Machine learning methods for better water quality prediction

Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., Elshafie, A
Journal of Hydrology, 578, 124084.

Name : Rui-Yu Zhang
Advisor : Jui-Sheng Chen
Date :2024/10/11

Outline

Why we need artificial neural networks(ANN)?

- Compared to traditional water quality models, artificial neural networks(ANN) demonstrate greater flexibility and accuracy in addressing complex water quality prediction issues. They are well suited for applications involving large datasets, achieving the goal of saving time and money.

Abstract

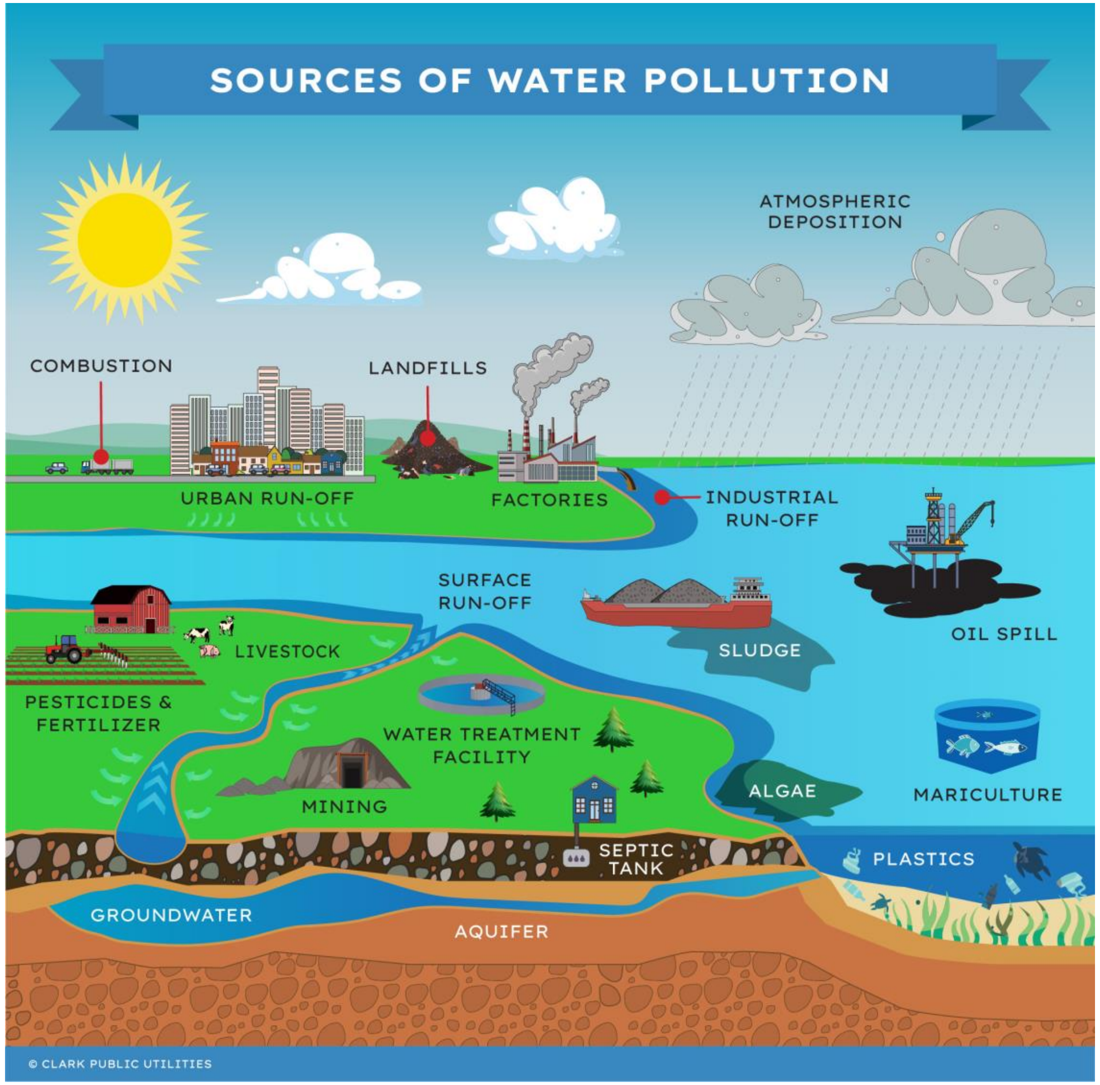
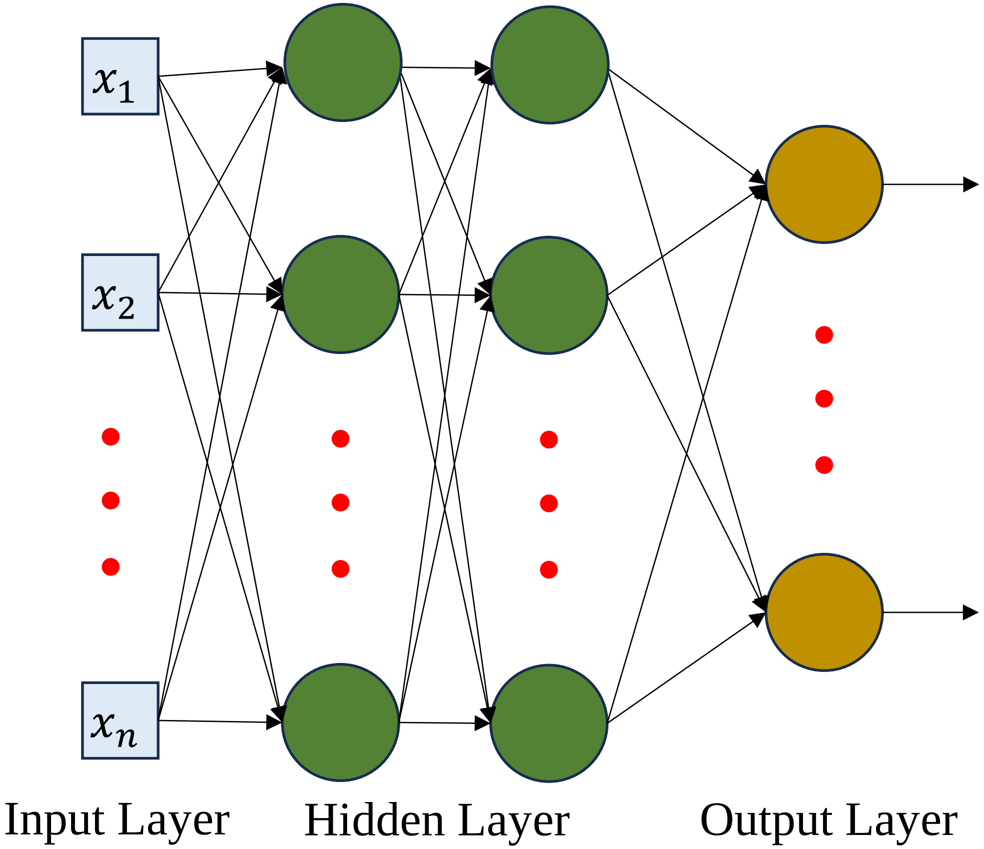


Fig. from https://powerzone.clarkpublicutilities.com/learn-about-water/quality/

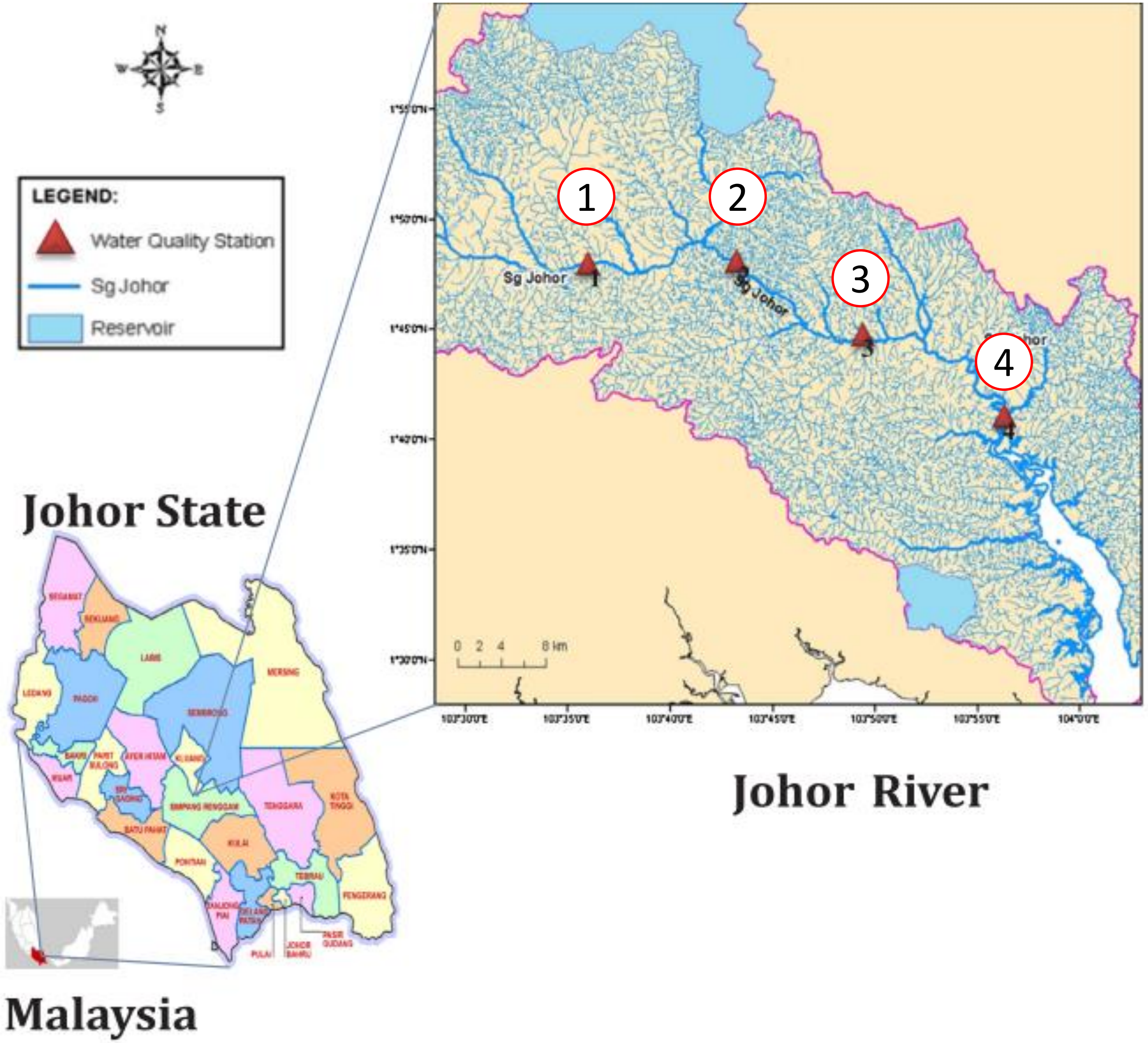


Neural Network Structure

Objective

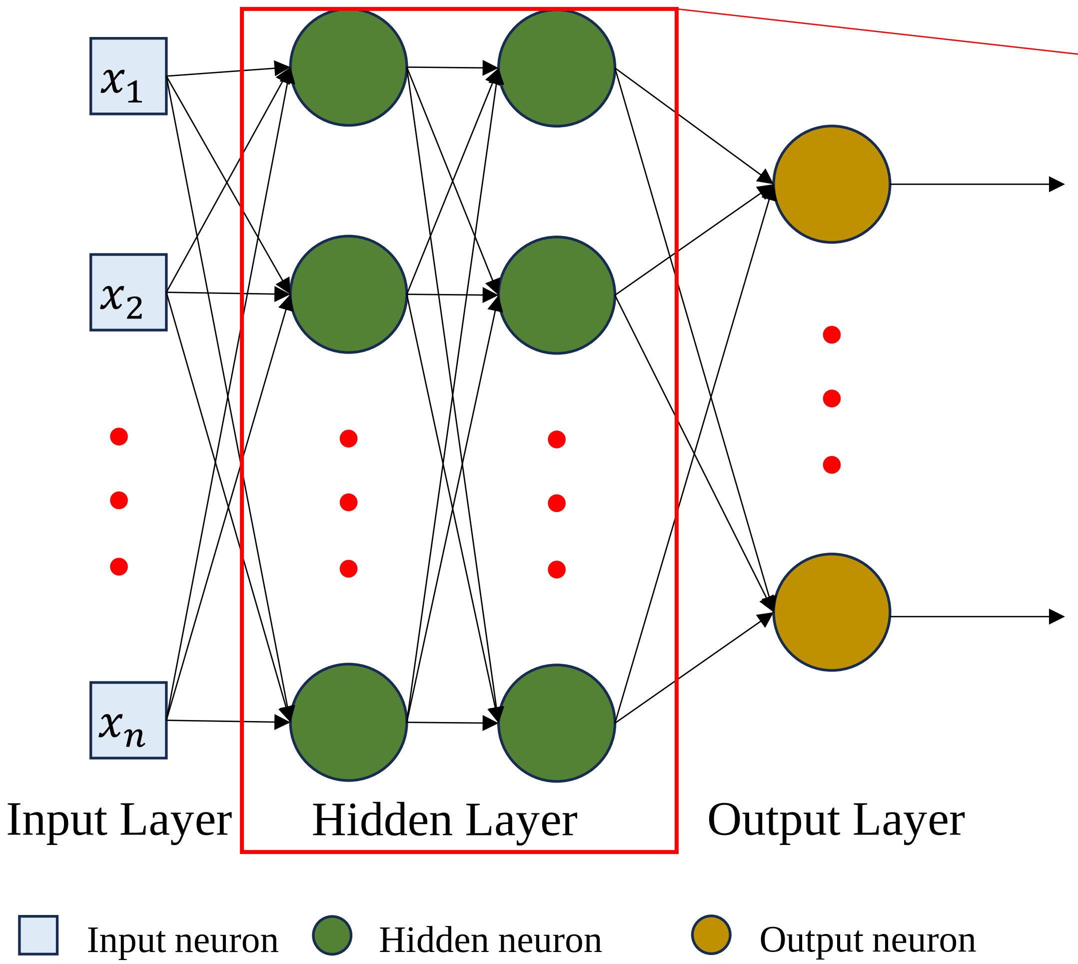
- Develop a computationally efficient and robust method for estimating water quality variables, thereby reducing the labor and cost associated with measuring these parameters.

Study area



- River basin : 2600 km^2
- Long : 123 km
- Originate : Mount Gunung Belumut and flows into the Johor Strait
- Pollution Sources: Agriculture, Household, and Industry

Multi-Layer Perceptron Neural Network(MLPNN)



- A multi-layer perceptron neural network architecture

$$y = f\left(\sum_j w_j x_j + b\right)$$

- Activation function in hidden layer

$$f = \frac{1}{1 + e^{-x}}$$

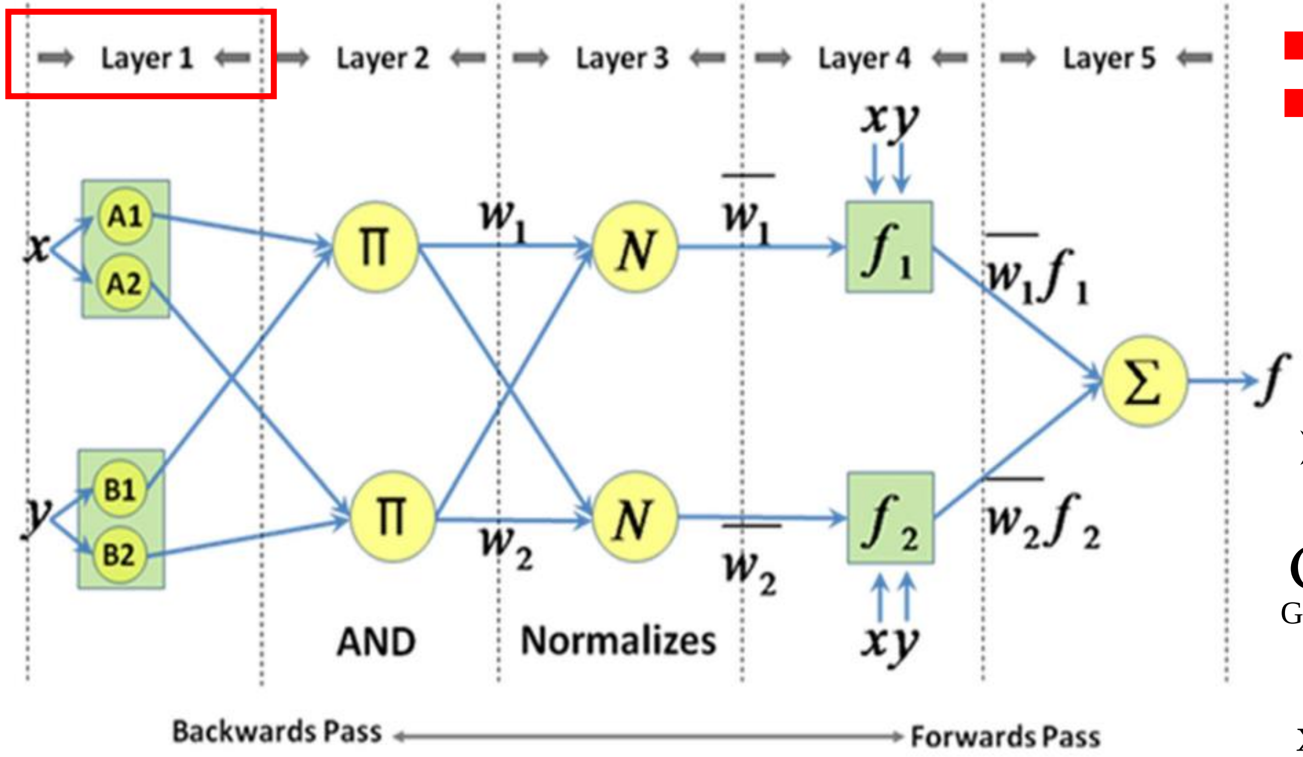
$$f = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- Activation function in output layer

$$f = x$$

x_j : output from the previous layer's j node
w_j : connection weight between the current node and j node
b : current node's bias
f : activation function

Adaptive Neuro-Fuzzy Inference System (ANFIS)



- Rule 1: $If\ x\ is\ A_1\ and\ y\ is\ B_1, then\ f_1 = p_1x + q_1y + r_1$
- Rule 2: $If\ x\ is\ A_2\ and\ y\ is\ B_2, then\ f_2 = p_2x + q_2y + r_2$

A_1, A_2 and B_1, B_2 : membership functions pertaining to inputs x and y

p_i, q_i and r_i (i = 1 or 2) : linear parameters pertaining to the model's consequent part

➢ Membership functions

$$\text{Gbell MF }(x;a,b,c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^b}$$

Generalized Bell Membership Function

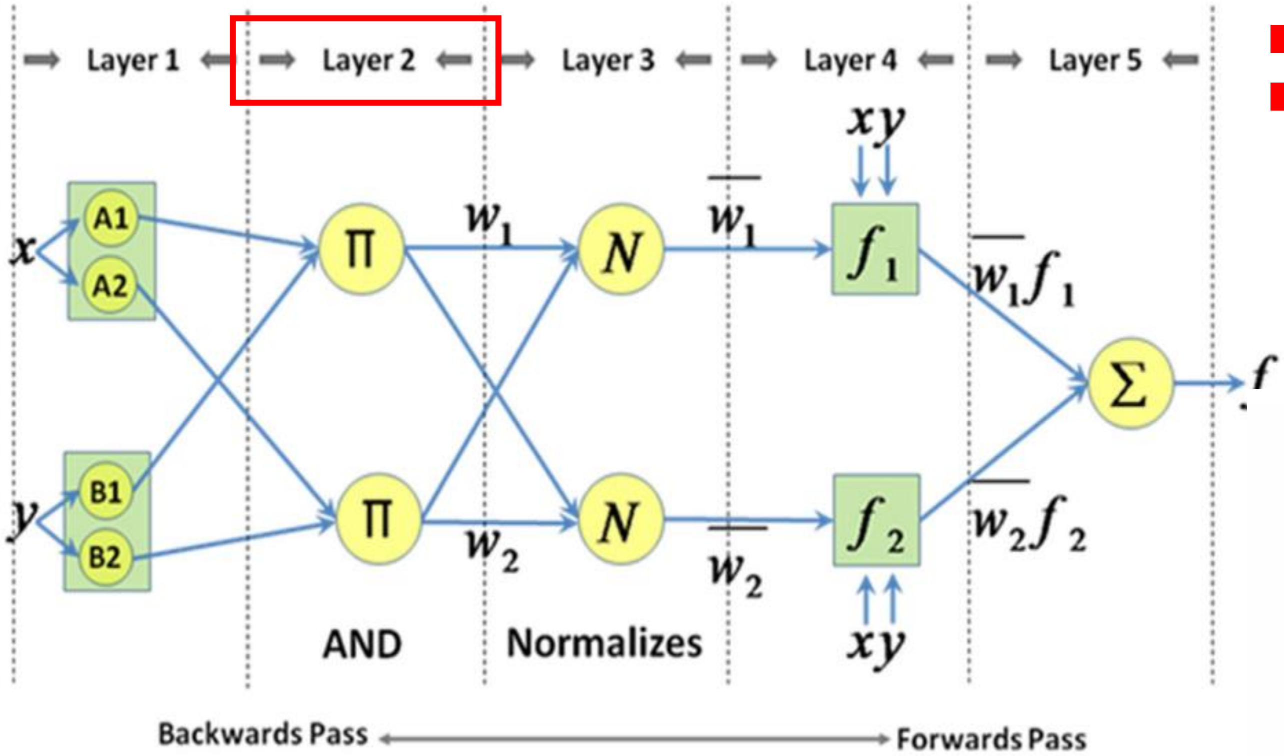
x : input variable

a : parameter controlling the width of the function

b : parameter controlling the slope of the function

c : parameter controlling the center of the function

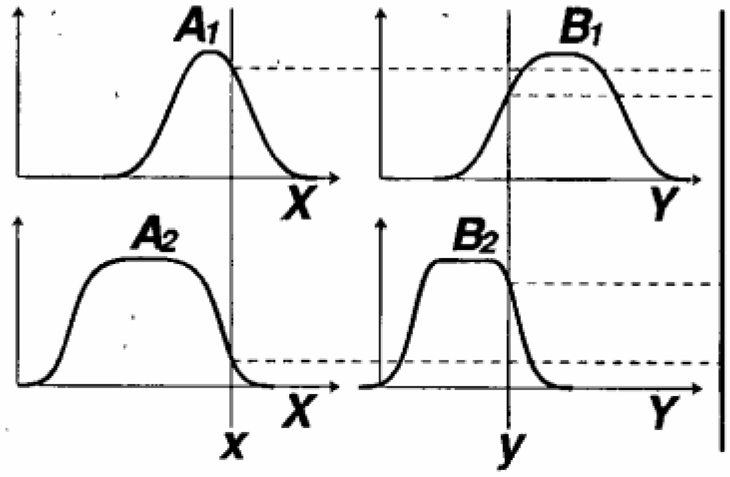
Adaptive Neuro-Fuzzy Inference System (ANFIS)



- Rule 1: $If\ x\ is\ A_1\ and\ y\ is\ B_1, then\ f_1 = p_1x + q_1y + r_1$
- Rule 2: $If\ x\ is\ A_2\ and\ y\ is\ B_2, then\ f_2 = p_2x + q_2y + r_2$

A_1, A_2 and B_1, B_2 : membership functions pertaining to inputs x and y

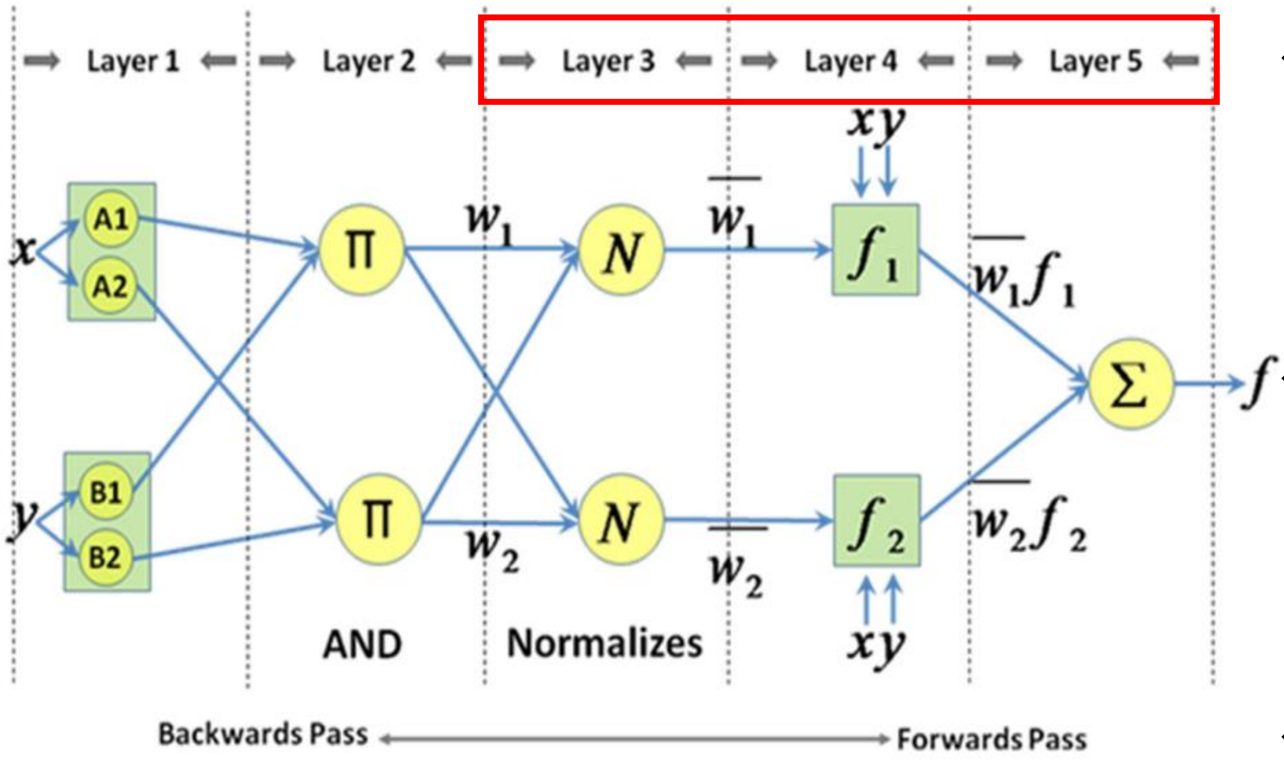
p_i, q_i and r_i (i = 1 or 2) : linear parameters pertaining to the model's consequent part



$w_1 = A_1 * B_1$

$w_2 = A_2 * B_2$

Adaptive Neuro-Fuzzy Inference System (ANFIS)



◆Layer 3

$$\overline{w_1} = \frac{w_1}{w_1 + w_2} \quad \overline{w_2} = \frac{w_2}{w_1 + w_2}$$

◆Layer 4

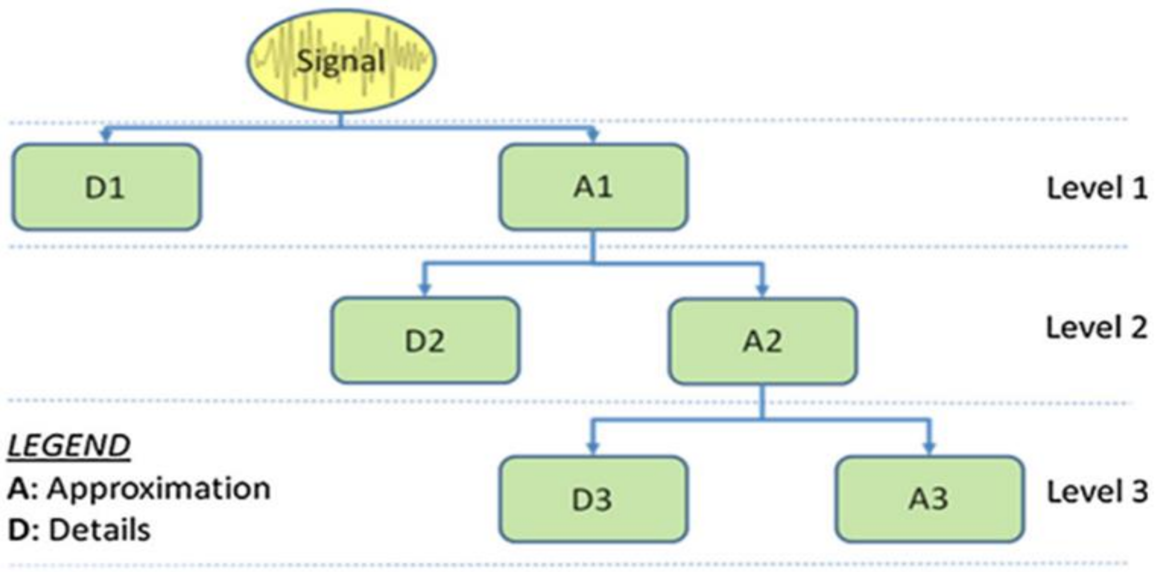
$$\overline{w_1}f_1 = \overline{w_1}(p_1x + q_1y + r_1)$$

$$\overline{w_2}f_2 = \overline{w_2}(p_2x + q_2y + r_2)$$

◆Layer 5

$$f = \frac{w_1f_1 + w_2f_2}{w_1 + w_2} = \overline{w_1}f_1 + \overline{w_2}f_2$$

Wavelet de-noising technology(WDT)



$$DWT(m,n) = 1/2\sqrt{2^m}\sum_k x[k]\psi[2^{-m}n-k]$$

ψ : mother wavelet
n : shift
m : scaling
k : shifting indices

- A schematic representation of the pyramid structure representing the WMRA

D1, D2 and D3 : high-frequency band outputs
A1, A2 and A3 : low-frequency band outputs

Input variables and data processing

Input parameters used in previous studies for the ANN model.

Author(s) and year	Input variable	Location(s)
Rabia (Koklu, 2006)	BOD, Temp, Water discharge, NO2-N, NO3-N	N/A
Kuo *et al.* (Kuo et al., 2007)	pH, Chl-a, NH4N, No3N, temp, month	Te-Chi Reservoir, Taiwan
Ying *et al.* (Zhao et al., 2007)	Turbidity, Temp, pH, Hardness, Alkalinity, Chloride, NH_4-N, NO_2-N	Yuqiao reservoir, China
Palani *et al.* (Palani et al., 2008)	DO, Chl-a, temp	Singapore coastal, Singapore
Zaqoot *et al.* (Zaqoot et al., 2009)	Conductivity, Turbidity, Temp, PH, Wind speed	Mediterranean Sea along Gaza, Palestine
Singh *et al.* (Singh et al., 2009)	pH, TS, T-AlK, T-Hard, CL, PO4, K, Na, NH4N, No_3N, COD	Gomti, India

Based on different research, a total of 12 water quality parameters were selected for ANN modeling

Temperature, Conductivity, Salinity, Turbidity, NO3, PO4, CI, K, Na, Mg, Fe and Escherichia coli

Predict water quality parameters

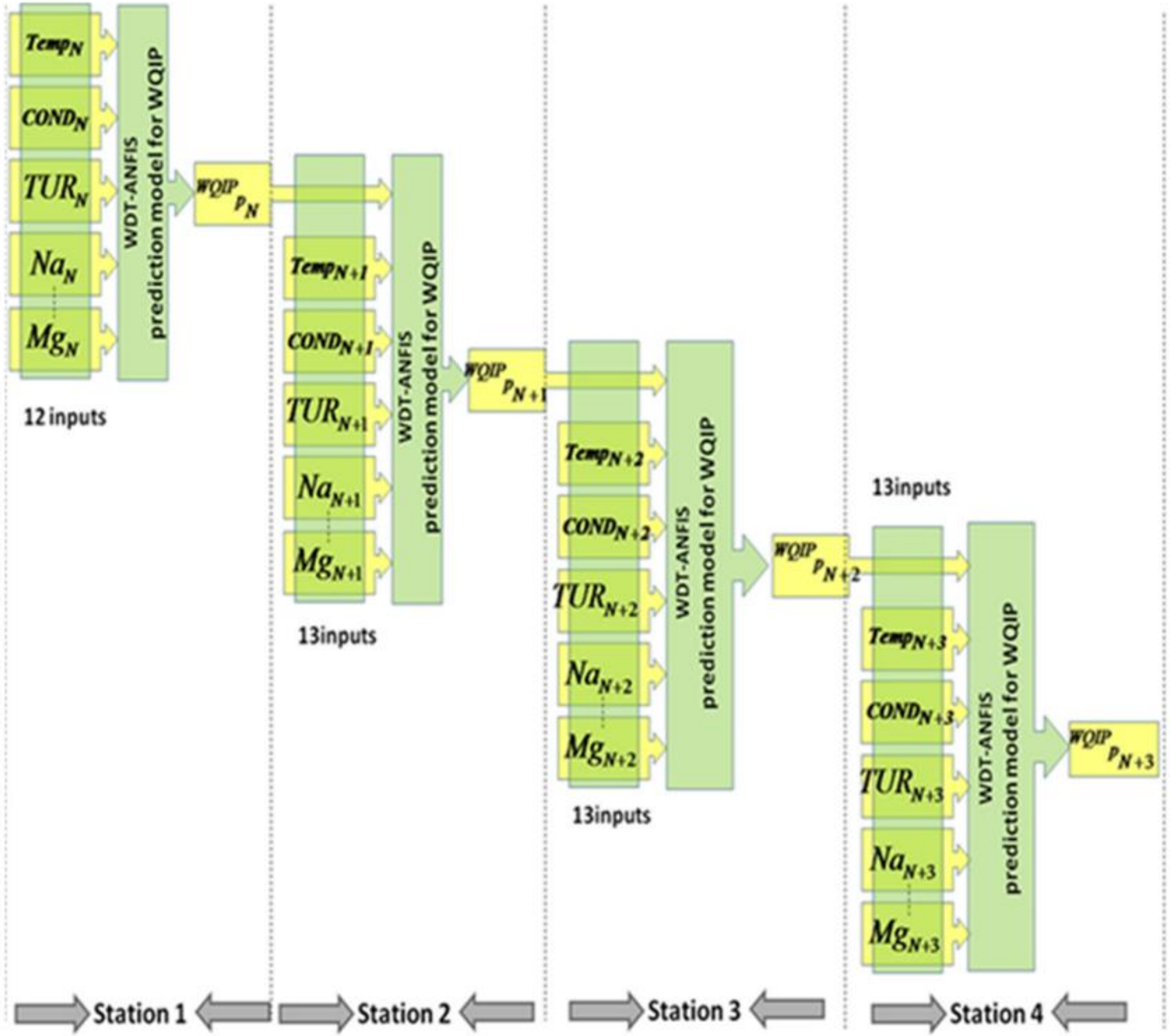
pH, Suspended solid, Ammonia nitrogen

Model performance evaluation

Name	Formula	Purpose	Range
Coefficient of efficiency (CE)	$1 - \frac{\sum_{i=1}^{n}(x_m - x_p)^2}{\sum_{i=1}^{n}(x_m - \bar{x}_m)^2}$	Evaluate the model's performance	-∞≤value≤1
Mean Square Error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(x_m - x_p)^2$	Examine the fit between the network's output and the expected output	0≤value<∞

n : the number of observations
x_m and x_p : measured and predicted parameters
$\bar{x}_m$: average of measured parameter

Different scenarios



- Schematic representation of the proposed networks for Scenario 2

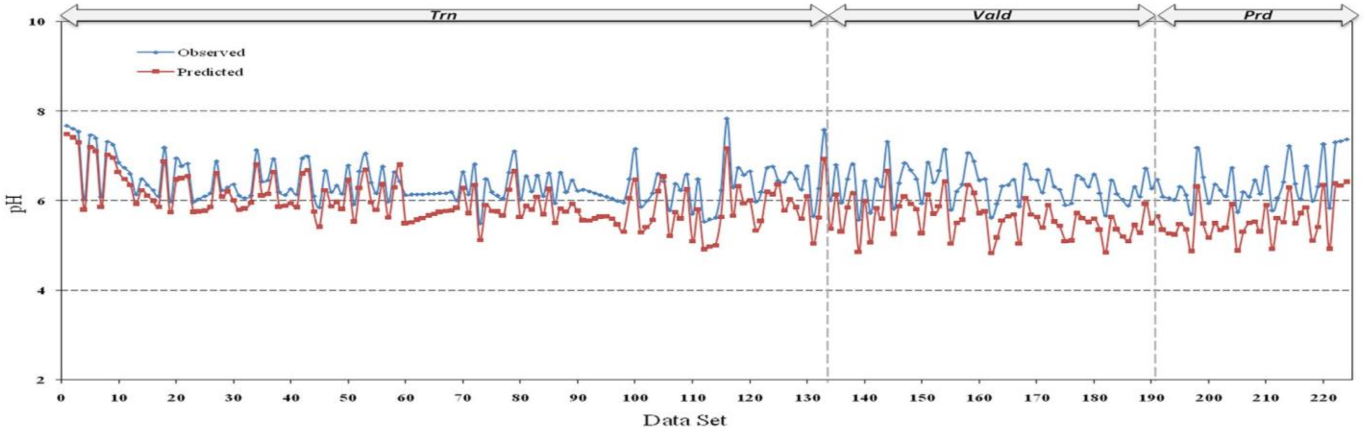
- Scenario I : Different twelve input parameters were used that have been acquired at the same station.

- Scenario II : Developed as, in addition to the same twelve water quality parameters used as inputs in scenario I, the value of AN parameter that has been acquired from the upstream station will be added.

$$WQIP_{N+1} = f_{WDT-ANFIS}(TEMP_N + COND_N + SAL_N + TUR_N + NO_{3N} + CI_N + PO_{4N} + Fe_N + K_N + Mg_N + Na_N + E\_coli_N)$$ N = 1, 2, 3, 4

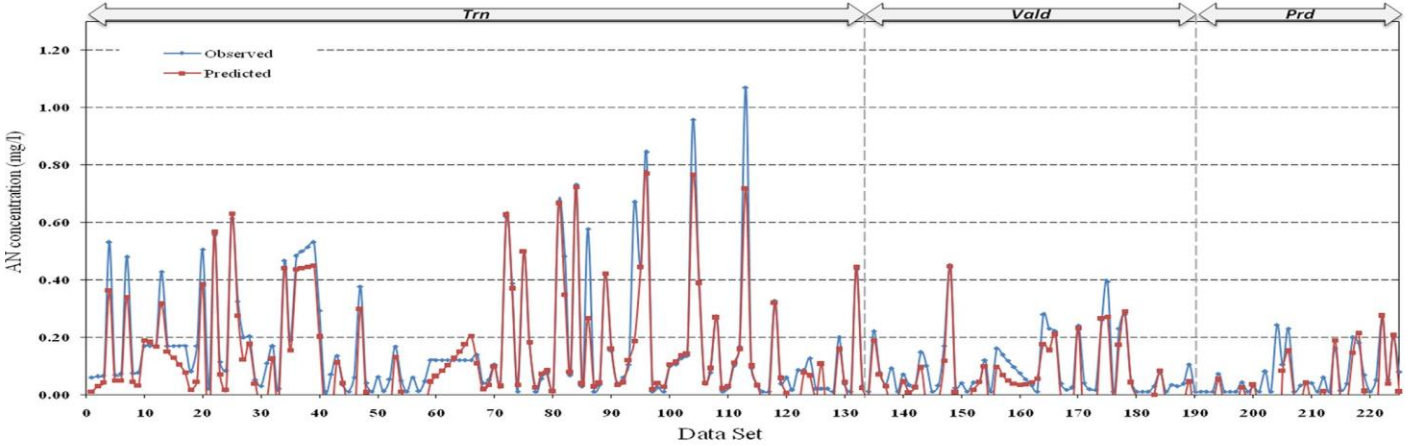
$WQIP_{N+1}$: water quality index parameters pertaining to station N
$f_{WDT-ANFIS}$: non-linear function predictor built via the WDT-ANFIS network
N : station

Water quality prediction model of MLPNN

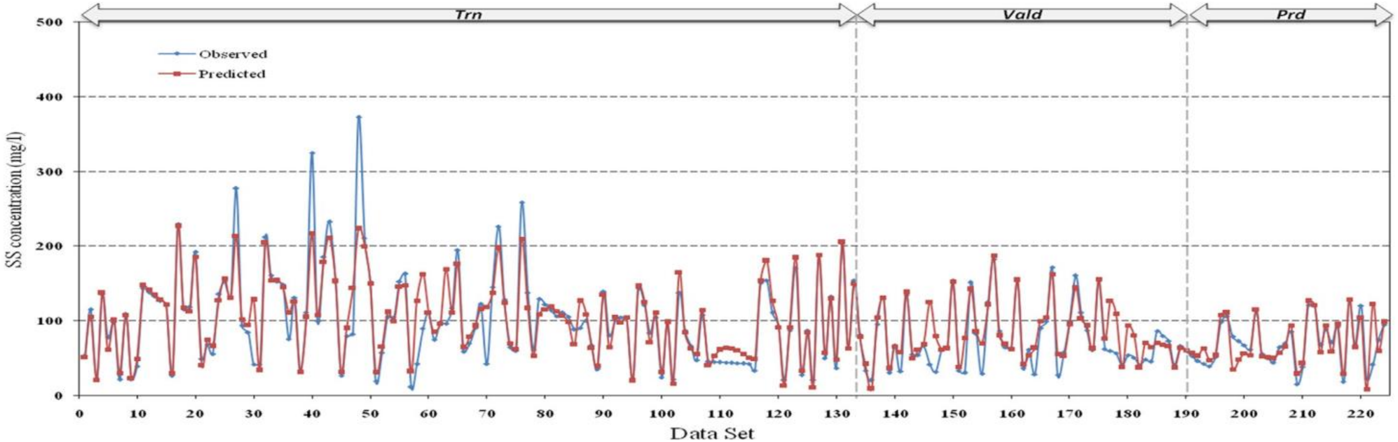


pH

The model's predictive performance was relatively good during the training phase, but its accuracy was lower in the validation and testing phases.



Suspend solid

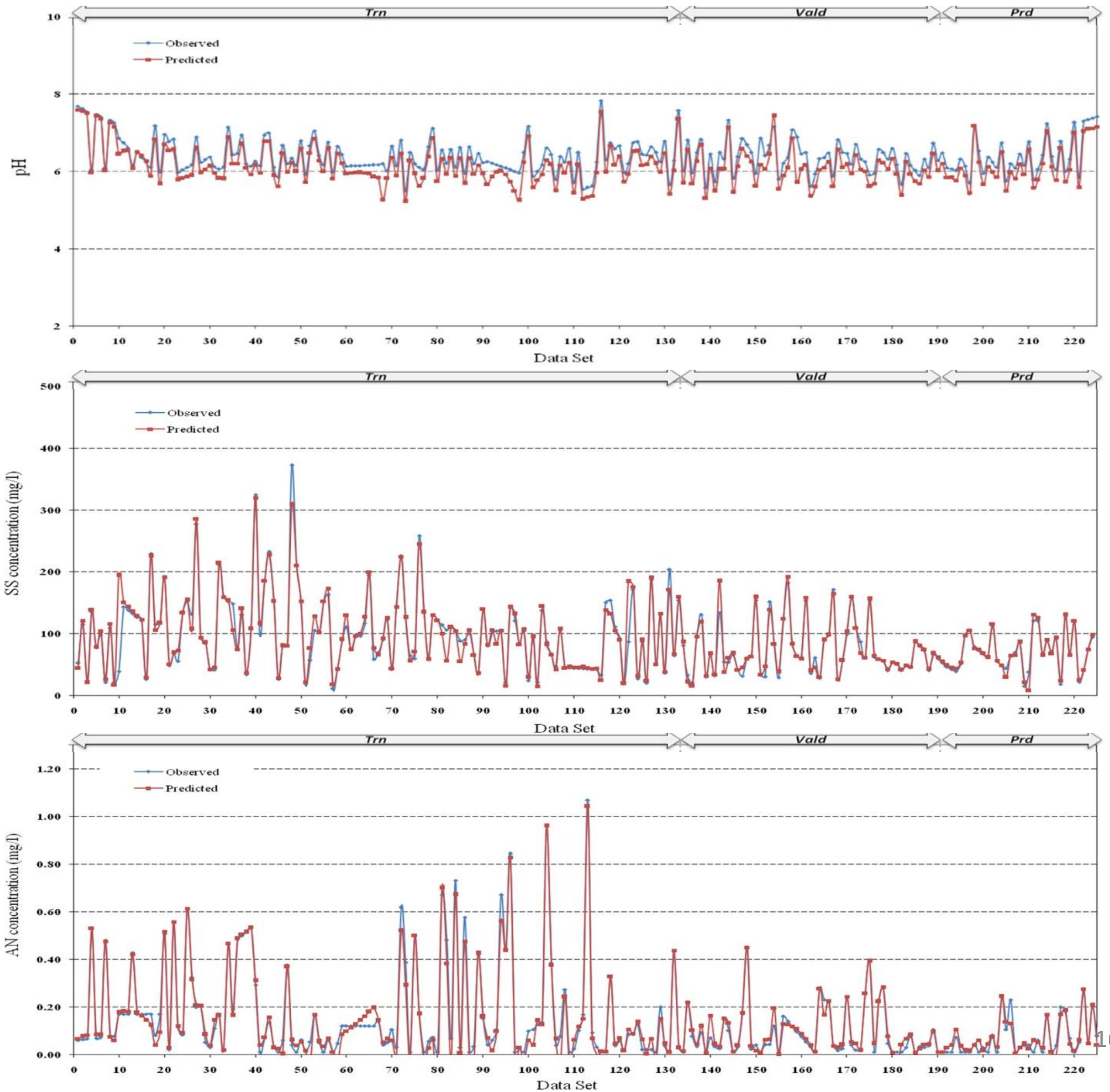


Ammonia nitrogen

- Performance of the MLPNN model: A comparison between the predicted and observed values.

Water quality prediction model of ANFIS

The Results for all three parameters were better compared to MLPNN, making this model the preferred choice for denoising.



pH

Suspend solid

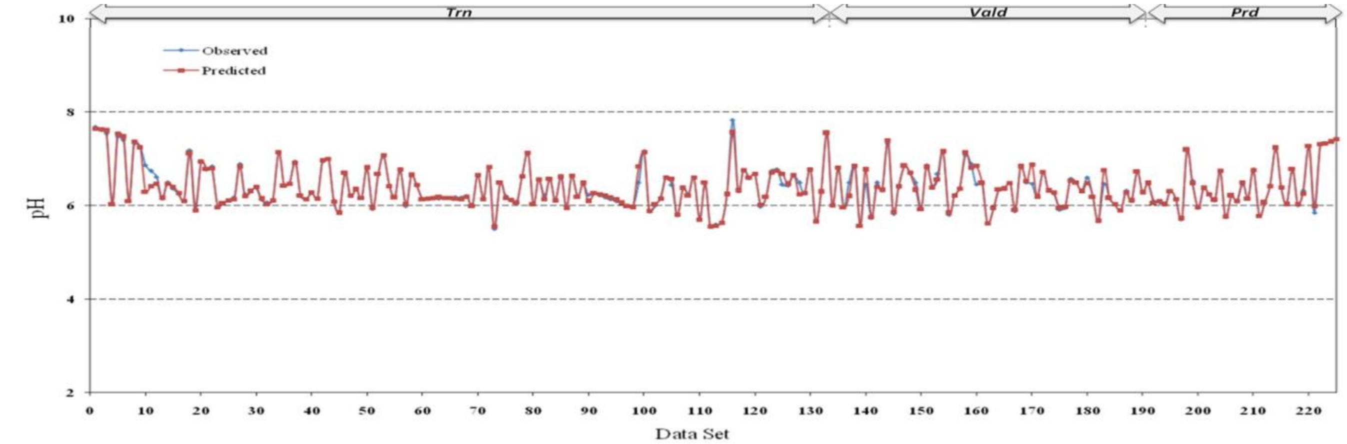
Ammonia nitrogen

- Performance of the ANFIS model: A comparison between the predicted and observed values.

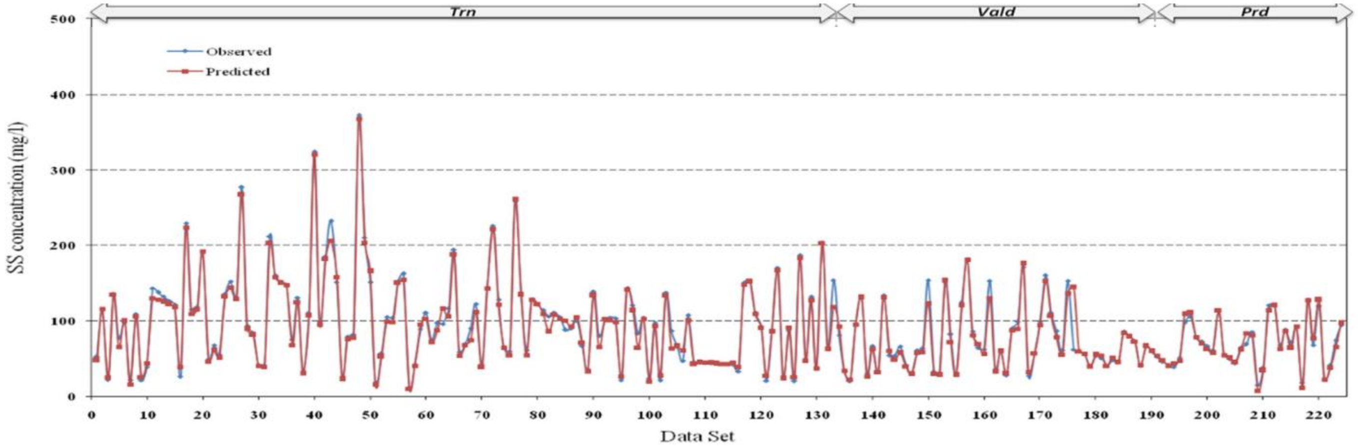
Water quality prediction model of WDT-ANFIS

Due to objective or subjective errors, the raw data must be denoised, and the Results show higher accuracy after the process.

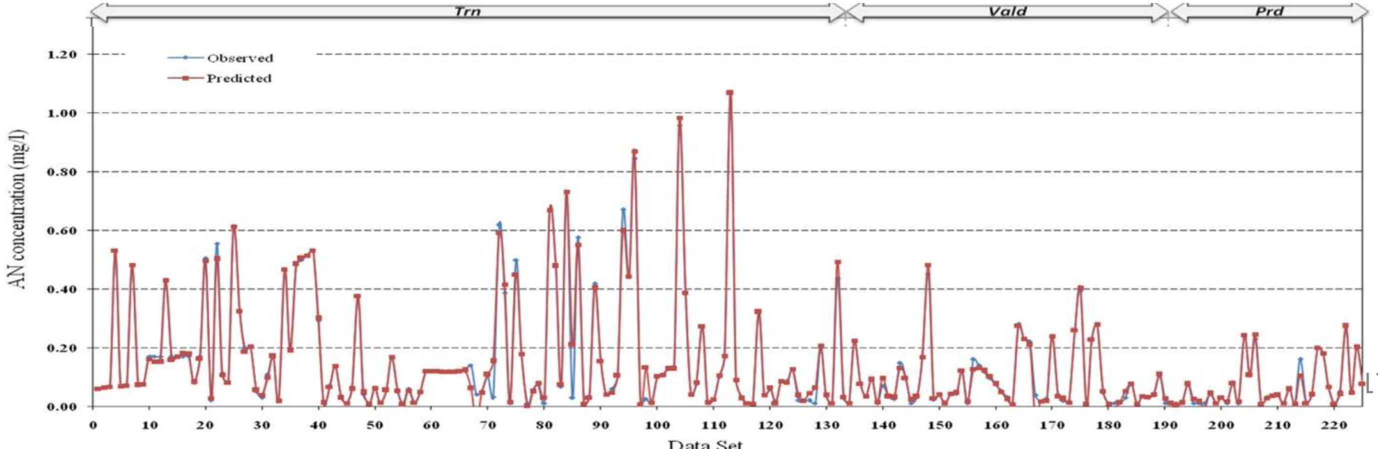
- Performance of the WDT-ANFIS model: A comparison between the predicted and observed values



pH



Suspend solid



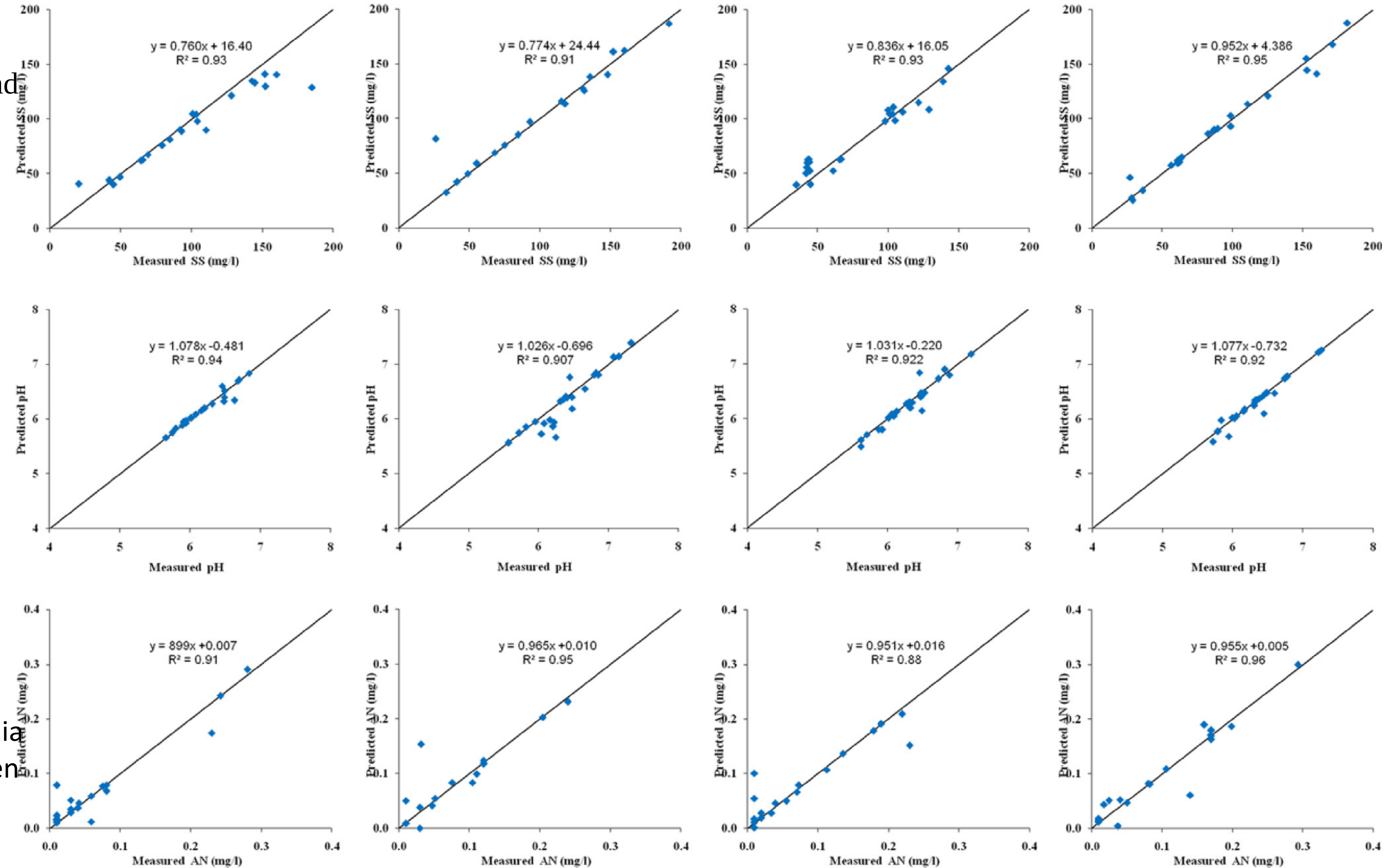
Ammonia nitrogen

Model validation

Suspend solid

pH

Ammonia nitrogen



- WDT-ANFIS model verification for each water quality parameter at each station

- This study proposes the use of an enhanced Wavelet De-noising technique with the Neuro-Fuzzy Inference System (WDT-ANFIS) to predict water quality parameters using historical data. The effectiveness of these models is examined to forecast key parameters that may be impacted by urbanization around the river.

- The WDT-ANFIS approach outperforms the standard ANFIS, improving prediction accuracy for each water quality parameter, making it the optimal network architecture.

Thanks for your listening