation of cross-validation data concurrent with training data provided for an unbiased estimation of prediction.

For training, exemplars were repeatedly presented to a network utilizing a constant momentum rate (0.7) and varied step sizes (1.0 and 0.1 for the HL and output layer, resp.), with weights adjusted after presentation of all exemplars (or “epoch”) to minimize the mean square error (MSE). The MSE was computed concurrently for the cross-validation exemplars. Training was terminated prior to the maximum of 1000 epochs if the MSE within either the training or cross-validation data began to increase (i.e., an indication that the network began to memorize the data [39, 40]). Values of MSE for both training and cross-validation data subsets converged to global minimums over multiple epochs (e.g., Figure 3(a)), indicating that MLPs provided adequate estimates of prediction (e.g., Figure 3(b)). Upon applying trained networks to testing data subsets, modeled response values closely approximated simulated (measured) values (e.g., Figure 3(c)).

6.2. Determining and Evaluating SBSA. For each simulated data subset, values of singular input variables of interest were deviated across predefined states of their data range. The remaining input variables were fixed at values corresponding to the particular state that the value of the input variable of interest resided within. When the state for the variable of interest changed, values of the remaining predictors also changed to values within their respective distributions corresponding to the “new” state for the variable of interest. Following this approach for each of the 50 simulations, weighted mean values for ancillary variables were generated for singular variables of interest within each of the designated states across its distribution (e.g., Table 5; see above). These

data then were implemented within the aforementioned trained/validated ANNs.

State-based sensitivity values were derived for all input variables (following (4) and (5)). The accuracy of ranking the input variable importance then was estimated via Gower’s coefficient [8, 41]. Briefly, for inputs ($x_{1,\dots,5}$) within a simulation subset (z), the agreement between the modeled and actual variable ranks was tabulated as a binary number (“1” if correct and “0” if incorrect). The accuracy in ranking inputs across 50 simulations then was derived as

$$\text{Gower's coefficient (as\%)} = 100 \cdot \left(\frac{1}{50} \right) \cdot \sum_{z=1}^{50} x_{1,\dots,5}. \quad (7)$$

In this manner, the correctness in ranking predictors for each of SBSA, local SA (following (1)), and CWA (following (2) and (3)) was estimated, affording the means with which to compare methodology effectiveness (i.e., accuracy). The absolute deviation of the modeled rank from the actual rank (i.e., precision) also was tabulated.

Correctness in ranking predictive significance of input variables was dissimilar among SBSA, local SA, and CWA (Figure 4), with the accuracy of SBSA (Gower’s coefficient = 0.83) being twofold greater than those provided by local SA and CWA (coefficients of 0.42 and 0.43, resp.).

SBSA correctly ranked the initial three variables having the greatest predictive significance (i.e., x_1 , x_2 , and x_3) in 95% of simulations; thereafter, accuracy diminished to approximately 65% for the rankings of x_4 and x_5 (Figure 4(a)). Notably, SBSA yielded minimal deviations in estimation of rank importance across variables (Figure 4(b)). Local SA was comparable with SBSA in ranking the two most significant predictors (x_1 , x_2). However, the ranking of x_3 correctly occurred in only 20% of simulations with zero instances of correctness for inputs x_4 and x_5 . In contrast, CWA correctly ranked x_1 and x_2 , in only 84 and 52% of instances, respectively. Accuracy subsequently diminished to approximately 28% for inputs x_4 and x_5 . Notwithstanding the substantial deviation in ranking x_3 by local SA (see Figure 4(b)), CWA produced the greatest holistic deviation of modeled rankings from actual rankings (i.e., the least precision) for predictors.

7. Discussion

Through delineation of inputs having the greatest impact upon the variability of modeled outcomes, SA offers an interpretable metric of how well predictor variables embody the behavior of the modeled process. Because SA is used to characterize regions of maximal variation within a model domain, it can be considered a byproduct of uncertainty analysis [42]. Local SA is easy to manipulate through minimal computational effort; as such, it has had universal, ostensibly indiscriminate application within statistical and mechanistic modeling practices spanning diverse disciplines [14, 19]. Yet, local SA only provides information concerning how network outcomes may (or may not) be altered via deviation of a single predictor. Such a “one parameter at a time” approach is realistic only when the relationships between predictor

TABLE 5: Resultant mean values for input variables, x_2 to x_5 , and the modeled output, y , within the case study simulation, #37 (see Figure 3), when values of x_1 lie within distinct “states” across its distribution (refer to Figure 2). n : the instances of data vectors within a distinct x_1 “state” (total $n = 150$).

n	State of x_1	x_2	x_3	x_4	x_5	y
2	-2.5	0.000	-0.313	-0.250	-0.313	-1.750
7	-1.75	-0.446	-0.161	-0.286	-0.268	-1.536
18	-1.25	0.118	-0.167	0.125	-0.375	-0.771
9	-0.875	-0.653	-0.083	-0.167	0.194	-1.111
10	-0.625	-0.300	-0.538	-0.050	-0.438	-0.575
18	-0.375	-0.007	-0.382	-0.458	0.111	-0.347
13	-0.125	-0.154	0.269	0.173	-0.115	-0.058
15	0.125	0.158	0.042	-0.383	-0.083	0.258
12	0.375	0.031	-0.031	0.406	-0.469	0.458
11	0.625	0.091	0.045	-0.250	0.011	0.420
8	0.875	0.281	0.297	0.578	0.344	0.797
17	1.25	0.096	0.272	0.125	0.434	0.838
6	1.75	0.313	0.500	0.021	0.313	1.500
4	2.5	0.344	0.844	0.656	1.281	1.688

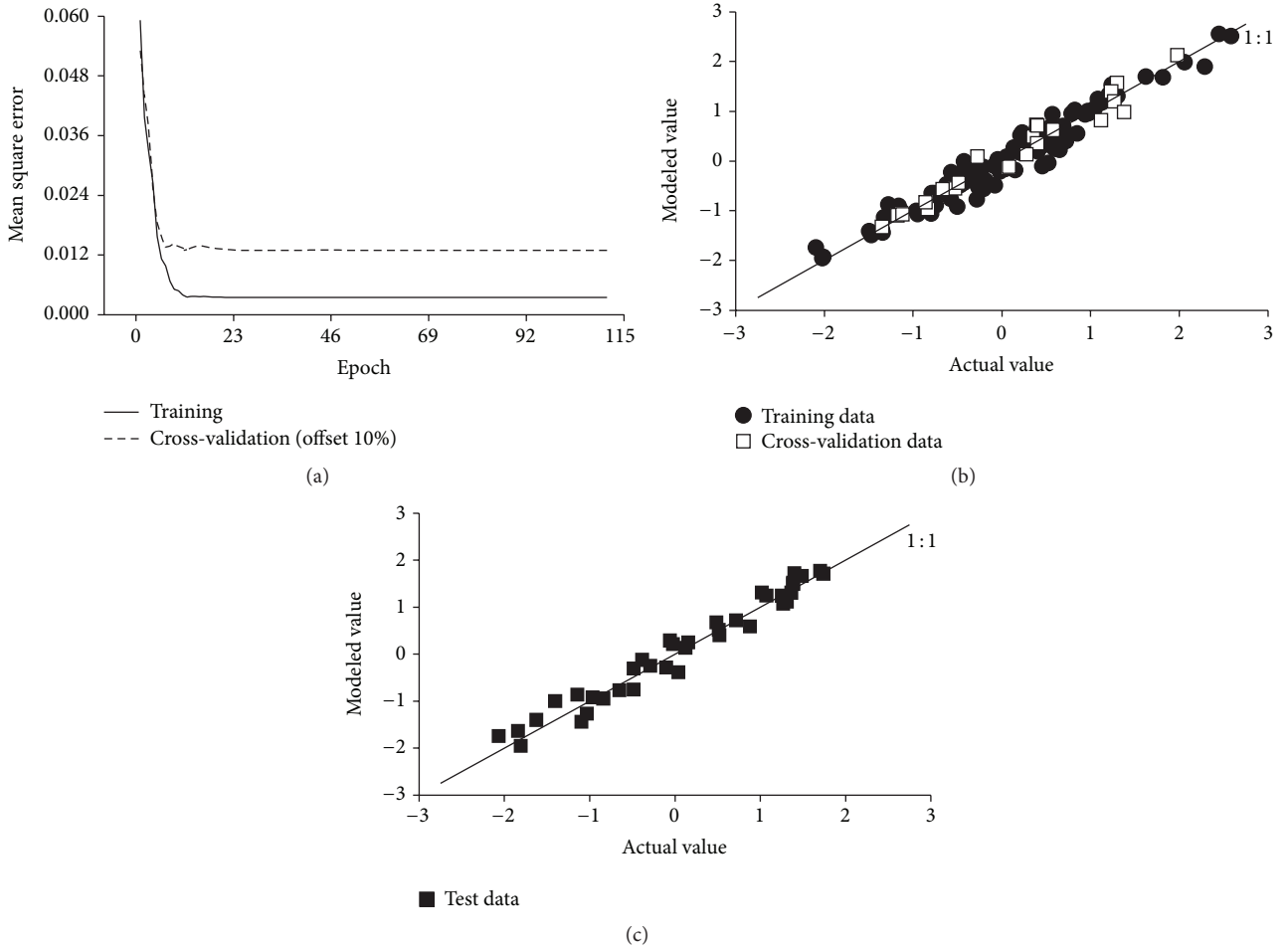


FIGURE 3: Outcomes of a multilayer perceptron for a representative simulation (Model 37; see text) within Case Study. (a) Mean square error associated with training and cross-validation data subsets during network training. (b and c) Modeled values as a function of actual values in (b) training and cross-validation data ($n = 90$ and 22 , resp.). (c) Test data ($n = 38$) subsets. The dashed line represents a 1:1 relationship.

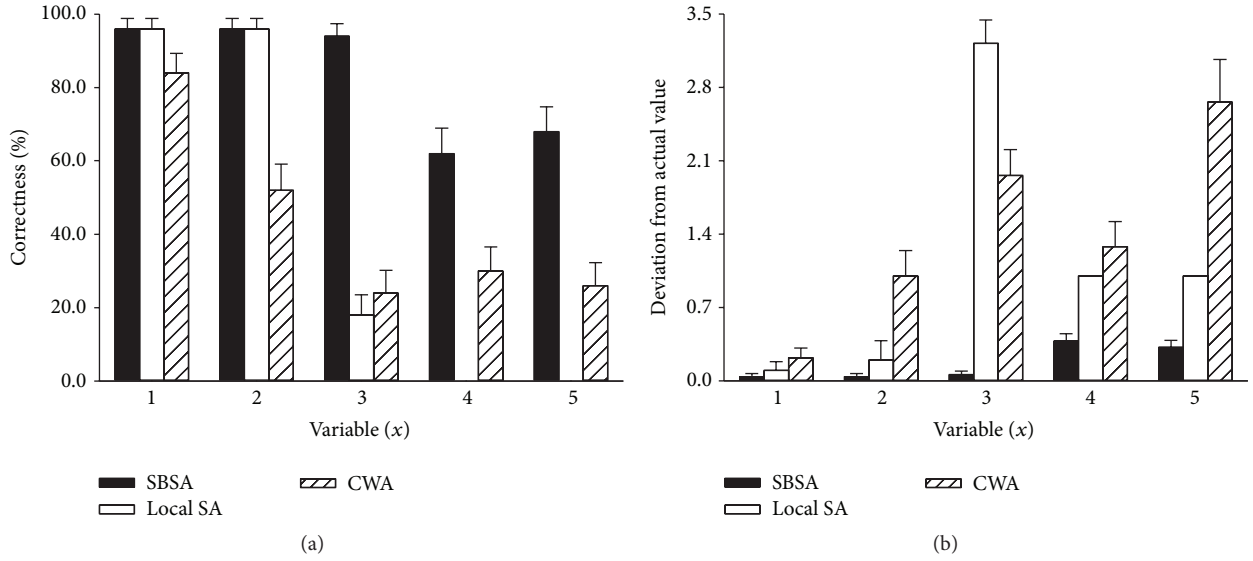


FIGURE 4: (a) Correctness of state-based sensitivity, local sensitivity, and connected weights analyses (SBSA, local SA, and CWA, resp.) in ranking singular variables within the Case Study, as compared to actual variable importance ranks. (b) Absolute deviations of modeled rankings (arising from SBSA, local SA, and CWA) for variables from actual rankings. Data are means \pm standard error; $n = 50$.

and response variables are linear, and/or variable averaging is reasonable for the modeled process [13, 43].

The proposed SBSA incorporated a global perspective where values of all input variables deviated, corresponding to the variable of interest, yet retained the computational ease and simplicity of comprehension for local SA. From this, SBSA not only affords the identification of interacting predictors (or potentially insignificant predictors) within networks, but also provides a quantitative means upon which users can base network optimization, validation, and parsimony via variable “pruning” [13, 31, 44]. Notably, SBSA retained the generalized correlation structure among predictors, enabling direct absolute comparison(s) of interacting predictor importance. In doing so, SBSA provided a realistic representation of the predictive influences for multiple inputs (with minimal deviation in the predictive significance of inputs; see Figure 4) while negating the unrealistic premise underlying local SA (i.e., altering one input while keeping values of other attributes “fixed”).

To assess the efficiency of SBSA, results arising from an artificial database having correlated predictors were compared to those generated via local SA and CWA. SBSA provided an accurate and precise portrayal of holistic variable interactions (cf. [27, 45]) and, as such, a superior approximation of database variability to those achieved via the alternative methodologies. The SBSA application was demonstrated in conjunction with artificial intelligence due to the superior capabilities of ANNs in modeling complex, nonlinear data relationships. Quantifying the importance(s) of input variables upon a modeled output is a significant problem that users must address when identifying the functionality and interpreting the predictive uncertainty of a trained network. The accounting of interaction among

database variables and the robustness of SBSA for computing and subsequently ranking the predictive importance of correlated network predictors was impressive. Moreover, the ability to extend SA applications to predictor variable states where insufficient input data existed (see Table 3) was a benefit. Though originally derived for ANN applications, SBSA also is applicable for defining complex, interacting, variable relationships within alternative parametric modeling efforts (e.g., multiple linear regression).

SBSA is an extended, albeit global derivation of traditional SA and can be classified as a regression-based approach (i.e., based upon multiple linear regression of the response variable upon the predictor vector). Similar to local SA, SBSA typically requires minimal analytic effort. Yet, depending upon the number of distinct states chosen for characterization of a variable’s data range, the computational complexity for SBSA can become cumbersome. As users increase the number of variable states, calculations become more complicated (and with a large number of states potentially unwieldy). Conversely, as the number of variable states decreases, SBSA becomes more manageable and increasingly approximates local SA methodology. In instances where only one variable state is recognized, SBSA will emulate local SA.

Multiple dissimilar databases typically have been utilized in studies intent on denoting differences among sensitivity methodologies (e.g., [3, 4, 7, 18, 38, 46–48]). Olden et al. [8] utilized a similar approach to that presented here (i.e., a synthetic database where predictor-response relationships were known *a priori* and used to evaluate multiple analytics estimating predictor importance for a trained ANN). In evaluating the effectiveness of genetic algorithms, partial derivatives, input perturbation, the profile method, SA by forward-backward stepwise addition/elimination, and

Garson's Algorithm for differentiating predictor influences, Olden et al. [8] stated that, due to disparate data structures, a user's ability to identify the "best" analytic for assessing predictor importance among studies is difficult. Nevertheless, it was concluded that CWA provided "... the best overall performance compared to the other approaches in terms of its accuracy (i.e. the degree of similarity between true and estimated variable ranks) and precision (i.e. the degree of variation in accuracy) ...". Interestingly, the exact database within the aforementioned assessment was assessed here and although the accuracy of CWA rivaled that of local SA, CWA generally displayed less precision in the rankings of predictive significance across inputs. The accuracy of SBSA was twofold greater than those of local SA and CWA, yielding minimal deviation in modeled ranking (from actual) across predictors.

A disadvantage of CWA is the inability to identify the absolute influence that a predictor variable has upon modeled outcomes. Specifically, the explanatory ability of CWA is limited to ranking the relative importance of inputs that neither are based upon variable means nor provide results for summarizing (and illustrating) predictor variability within an ANN. To accomplish detection of input-output relationships of a predictor variable across its data range via CWA, practitioners would need to separate the database into distinct subsets, via mechanical means and/or Monte Carlo resampling, which reflect a single- (or multiple-) specified range(s)/deviation(s) of predictor variability. Such a technical approach is analogous to that used in local SA where the effects of deviating predictor variables by one and two standard deviations (denoting "common" and "disturbance" variation, resp.) upon the response variable can be reported [16, 49]. A similar application of CWA would require the manual "splitting" of the holistic database into multiple (representative) subsets, with derivation of separate ANNs having individualized training procedures for each data grouping.

Here, a global SA that provided a realistic insight into the significance of correlated predictor variables upon a modeled response was developed. Within its calculations, SBSA maintained predictor relationships, challenging the premise of local SA that variable attributes are linear and independent. When validated with ANN models of a synthetic database, SBSA outperformed local SA and CWA in predictive accuracy and precision. Despite the predictive capability of ANNs, most researchers entrust low confidence to their use as empirical models due to the lack of transparency in ascertaining predictor significance [38, 50]. SBSA helps to overcome this "black-box" nature of MLPs by improving upon the identification of interacting predictor influences and the delineation of their uncertainty domains (thereby simplifying interpretation of modeled outcomes). From this, it is anticipated that user skepticisms for ANNs as a common modeling practice will be reduced.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] L. N. de Castro, "Fundamentals of natural computing: an overview," *Physics of Life Reviews*, vol. 4, no. 1, pp. 1–36, 2007.
- [2] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [3] S. J. Kemp, P. Zaradic, and F. Hansen, "An approach for determining relative input parameter importance and significance in artificial neural networks," *Ecological Modelling*, vol. 204, no. 3–4, pp. 326–334, 2007.
- [4] E. Borgonovo, W. Castaings, and S. Tarantola, "Model emulation and moment-independent sensitivity analysis: an application to environmental modelling," *Environmental Modelling & Software*, vol. 34, pp. 105–115, 2012.
- [5] G. Weckman, D. Millie, C. Ganduri et al., "Knowledge extraction from the neural 'black box' in ecological monitoring," *Journal of Industrial and Systems Engineering*, vol. 3, no. 1, pp. 38–55, 2009.
- [6] D. F. Millie, G. R. Weckman, W. A. Young, J. E. Ivey, H. J. Carrick, and G. L. Fahnenstiel, "Modeling microalgal abundance with artificial neural networks: demonstration of a heuristic 'Grey-Box' to deconvolve and quantify environmental influences," *Environmental Modelling and Software*, vol. 38, pp. 27–39, 2012.
- [7] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," *Ecological Modelling*, vol. 160, no. 3, pp. 249–264, 2003.
- [8] J. D. Olden, M. K. Joy, and R. G. Death, "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data," *Ecological Modelling*, vol. 178, no. 3–4, pp. 389–397, 2004.
- [9] O. Pastor-Bárcenas, E. Soria-Olivas, J. D. Martín-Guerrero, G. Camps-Valls, J. L. Carrasco-Rodríguez, and S. Del Valle-Tascón, "Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling," *Ecological Modelling*, vol. 182, no. 2, pp. 149–158, 2005.
- [10] W.-S. Lee, Y.-S. Kwon, Y.-S. Park, T.-S. Chon, and J.-C. Yoo, "Evaluation of environmental factors to predict breeding success of Black-tailed Gulls," *Ecological Informatics*, vol. 1, no. 3, pp. 331–339, 2006.
- [11] A. M. Mouton, A. P. Dedecker, S. Lek, and P. L. M. Goethals, "Selecting variables for habitat suitability of *Asellus* (Crustacea, Isopoda) by applying input variable contribution methods to artificial neural network models," *Environmental Modeling and Assessment*, vol. 15, no. 1, pp. 65–79, 2010.
- [12] J. Cariboni, D. Gatelli, R. Liska, and A. Saltelli, "The role of sensitivity analysis in ecological modelling," *Ecological Modelling*, vol. 203, no. 1–2, pp. 167–182, 2007.
- [13] A. Saltelli, S. Tarantola, and F. Campolongo, "Sensitivity analysis as an ingredient of modeling," *Statistical Science*, vol. 15, no. 4, pp. 377–395, 2000.
- [14] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice*, John Wiley & Sons, Sussex, UK, 2004.
- [15] A. Saltelli, M. Ratto, T. Andres et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Chichester, UK, 2008.
- [16] D. F. Millie, G. R. Weckman, W. A. Young et al., "Coastal 'big data' and nature-inspired computation: prediction potentials, uncertainties, and knowledge derivation of neural networks for an algal metric," *Estuarine, Coastal and Shelf Science*, vol. 125, pp. 57–67, 2013.
- [17] D. F. Millie, G. R. Weckman, G. L. Fahnenstiel et al., "Using artificial intelligence for cyanoHAB niche modeling: discovery

- and visualization of Microcystis–environmental associations within western Lake Erie,” *Canadian Journal of Fisheries & Aquatic Sciences*, vol. 71, no. 11, pp. 1642–1654, 2014.
- [18] S. Lek, A. Belaud, P. Baran, I. Dimopoulos, and M. Delacoste, “Role of some environmental variables in trout abundance models using neural networks,” *Aquatic Living Resources*, vol. 9, no. 1, pp. 23–29, 1996.
- [19] D. Yeung, I. Cloete, D. Shi, and W. Ng, *Sensitivity Analysis for Neural Networks*, Neural Computing Series, Springer, Berlin, Germany, 2010.
- [20] A. P. Engelbrecht, I. Cloete, and J. M. Zurada, “Determining the significance of input parameters using sensitivity analysis,” in *From Natural to Artificial Neural Computation*, J. Mira and F. Sandoval, Eds., vol. 930 of *Lecture Notes in Computer Science*, pp. 382–388, 1995.
- [21] M. Wang, J. Wright, R. Buswell, and A. Brownlee, “A comparison of approaches to stepwise regression for global sensitivity analysis used with evolutionary optimization,” in *Proceedings of the 13th Conference of International Building Performance Simulation Association (BS '13)*, pp. 2551–2558, Chambéry, France, 2013.
- [22] S. Wagner, “Global sensitivity analysis of predictor models in software engineering,” in *Proceedings of the IEEE International Workshop on Predictor Models in Software Engineering, PROMISE'07: ICSE Workshops 2007*, Minneapolis, Minn, USA, May 2007.
- [23] R. Muñoz-Carpena, Z. Zajac, and Y.-M. Kuo, “Global sensitivity and uncertainty analyses of the water quality model VFMOD-W,” *Transactions of the ASABE*, vol. 50, no. 5, pp. 1719–1732, 2007.
- [24] E. Kent, S. Neumann, U. Kummer, and P. Mendes, “What can we learn from global sensitivity analysis of biochemical systems?” *PLoS ONE*, vol. 8, no. 11, Article ID e79244, 2013.
- [25] R. Confalonieri, G. Bellocchi, S. Bregaglio, M. Donatelli, and M. Acutis, “Comparison of sensitivity analysis techniques: a case study with the rice model WARM,” *Ecological Modelling*, vol. 221, no. 16, pp. 1897–1906, 2010.
- [26] J. Yang, “Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis,” *Environmental Modelling and Software*, vol. 26, no. 4, pp. 444–457, 2011.
- [27] A. Saltelli, F. Campolongo, and J. Cariboni, “Screening important inputs in models with strong interaction properties,” *Reliability Engineering and System Safety*, vol. 94, no. 7, pp. 1149–1155, 2009.
- [28] B. Iooss and P. Lemaître, “A review on global sensitivity analysis methods,” in *Uncertainty Management in Simulation-Optimization of Complex Systems*, G. Dellino and C. Meloni, Eds., vol. 59, pp. 101–122, Springer, New York, NY, USA, 2015.
- [29] J. S. Coggan, T. M. Bartol, E. Esquenazi et al., “Evidence for ectopic neurotransmission at a neuronal synapse,” *Science*, vol. 309, no. 5733, pp. 446–451, 2005.
- [30] I. Sobol, “Sensitivity analysis for non-linear mathematical models,” *Mathematical Modeling & Computational Experiment*, vol. 1, pp. 407–414, 1993.
- [31] U. Reuter and M. Liebscher, *Global Sensitivity Analysis in View of Nonlinear Structural Behaviour*, LS-DYNA Anwenderforum, Bamberg, Germany, 2008.
- [32] C.-S. Zhan, X.-M. Song, J. Xia, and C. Tong, “An efficient integrated approach for global sensitivity analysis of hydrological model parameters,” *Environmental Modelling and Software*, vol. 41, pp. 39–52, 2013.
- [33] R. I. Cukier, H. B. Levine, and K. E. Shuler, “Nonlinear sensitivity analysis of multiparameter model systems,” *Journal of Computational Physics*, vol. 26, no. 1, pp. 1–42, 1978.
- [34] S. L. Özdesmi and U. Özdesmi, “An artificial neural network approach to spatial habitat modelling with interspecific interaction,” *Ecological Modelling*, vol. 116, no. 1, pp. 15–31, 1999.
- [35] J. D. Olden, “An artificial neural network approach for studying phytoplankton succession,” *Hydrobiologia*, vol. 436, pp. 131–143, 2000.
- [36] D. F. Millie, G. R. Weckman, R. J. Pigg et al., “Modeling phytoplankton abundance in Saginaw Bay, Lake Huron: using artificial neural networks to discern functional influence of environmental variables and relevance to a great lakes observing system,” *Journal of Phycology*, vol. 42, no. 2, pp. 336–349, 2006.
- [37] D. F. Millie, G. R. Weckman, H. W. Paerl et al., “Neural net modeling of estuarine indicators: hindcasting phytoplankton biomass and net ecosystem production in the Neuse (North Carolina) and Trout (Florida) Rivers, USA,” *Ecological Indicators*, vol. 6, no. 3, pp. 589–608, 2006.
- [38] J. D. Olden and D. A. Jackson, “Illuminating the ‘black box’: a randomization approach for understanding variable contributions in artificial neural networks,” *Ecological Modelling*, vol. 154, no. 1–2, pp. 135–150, 2002.
- [39] C. Karul, S. Soyupak, A. F. Çilesiz, N. Akbay, and E. Germen, “Case studies on the use of neural networks in eutrophication modeling,” *Ecological Modelling*, vol. 134, no. 2–3, pp. 145–152, 2000.
- [40] H. Gurbuz, E. Kivrak, S. Soyupak, and S. V. Yerli, “Predicting dominant phytoplankton quantities in a reservoir by using neural networks,” *Hydrobiologia*, vol. 504, pp. 133–141, 2003.
- [41] J. C. Gower, “A general coefficient of similarity and some of its properties,” *Biometrics*, vol. 27, no. 4, pp. 857–872, 1971.
- [42] E. Smith, “Uncertainty analysis,” in *Encyclopedia of Environmental Metrics*, A. H. El-Shaarawi and W. W. Piegorsch, Eds., pp. 2283–2297, John Wiley & Sons, Chichester, UK, 2002.
- [43] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, “Sensitivity analysis practices: strategies for model-based inference,” *Reliability Engineering and System Safety*, vol. 91, no. 10–11, pp. 1109–1125, 2006.
- [44] S. Holtzman, *Intelligent Decision Systems*, Addison-Wesley, Reading, Mass, USA, 1989.
- [45] M. Gevrey, I. Dimopoulos, and S. Lek, “Two-way interaction of input variables in the sensitivity analysis of neural network models,” *Ecological Modelling*, vol. 195, no. 1–2, pp. 43–50, 2006.
- [46] I. Dimopoulos, J. Chronopoulos, A. Chronopoulou-Sereli, and S. Lek, “Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece),” *Ecological Modelling*, vol. 120, no. 2–3, pp. 157–165, 1999.
- [47] Y.-S. Park, J. Rabinovich, and S. Lek, “Sensitivity analysis and stability patterns of two-species pest models using artificial neural networks,” *Ecological Modelling*, vol. 204, no. 3–4, pp. 427–438, 2007.
- [48] C. Vasilakos, K. Kalabokidis, J. Hatzopoulos, and I. Matsinos, “Identifying wildland fire ignition factors through sensitivity analysis of a neural network,” *Natural Hazards*, vol. 50, no. 1, pp. 125–143, 2009.
- [49] K.-S. Jeong, D.-K. Kim, and G.-J. Joo, “River phytoplankton prediction model by Artificial Neural Network: model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system,” *Ecological Informatics*, vol. 1, no. 3, pp. 235–245, 2006.
- [50] S. Lek and J. F. Guégan, “Artificial neural networks as a tool in ecological modelling, an introduction,” *Ecological Modelling*, vol. 120, no. 2–3, pp. 65–73, 1999.

