

SATELLITE-DRIVEN ASSESSMENT OF TOTAL ORGANIC CARBON AND DISSOLVED SOLIDS USING ARTIFICIAL INTELLIGENCE: CODE DEVELOPMENT AND IMPLEMENTATION

Karim M. El Zahar¹, Hafez A. Afify², Essam Sharaf El Din³

¹Demonstrator, Public Works Department, Faculty of Engineering, Tanta University, Gharbia, Egypt.

²Professor of Surveying and Photogrammetry, Public Works Engineering Department, Faculty of Engineering, Tanta University, Gharbia, Egypt.

³Associate Professor, Public Works Engineering Department, Faculty of Engineering, Tanta University, Gharbia, Egypt.

ABSTRACT

Water is a fundamental resource for life and the progression of human civilization. The escalating global population and its growing demand for clean water are posing significant challenges to the existing water resources. Regular monitoring of water bodies and large-scale estimation of Surface Water Quality Parameters (SWQPs) is crucial; conversely, remote sensing provides extensive spatial and temporal coverage. This study integrates satellite data, Artificial Neural Networks (ANNs), and ground truth water quality data for modelling SWQPs. Initially, atmospheric correction algorithms, including Fast Line of Sight Atmospheric Analysis of Hypercubes (FLAASH), Quick Atmospheric Correction (QUAC), Dark Object Subtraction (DOS), and Atmospheric Correction (ATCOR), have been employed to produce surface

reflectance values of the water region. Then, a python code was developed to determine the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) values between each atmospherically corrected algorithm and the Landsat8 reference data. Secondly, this study compiles the most precise atmospherically corrected algorithm FLAASH with ANN to construct a Landsat8-based Backpropagation Neural Network (BPNN) model for modeling Total Organic Carbon (TOC) and Total Dissolved Solids (TDS). Thirdly, the developed BPNN models were trained with varying epochs (800 to 10,000) to assess their impact on the training process. Adjusting the epochs to 1200 guarantees that the models remain cost-effective and efficient while maintaining accuracy, thereby minimizing processing time and computational costs for future initiatives.

Keywords: SWQPs, ANNs, FLAASH, Landsat8, BPNN

Cite this Article: Karim M. El Zahar, Hafez A. Afify, Essam Sharaf El Din. (2025). Satellite-Driven Assessment of Total Organic Carbon and Dissolved Solids using Artificial Intelligence: Code Development and Implementation. *International Journal of Civil Engineering and Technology (IJCET)*, 16(1), 31-58.

https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_16_ISSUE_1/IJCET_16_01_003.pdf

1. Introduction

The growing population's demand for clean water threatens current supplies [1]. Traditional in-situ water quality assessment techniques, while accurate for specific locations, are time-consuming, costly, and labor-intensive, failing to provide a comprehensive overview of entire water bodies. This makes timely reporting and prediction of water quality challenging over large areas. However, remote sensing has emerged as an accessible and efficient tool for water quality monitoring. Although its accuracy may not match conventional methods, remote sensing provides time and cost efficiency over large areas and enables regular monitoring of remote regions. Consequently, it is increasingly used for monitoring and estimating surface water pollution [2].

Pre-processing of raw remote sensing data is essential before interpretation or analysis to correct errors and enhance data processing capabilities. Although data suppliers may perform some steps, users often need to execute additional pre-processing on the available data [3]. Atmospheric correction methods, including FLAASH, QUAC, DOS, and ATCOR, have been

implemented to minimize atmospheric interference and align corrected image reflectance with actual ground conditions

Due to the complex nature of water quality, which often lacks simple linear relationships with satellite spectral data, mapping SWQPs using ANNs based on remotely sensed imagery is effective. ANNs model accurate nonlinear relationships, providing a reliable means to estimate SWQPs [4], whereas conventional regression techniques often prove insufficient. Consequently, contemporary studies focus on learning-based techniques for more precise SWQP retrieval. The fundamental ANN, the Multi-Layer Perceptron (MLP), a type of Feedforward Neural Network (FNN), predicts relationships between SWQP concentrations and atmospherically corrected satellite data. In FNNs, information flows unidirectionally from input nodes to output nodes through hidden nodes, and they are frequently used in classification and pattern recognition tasks. MLP training employs backpropagation, which adjusts network weights by efficiently computing the gradient of the loss function with respect to the weights for individual input-output samples [5].

The primary objectives of this study are to: (1), conduct a comprehensive statistical analysis comparing Level 2 Landsat8 data with atmospherically corrected images derived through various techniques, including FLAASH, QUAC, DOS, and ATCOR. (2), develop a novel MATLAB code crafted specifically within the MATLAB environment. This advanced code compiles the most accurately atmospherically corrected image data alongside in-situ measurements to construct Landsat8-based BPNNs for estimating TDS and TOC over the watersheds within the study area. (3), train the BPNN models using different numbers of epochs for each SWQP. For each configuration, the processing time required to complete the training process was precisely recorded, aiming to identify the optimal number of epochs that can balance model accuracy and processing time. And (4), compile the most efficient Landsat8-based BPNN for each SWQP, incorporating a water mask to present a spatial distribution map. This map will illustrate the spatiotemporal variation of each SWQP across each pixel of the study area

2. Materials and methods

Fig. 1 Presents a flowchart detailing the extraction of SWQP concentrations using satellite imagery and a Landsat8-based BPNN. This section examines the New Brunswick study area and outlines pre-processing steps, including radiometric calibration, geometric correction, atmospheric corrections to enhance water quality accuracy, in-situ water measurements, and model performance evaluation.

2.1. Selected study area

Fig. 2 illustrates the water bodies incorporated in this study, displayed within the study area boundaries, and overlaid on the mosaic generated from Landsat8 OLI satellite imagery. New Brunswick, which is located on the Atlantic coast, borders the states of Maine in the United States to the west, the Bay of Fundy to the southeast, the Gulf of Saint Lawrence to the northeast, and Quebec to the north. The province's abundant diversity of water bodies is maintained by precipitation that falls throughout the year. Notably, the main discharge basins are the Bay of Fundy to the south and the Gulf of Saint Lawrence to the east and north. Within New Brunswick, numerous lakes dot the landscape, with Grand Lake standing out as the largest freshwater lake. Additionally, the province features several waterfalls, including the impressive Grand Falls, which holds the distinction of being the largest waterfall in the region [6]. The study focuses on assessing the SWQPs in the watersheds located within the province's borders.

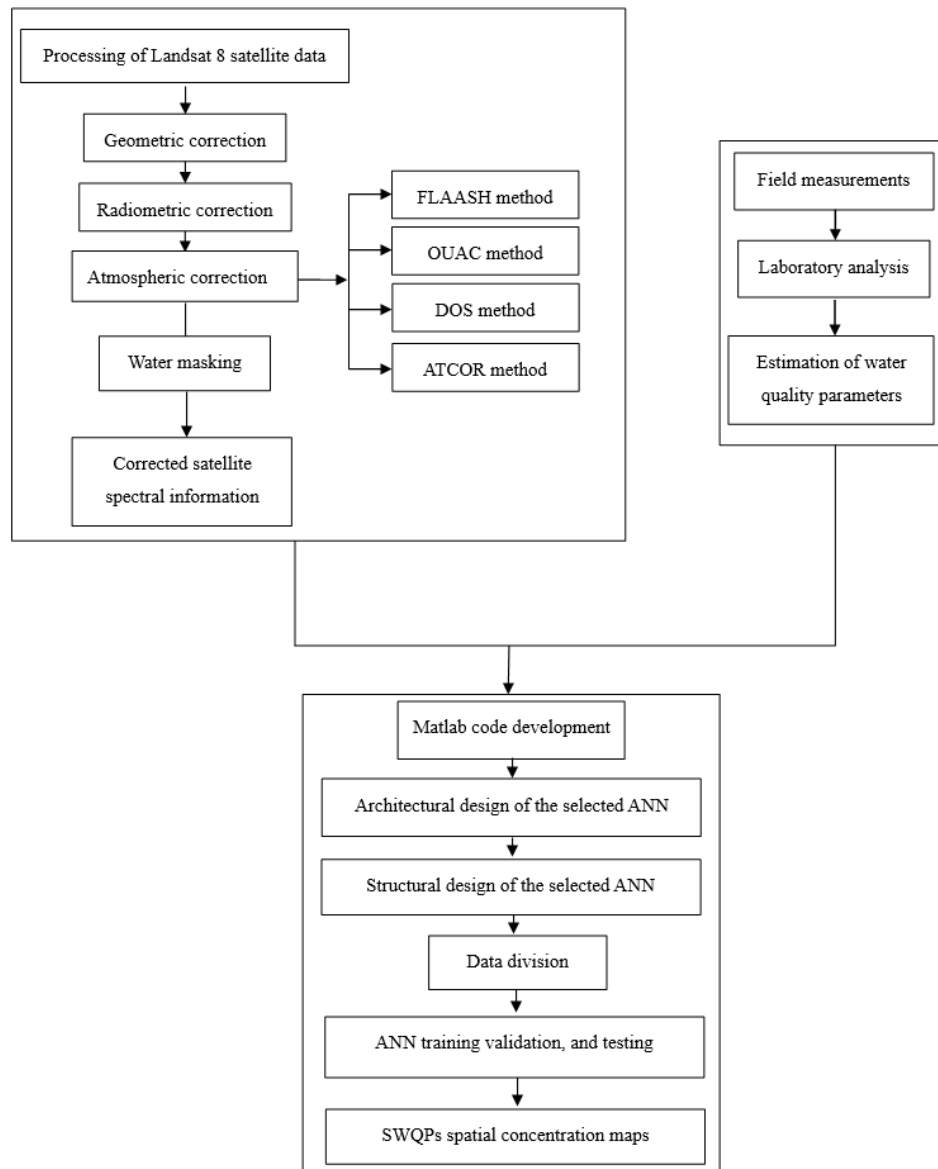


Figure 1: The proposed methodology flowchart.

2.2. Satellite image data

In this study, two different datasets of images have been employed. Each dataset consists of seven Landsat8 OLI satellite images, which are freely available on two levels: Level 1T and Level 2. The Level 1T product, which has been used in this study, is an image that has been geometrically corrected and rectified to the World Geodetic System 1984 (WGS 84) datum and the Universal Transverse Mercator (UTM) projection. It is a terrain-corrected image that has been corrected for sensor orientation and other factors that can affect the quality of the image. The Level 2 product includes the atmospherically corrected surface reflectance with a thirty

meters spatial resolution. In this research work, the Level 2 product served as the study's reference data in order to identify the most accurate atmospheric correction method.

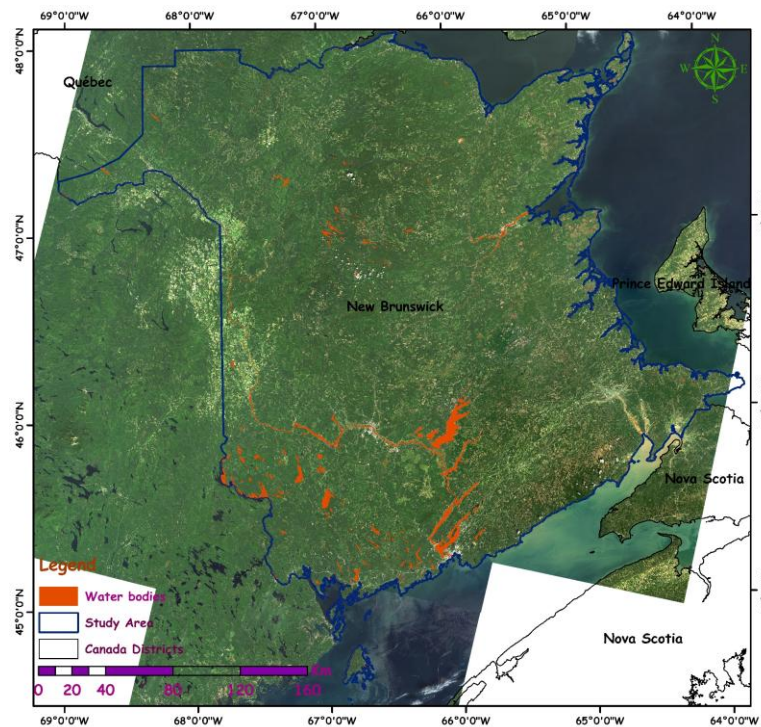


Figure 2: The selected study area along with waterbodies highlighted in red color.

The images used in the study have been acquired at different dates and different zones, as depicted in Table 1. The dates for the utilized images were selected with utmost care to ensure that the cloud cover was minimal. The dates of the image acquisition matched the available dates of the water sampling process to ensure that the measurements were as accurate as possible.

3. Processing of landsat8 satellite data

Raw image data requires pre-processing before interpretation or analysis. Although data providers may perform some steps, users must often complete additional pre-processing. This section discusses the applied operations, including geometric correction, radiometric calibration, and atmospheric correction [3], [7].

Table 1: The World Reference System (WRS), UTM zone, and the dates of the selected datasets for the study area.

WRS Path- Row	UTM zone	Date	
		The 1 st dataset	The 2 nd dataset
Path 9 – Row 27	20	13-9-2017	29-10-2022
Path 9 – Row 28	20	21-5-2016	11-9-2022
Path 10 – Row 27	19	25-8-2019	4-10-2022
Path 10 – Row 28	19	25-8-2019	2-9-2022
Path 10 – Row 29	19	8-7-2019	1-6-2023
Path 11 – Row 27	19	14-9-2018	23-5-2023
Path 11 – Row 28	19	14-9-2018	23-5-2023

3.1. Geometric correction

Fourteen Landsat8 OLI processing Level 1T images were used. They have been geometrically corrected and rectified to the WGS 84 datum and UTM projection.

3.2. Radiometric correction

Radiometric calibration transforms Digital Numbers (DNs) into physical units, such as radiance (W/m²/sr/μm) or Top-Of-Atmosphere (TOA) reflectance, ensuring data accuracy and consistency. This process corrects systematic errors and temporal variations, enabling reliable comparisons of images captured at different times. By converting DNs to physical units, it represents the actual electromagnetic radiation at the sensor's ground-equivalent location during image capture. Calibration techniques vary depending on the sensor and analysis requirements

The general equations used in radiometric calibration are as follows:

- **Conversion to Radiance:** The DNs in an image are first converted to radiance using the following equation (1):

$$L = G * DN + B \quad (1)$$

Where L: the spectral radiance (W/m²/sr/μm), G: the gain or radiance scaling factor, DN: the digital number of the pixel, and B: the bias or radiance offset.

- **Conversion to Reflectance:** The radiance is then converted to reflectance using the following equation (2):

$$\rho = \frac{\pi * L * d^2}{E_{sun} * \cos(\theta)} \quad (2)$$

Where ρ is the unitless solar reflectance, L is the spectral radiance ($\text{W/m}^2/\text{sr}/\mu\text{m}$), d is the Earth-Sun distance in astronomical units, E_{sun} is solar atmospheric irradiance, and θ is the solar zenith angle in degrees.

3.3. Atmospheric correction

Several atmospheric correction techniques were applied to Level 1T images to minimize atmospheric effects. The following section examines the employed algorithms in detail, highlighting their mechanisms for mitigating atmospheric scattering and absorption to enhance the accuracy and reliability of image analysis. Four different atmospheric correction algorithms, such as FLAASH, QUAC, DOS, and ATCOR, have been implemented to ensure that spectral values accurately represent ground conditions. The discussion covers their theoretical foundations, mathematical formulations, and practical implementation, along with a comparative analysis of their strengths, limitations, and suitability for different imagery datasets.

3.3.1. Fast Line of Sight Atmospheric Analysis of Hypercubes (FLAASH)

FLAASH, developed by Spectral Sciences, Inc. (SSI) and the Air Force Research Laboratory (AFRL/VS), is an atmospheric correction algorithm designed for hyperspectral and multispectral imaging from visible to shortwave infrared wavelengths up to $3 \mu\text{m}$ [8]. Its main objectives include accurate extraction of atmospheric parameters (e.g., aerosol, cloud cover, surface pressure, and water vapor) and converting radiance-at-detector data into surface reflectance values using a correction matrix [9].

A key feature of FLAASH is its user-friendly graphical interface for spectral calculations and its data simulation capabilities available in the paid ENVI version. The program estimates aerosol content and provides derived data, such as column water vapor, surface altitude, and cloud-mask images. It generates atmospherically corrected images, specifically surface spectral reflectance, using available Look-Up Tables (LUTs) and includes a "polishing" technique for image sharpening and adjacency correction in Multi-Spectral Imaging (MSI) data [10].

With little heat emission, FLAASH is intended to process radiance spectral bands from mid-infrared to Ultraviolet. Equations (3), and (4) can be used in this context to parameterize the spectral radiance (L) at a sensor pixel, where all variables are wavelength-dependent [9], [11].

$$L = \left(\frac{A \rho}{1 - \rho e S} \right) + \left(\frac{B \rho e}{1 - \rho e S} \right) + L_a \quad (3)$$

$$L_e = \left(\frac{(A+B) \rho e}{1 - \rho e S} \right) + L_a \quad (4)$$

Where L : the light received by the sensor for a single pixel, ρ : the reflectance of the surface of the pixel, S : the process of accounting for the total reflection and scattering of sunlight by particles in the atmosphere, ρ_e : the average reflectance of the surface for the pixel and its surroundings, L_a : the amount of light that is scattered back towards the sensor by particles in the atmosphere, L_e : the average spatial radiance image, and A and B : coefficients that are surface-independent and vary with atmospheric and geometric conditions.

The term $(\frac{A\rho}{1-\rho_e S})$ represents the surface-reflected light intensity (from scattered skylight and direct sunlight) reaching the sensor, while $(\frac{B\rho_e}{1-\rho_e S})$ accounts for atmospheric-reflected light directed back to the sensor [8]. Accurate water vapor estimation relies on the ratio of "absorption" radiances (within absorption bands) to "reference" radiances (at band edges) linked to the water vapor column. A 3D Look-Up Table (LUT) models water vapor across absorption and reference bands. Initially, adjacency effects are ignored, assuming uniform ρ_e values. Coefficients $(A+B, S, L_a)$ generate absorption and reference radiances (L) for various reflectance values, forming a 2D LUT with water vapor as the dependent variable [8].

3.3.2. Quick Atmospheric Correction (QUAC)

QUAC is a distinctive in-scene atmospheric correction technique developed for multispectral and hyperspectral imagery. Its distinctiveness lies in its capacity to ascertain atmospheric correction parameters straight from scene data using observed pixel spectra. The approach takes its cues from the empirical finding that the average spectrum of various material spectra, such as endmember spectra in a picture, tends to differ dramatically between scenes. This feature allows QUAC to retrieve reflectance spectra with reasonable accuracy, even in situations where the sunlight intensity is unknown, or the sensor is not properly calibrated. Understanding these principles is crucial in fields like remote sensing and atmospheric science. Notably, QUAC boasts computational speed, making it a potential choice for real-time applications [12].

Unlike physically based approaches, QUAC requires minimal metadata, relying only on an approximate characterization of sensor band positions (center wavelengths) and their radiometric calibration. While QUAC may sacrifice some precision compared to physically based methods, its computational efficiency makes it an attractive option [13]. In Fig. 3, the basic mechanics of QUAC are depicted. The total of the three indicated lines is the observed spectral radiance, ρ_{obs} , for a pixel with surface reflectance, ρ_{true} . Solar photons can follow

three different paths, A, B, and C, to reach the observer sensor. ρ_{true} is the observed surface pixel's reflectance, ρ_{ave} is the average reflectance of the pixels nearby, and ρ_{obs} is the sensor radiance that corresponds to the observed surface pixel. Equations (5), and (6) clarify these relationships [14].

$$\rho_{obs} = (A + C\rho_{ave}) + B\rho_{true} \quad (5)$$

$$\rho_{true} = \text{Gain} (L_{obs} - \text{offset}) \quad (6)$$

Where ρ_{true} : the reflectance of the observed surface pixel, ρ_{ave} : the average reflectance of the neighboring pixels, ρ_{obs} : the sensor radiance of the observed surface pixel, Offset: the minimum pixel value for each band, Gain: $1/B$, and A, B, and C: parameters that are retrieved from the in-scene spectral data.

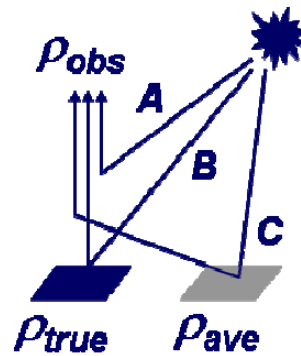


Figure 3: The contributions of radiation transfer to the sensed apparent reflectance [15].

The elements in $(A + C\rho_{ave})$ are nearly constant across an image, acting as a shared offset for all pixels. Surface reflectance can be expressed using the detected signal and atmospheric parameters, where the $\text{Gain} = 1/B$ and the $\text{Offset} = (A + C\rho_{ave})$ [15].

In a physically based atmospheric correction, parameters A, B, and C are derived by comparing observed spectral features to those predicted by Radiative Transfer (RT) calculations. RT models simulate light interaction with the atmosphere and surface, factoring in atmospheric composition, aerosols, and surface properties.

The goal is to determine A, B, and C to align observed spectral data with RT-simulated features, ensuring the model accurately reflects physical processes affecting radiance. These parameters are then applied to correct radiance values, yielding precise surface reflectance

estimates. While rigorous, this approach requires detailed knowledge of atmospheric conditions and sensor characteristics [14].

3.3.3. Dark Object Subtraction (DOS)

When processing remotely sensed data, the DOS method can be applied by subtracting a constant value, referred to as the DN, from each pixel in the image. This method assumes that haze, caused by atmospheric light scattering, is uniform across the image. While this assumption may not always hold due to atmospheric variability, a first-order correction via DOS is often preferable to no correction. Each spectral band requires unique constants specific to the image [16].

The DOS approach assumes that some pixels in the image represent completely black areas with 0% reflectance. However, due to atmospheric scattering, shadowed pixels may have non-zero DN values. To correct for first-order scattering, the DN value of each spectral band must be subtracted. This value can be identified using various techniques, including histogram analysis [17].

In histogram analysis, the DN frequency histogram of the image is examined to identify the ideal DN value. For visible spectrum bands, haze causes histogram shifts toward higher DN values. These histograms often display a sharp increase in pixel counts at non-zero DN levels, representing haze in the band. To ensure accuracy, the histogram should be constructed from the entire image or a significant portion. Using histograms from localized areas may lead to overcorrection and failure to detect true minima or dark objects [13].

3.3.4. Atmospheric Correction (ATCOR)

ATCOR, integrated into PCI Geomatica, offers workflows for atmospheric correction, including (1) Top-of-the-Atmosphere Reflectance, converting pixel values to reflectance above the atmosphere by normalizing radiance values with minimal data; (2) Haze Removal and Cloud Masking, generating masks for clouds and water, removing haze, and preparing images for correction or mosaicking, producing adjusted images with raw DN values and pre-classification masks; and (3) Ground Reflectance Atmospheric Correction, calculating ground-level reflectance by eliminating atmospheric effects, accounting for topography, water vapor, and aerosols for accurate analysis under diverse atmospheric conditions [18].

The information provided is subject to updates in PCI Geomatica software. Atmospheric impacts can be mitigated by computing ground-level reflectance values using the PCI Geomatics ATCOR Ground Reflectance procedure, which automatically utilizes metadata from the input scene.

ATCOR employs MODerate resolution atmospheric TRANsmission-5 (MODTRAN-5), while FLAASH uses MODTRAN-4, both widely used for simulating electromagnetic radiation propagation in the 0.2–100 μm spectral range. These Radiative Transfer Models (RTMs) enhance understanding of atmospheric radiative processes critical to remote sensing. Comparative studies of RTMs, including 6SV v2.1, libRadtran v2.0.2, and MODTRAN (v5 and v6) evaluate their performance and identify model-specific variations. While explicit equations are not detailed, these methods involve complex computations related to sensor design, sensitivity analysis, energy balance models, atmospheric correction, aerosol and gas retrieval, and scene modeling [19], [20].

Atmospheric correction algorithms, including FLAASH, QUAC, and ATCOR, are applied individually to each image before mosaicking. This is because variations in sun azimuth and elevation arise from differences in satellite and sun viewing angles during scene acquisition. These angles are influenced by time-dependent factors such as the Landsat8 satellite orbit, Earth's rotation, and its orbit around the sun. Such variations impact sun and sensor viewing angles, necessitating correction on a per-image basis to ensure accuracy.

3.5. Water Masking

A water mask, a binary image, distinguishes water bodies from land in satellite imagery, with water pixels labeled as '1' (white) and non-water pixels as '0' (black). It isolates water bodies to prevent their impact on reflectance values. The mask is created by applying a threshold based on water's spectral properties, forming a binary image that excludes non-water pixels. This process is essential for accurately mapping SWQP concentrations, improving data quality, ensuring reliability, and streamlining model processing by focusing solely on water pixels.

4. Field measurements

In this study, two sets of water sampling points, comprising ninety-six samples, were selected to maximize spatio-temporal variation across the study area and capture the widest range of SWQP concentrations. Fig.4 and 5 visually depict the sampling sites overlaid on the two generated water masks. Sampling was conducted in alignment with image acquisition, and TOC and TDS were measured at each station. These analyses adhered to the laboratory standards for water and wastewater established by the American Public Health Association [21]. TOC measurements were taken at Eighty-five locations, with some sites lacking TOC data.

5. Mapping SWQPs concentrations using the BBNN algorithm

5.1. Multi-Layer Perceptron (MLP)

An example of FNN is MLP: one input layer, one or more hidden layers, and one final output layer make up an MLP. Neurons in the input layer are matched to the input data's feature set. The computational nodes in the hidden layers use a nonlinear activation function, which helps the network learn from the error during the training phase. The output layer receives connections from the last hidden layer. The computation of a neuron in MLP is the same as the computation of a neuron in a simple ANN. MLPs are trained using backpropagation. Backpropagation in MLPs computes the error at the output and propagates it back across the network layers [22].

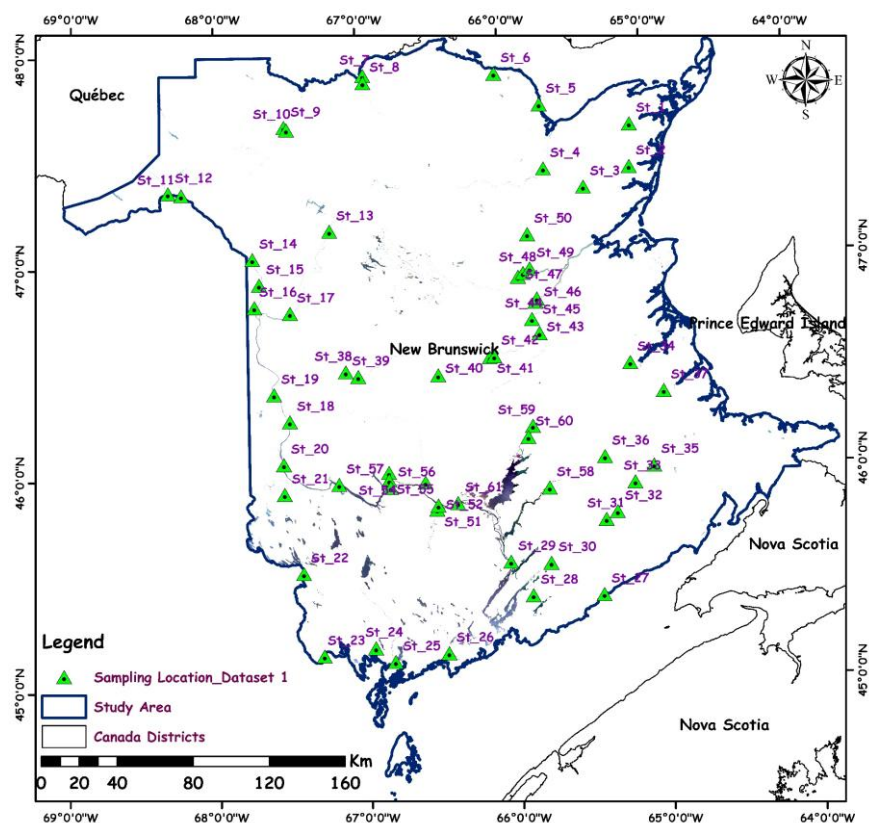


Figure 4: The 1st dataset of water sampling locations.

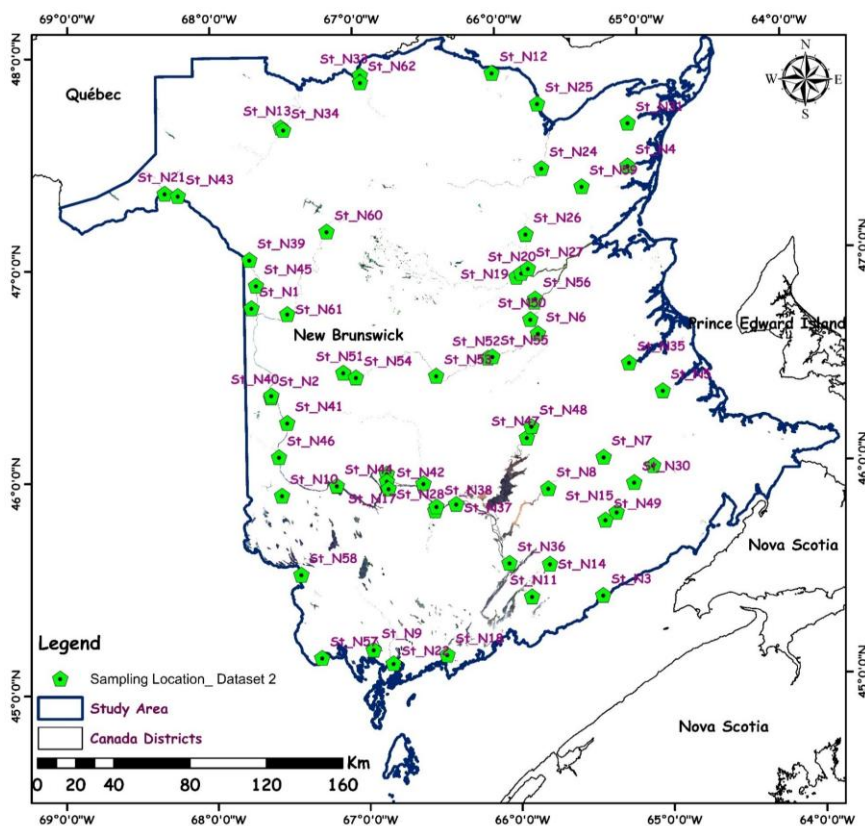


Figure 5: The 2nd dataset of water sampling locations.

One of the most popular learning algorithms in digital image processing is back propagation. It forwards input values through ANN layers, with hidden and output nodes weighted by connections [23]. The error signal, calculated as the difference between desired and actual outputs, as shown in equation (7), is minimized using gradient descent, often paired with backpropagation, to find global error minimal. The learning rate is applied in gradient descent at the hidden layer, where errors propagate from the output to the input layer. The network is trained to determine the optimal weight combination that, in theory, yields precise outputs that match the inputs.

$$L = \frac{1}{2} (\text{Desired output} - \text{Predicted output})^2 \quad (7)$$

5.2. Data division

The data was divided into calibration (training and validation) and testing sets. The training set trained the model by adjusting weights and biases, while the validation set tuned the learning rate, monitored overfitting, enabled early stopping, and selected the best model. It

ensured the model generalized rather than memorized the training data. The testing set provided a final, unbiased evaluation of the model's performance on unseen data [24].

5.3. characteristics of the selected ANN used in this research study

The BPNN code was developed in MATLAB R2021a, carefully considering the required architectural and structural design. The study utilized a four-layer FNN-MLP architecture consisting of an input layer, two hidden layers, and an output layer (Fig. 6). The input layer included seven neurons: Coastal Blue, Blue, Green, Red, NIR, SWIR 1, and SWIR 2, while the output layer handled one SWQP at a time to simplify the ANN and accelerate computations.

The network employed a linear aggregation function with varying activation functions across layers. The first hidden layer and output layer used the hyperbolic tangent (tanh) function, while the second hidden layer employed the sigmoid (logistic) function. Input neurons transmit weighted inputs to hidden and output layers; this basic operation is crucial for the execution of the network. The combination of sigmoid and tanh activation functions enhanced the model's ability to capture complex input-output relationships. This methodology effectively established strong correlations between SWQPs and satellite reflectance data.

The structure of a neural network, along with its architecture, delineates the functional form of the relationship between inputs and outputs. Optimizing the network structure involves selecting the number of hidden neurons and transferring functions to balance generalization, speed, and complexity. In this study, a trial-and-error approach was adopted, starting with two hidden neurons and incrementally increasing to thirty. Networks with fewer than fifteen neurons underfit the data, failing to learn patterns, while those with more than twenty-five neurons overfit and exhibited slower learning. It was observed that employing fewer than fifteen neurons led to an underfitting issue, which essentially means the network was unable to learn the desired patterns from the data. Conversely, utilizing more than twenty-five neurons resulted in slower learning and overfitting problems, where the model becomes overly complex and performs poorly on unseen data. Thus, a balance was sought to optimize the learning process and model performance.

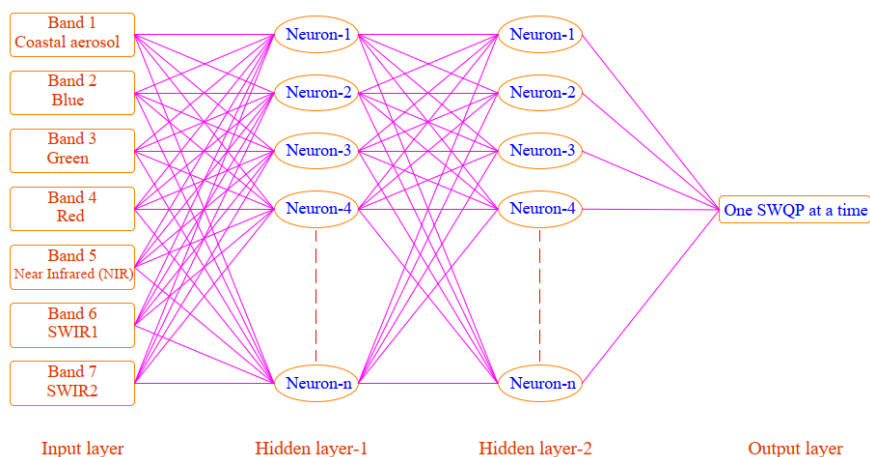


Figure 6: The architectural design of the nominated ANN.

For TOC, the first and second hidden layers used twenty and fifteen neurons, respectively, while TDS employed fifteen neurons for both layers. Cross-validation was employed to prevent overfitting during training, and a learning rate of 0.01 was chosen to ensure convergence to the global minimum. Learning rates below 0.001 reduced computational speed, while values above 0.10 impaired performance and generalization. This fine-tuned calibration of hidden neurons and learning rate was essential to optimizing the ANN's performance.

To evaluate the impact of the training process on model performance, processing time, and computational efficiency, trials were conducted with varying epochs, representing the number of full passes through the training dataset. Models were trained with 800, 1000, 1200, 1500, 2000, 3000, 4000, 5000, and 10,000 epochs for each SWQP, and the processing time for each configuration was recorded. This analysis aimed to determine the optimal number of epochs to balance accuracy and processing efficiency.

6. Results and discussion

6.1. SWQPs measurements

The concentrations of SWQPs have been examined to remove outliers. Based on a 95% confidence level, a Z critical value of 1.96 was chosen. Any data with Z values greater than 1.96 or less than -1.96 are considered outliers and removed [25]. Table 2 shows comprehensive descriptive statistics of the selected SWQPs. Concentrations ranged from 1.10 to 11.0 mg/L with a mean value of 5.316 mg/L for TOC and from 11.0 to 144.0 mg/L with a mean value of 56.3198 mg/L for TDS.

Table 2: Descriptive statistics of SWQPs.

SWQPs	TOC (mg/l)	TDS (mg/l)
No. of samples	85	96
Minimum	1.1	11.0
Maximum	11	144
Range	9.9	133
Mean (μ)	5.316	56.3198
Standard deviation	2.5977	30.5796

The dataset for the selected SWQPs was divided into 70% for training, 15% for testing, and 15% for validation, ensuring robust model evaluation. Table 3 provides detailed information on this division and the number of stations for each parameter.

6.2. Comparing different implemented atmospheric correction algorithms

A comparative analysis of mosaic images from different atmospheric correction methods was conducted using reference data. Two statistical measures, R^2 and RMSE, were computed pixel by pixel across all seven mosaic bands. Python 3.11, with its extensive libraries and ease of use, was employed for these calculations, alongside Microsoft Visual Studio 2015 for code editing, debugging, and version control. A code has been built to calculate the R^2 and RMSE values. The analysis allowed for a quantitative assessment of each correction method's performance, identifying the method producing images closest to the reference and the associated errors.

Table 3: Data for ANN training, validation, and testing.

Developed models	TOC BPNN model	TDS BPNN model
Total No. of samples	85	96
No. of training samples	59	68
No. of validation samples	13	14
No. of testing samples	13	14

Table 4 displays the R^2 and RMSE values for images corrected using different atmospheric correction methods compared to reference data. This tabular format simplifies the assessment of each method's performance, providing clear insights into the precision and reliability of various techniques when applied to satellite imagery.

Table 4: R^2 and RMSE for the different atmospherically corrected images.

Image bands in micrometer	FLAASH method		QUAC method		DOS method		ATCOR method	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Coastal aerosol	0.7731	0.01082	0.3878	0.01731	0.6258	0.03871	0.4732	0.03128
Blue	0.8283	0.00945	0.6909	0.01289	0.06556	0.02391	0.4624	0.01168
Green	0.8343	0.01092	0.8211	0.01051	0.5855	0.01232	0.1224	0.02913
Red	0.8334	0.00753	0.6944	0.01252	0.2884	0.01214	0.1161	0.01106
NIR	0.8582	0.04882	0.5285	0.07972	0.1441	0.09642	0.4829	0.12217
SWIR1	0.8112	0.01287	0.6343	0.02691	0.2337	0.06165	0.6242	0.01093
SWIR 2	0.6541	0.02070	0.5283	0.05087	0.2804	0.03334	0.6911	0.03704

The results reveal FLAASH as the most accurate method for atmospheric correction, with mean values of (0.7989 and 0.0173) for R^2 and RMSE, respectively. ATCOR follows with (0.6122, 0.0301), DOS yields (0.3175, 0.0398), and QUAC produces (0.4246, 0.0362). These findings align with prior research [26], highlighting FLAASH's superior accuracy. Despite the demonstrated accuracy of the FLAASH and ATCOR methods, it is noteworthy that the DOS and QUAC methods offer a more straightforward implementation and do not necessitate the input of intricate parameters. FLAASH and ATCOR generally have higher accuracy, but poor utility compared to QUAC and DOS methods since acquiring real-time parameters in the model is complex in practical operation.

6.3. The insights of BPNN models

The MATLAB code was developed to train the ANN based on the specified parameters outlined earlier, optimizing the network for the intended application. Fig. 7 shows a portion of the code for the TOC parameter, demonstrating the training process structure. To balance model accuracy and processing time, it was found that more epochs improved R^2 values and reliability but also increased processing time and cost. The BBNN models were trained under consistent conditions, including the same learning rate and number of stations for training, validation, and testing. The development was carried out on a machine with an Intel Core i7-7820HQ CPU at 2.90 GHz, 8 cores, and 32 GB of RAM.

The number of epochs was varied to assess its impact on the training process. Models were trained with different epoch counts for each SWQP, and processing time, R^2 , and RMSE values were recorded for each configuration. Results were presented in charts (Fig. 8 and 9) to

illustrate the relationship between processing time and R^2 values for each SWQP, simplifying trend identification and conclusions about the optimal number of epochs.

The results showed that as the number of epochs increased, R^2 improved, and RMSE decreased. However, this performance improvement came with higher processing time and cost. The literature review suggested that ANN models for SWQPs are reliable when R^2 values are around 0.90. In this study, models exceeded an R^2 of 0.90 at 1200 epochs, reaching 0.9134 during testing. The average processing time for all SWQPs was approximately 90 minutes. At 1500 epochs, R^2 increased to 0.9268, with a processing time of 110 minutes. The R^2 difference between the two models was 0.0152, while processing time differed by about 20 minutes.

```
1 % Load your input and target data
2 inputs = IN; % Your input data matrix
3 targets = OUT; % Your target data matrix
4
5 % Create a feedforward neural network
6 hiddenLayerSizes = [20, 15];
7 net = patternnet(hiddenLayerSizes);
8
9 % Set activation functions for each layer
10 net.layers{1}.transferFcn = 'tansig'; % tansigmoid for the first hidden layer
11 net.layers{2}.transferFcn = 'logsig'; % logsigmoid for the second hidden layer
12 net.layers{end}.transferFcn = 'tansig'; % tansigmoid for the output layer
13 % Randomly divide the data
14 net.divideFcn = 'dividerand';
15 net.divideParam.trainRatio = 70/100; % Ratio for training set
16 net.divideParam.valRatio = 15/100; % Ratio for validation set
17 net.divideParam.testRatio = 15/100; % Ratio for testing set
18
19 % Use Bayesian Regularization backpropagation
20 net.trainFcn = 'trainlm';
```

Figure 7: Segment of the developed MATLAB code for training the TOC ANN model.

The results showed an R^2 value of 0.9384 for 2000 epochs, with an average processing time of 2 hours and 5 minutes. This improved to 0.9459 at 3000 epochs, with a processing time of 3 hours and 20 minutes. The R^2 difference was minimal (0.0075), but processing time increased significantly by 1 hour and 26 minutes, nearly doubling between the two models, requiring greater labor and more advanced machine capabilities.

For models trained with 4000, 5000, and 10000 epochs, R^2 values were 0.9496, 0.9504, and 0.9559, respectively, with processing times of 4 hours 30 minutes, 5 hours 7 minutes, and 8 hours 19 minutes. As epochs increased, processing time rose significantly, correlating with higher computational costs. Despite the increased resources, the R^2 improvements were minimal. The extended processing time requires more labor and advanced machine capabilities,

resulting in higher operational costs, both in terms of human labor and the quality of machinery needed to support the intensive computational tasks.

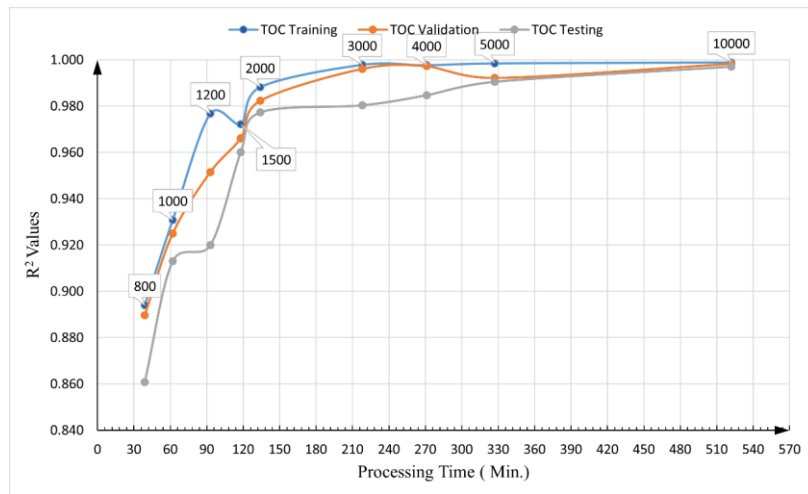


Figure 8: Graphical results for the developed TOC models showing processing time and R^2 for training, validation, and testing datasets at each epoch.

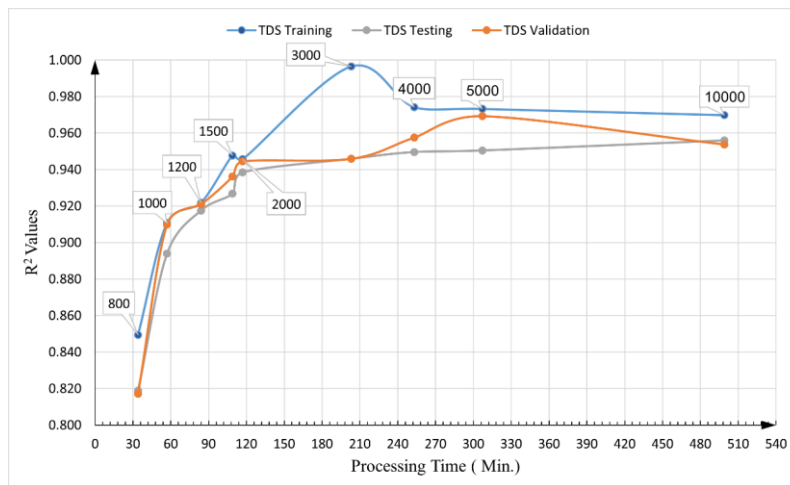


Figure 9: Graphical results for the developed TDS models showing processing time and R^2 for training, validation, and testing datasets at each epoch.

The results indicate that training models beyond 1200 epochs significantly increase processing time and computational costs without substantial gains in R^2 values. This conclusion stems from evaluating the resources required, including labor intensity and machine capabilities. Limiting epochs to 1200 ensures cost-effective, efficient models without sacrificing accuracy, minimizing processing time and overall costs. Models trained with 1200

epochs will be used for generating spatial maps for each SWQP. For the BPNN algorithm with 1200 epochs, the RMSE values for TOC were 0.4417, 0.5490, and 0.5735 mg/L for training, validation, and testing, respectively. For TDS, the RMSEs were 8.8243, 6.7279, and 8.8495 mg/L. The Landsat8-based BPNN showed significant efficiency in estimating SWQP concentrations, with R^2 values for TOC at 0.9768, 0.9515, and 0.9199 for training, validation, and testing, respectively. For TDS, the R^2 values were 0.9219, 0.9210, and 0.9176. These results are summarized in Table 5 and Fig. 10.

Table 5: Statistical measures between the actual and modeled concentrations of SWQPs using the developed Landsat8-based-BPNN at 1200 epoch.

1200 Epochs ANN models result							
SWQPs	Training		Validation		Testing		Approximate processing time
	R^2	RMSE	R^2	RMSE	R^2	RMSE	
TOC	0.9768	0.4417 mg/L	0.9515	0.5490 mg/L	0.9199	0.5735 mg/L	1 Hr 33 Min
TDS	0.9219	8.8243 mg/L	0.9210	6.7279 mg/L	0.9176	8.8495 mg/L	1 Hr 24 Min

6.4. The Landsat8-based-BPNN spatial concentration maps

After successfully executing the Landsat8-based BPNN model for each SWQP, a MATLAB code was used to apply these models to the final Landsat8 OLI image. The image, exported as a TIFF file from PCI Geomatica, was processed pixel by pixel. Each pixel was input into the developed Landsat8-based BPNN model within the MATLAB environment. The goal was to generate spatial concentration maps for each SWQP, including TOC and TDS, which are visually presented in Fig. 11 to Fig. 14 for both image datasets.

Spatial distribution maps reveal that TOC concentrations vary between 1.0 and 11.0 mg/L. Lower TOC concentrations were identified in the Canaan River, Digdeguash River, Magaguadavic River, and St. Croix River, likely due to lower concentrations of organic matter input and reduced human activities in these areas. Conversely, higher TOC concentrations were found in the Hammond River, Kennebecasis River, and across significant portions of Grand Lake, which may be attributed to higher levels of organic matter runoff from surrounding agricultural and urban areas.

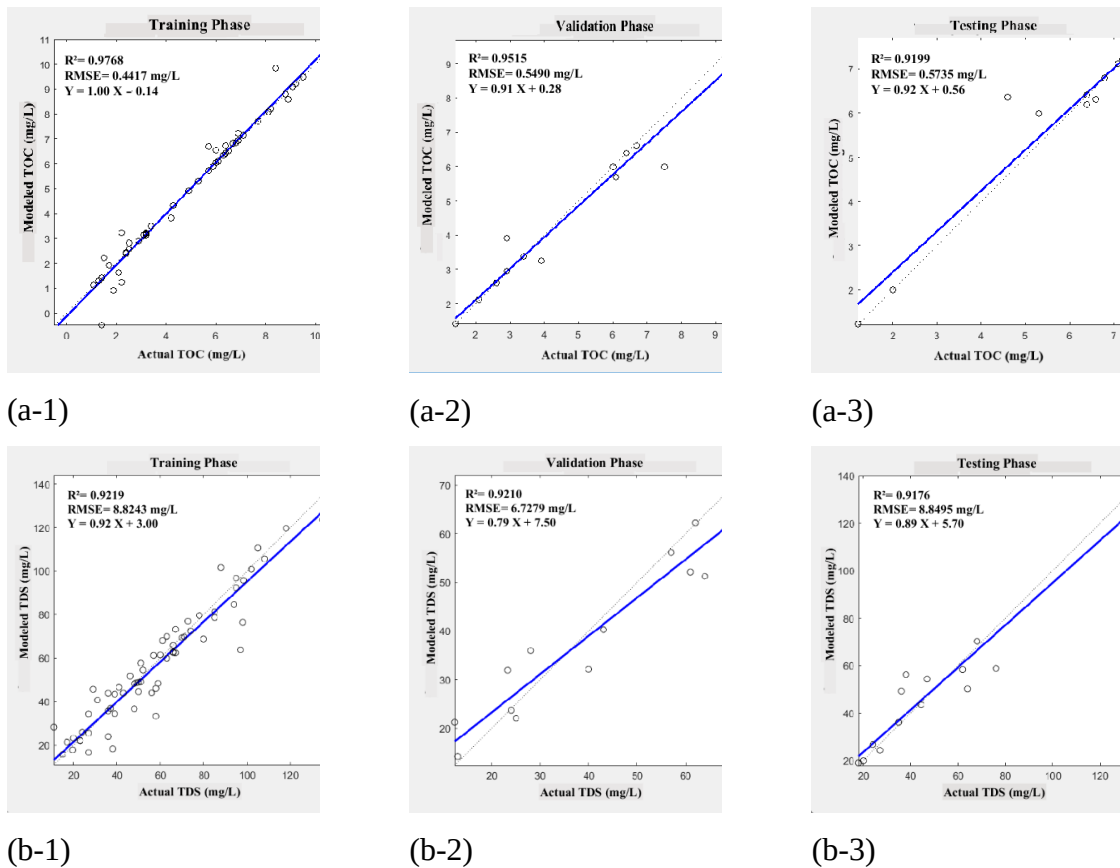


Figure 10: Graphical fit results of TOC (a-1), (a-2), and (a-3)), and TDS ((b-1), (b-2), and (b-3)), for training, validation, and testing of the developed Landsat8-based-BPNN at 1200 epochs.

The monitoring results indicate variability in TDS concentrations across different watersheds, ranging between 10.0 and 144.0 mg/L. Lower TDS concentrations were identified in the Saint John River and across significant portions of Grand Lake, likely due to limited industrial activities, lower levels of agricultural runoff, and better management practices that reduce the inflow of dissolved solids. Conversely, higher TDS concentrations were found in the Canaan River, Hammond River, Digdeguash River, and Magaguadavic River, which may be attributed to higher levels of agricultural runoff, industrial discharges, and increased urbanization contributing to greater amounts of dissolved solids entering these water bodies

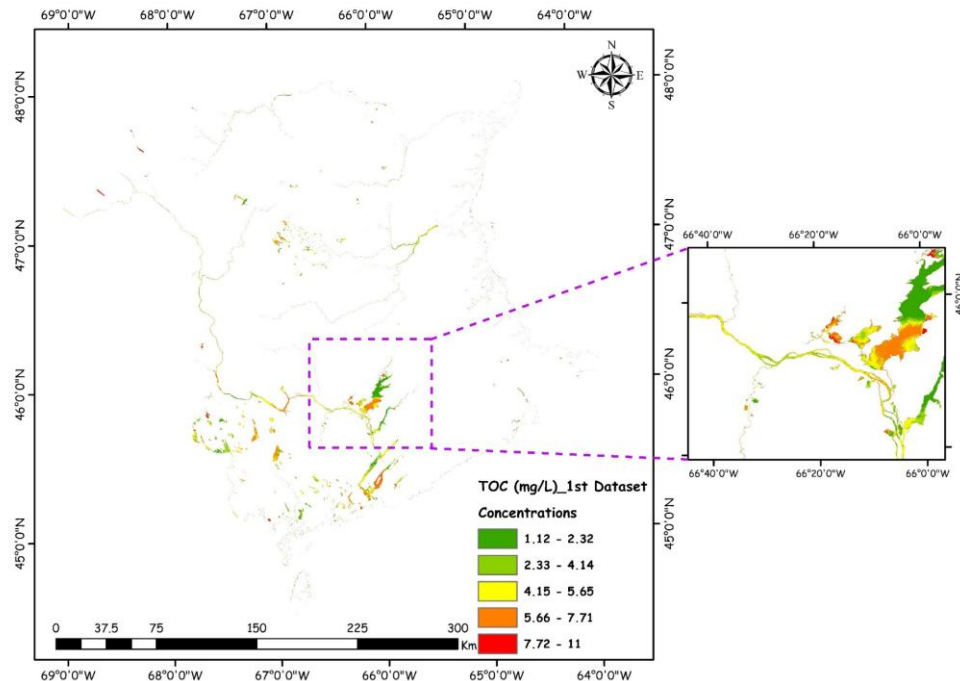


Figure 11: A spatial distribution map for TOC using the developed Landsat8-based-BPNN for the 1st dataset of images.

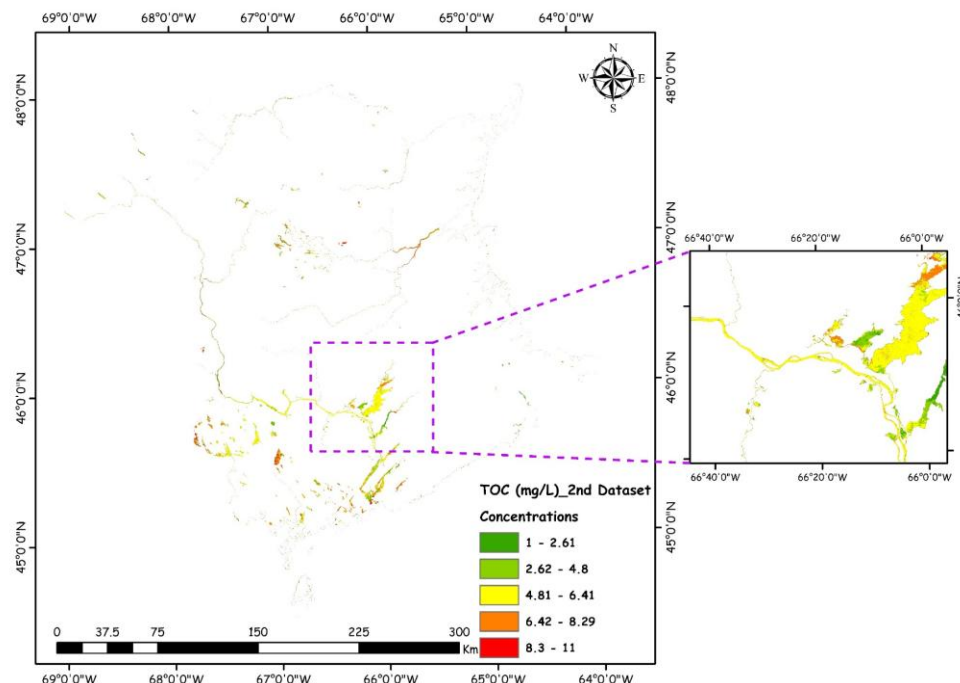


Figure 12: A spatial distribution map for TOC using the developed Landsat8-based-BPNN for the 2nd dataset of images.

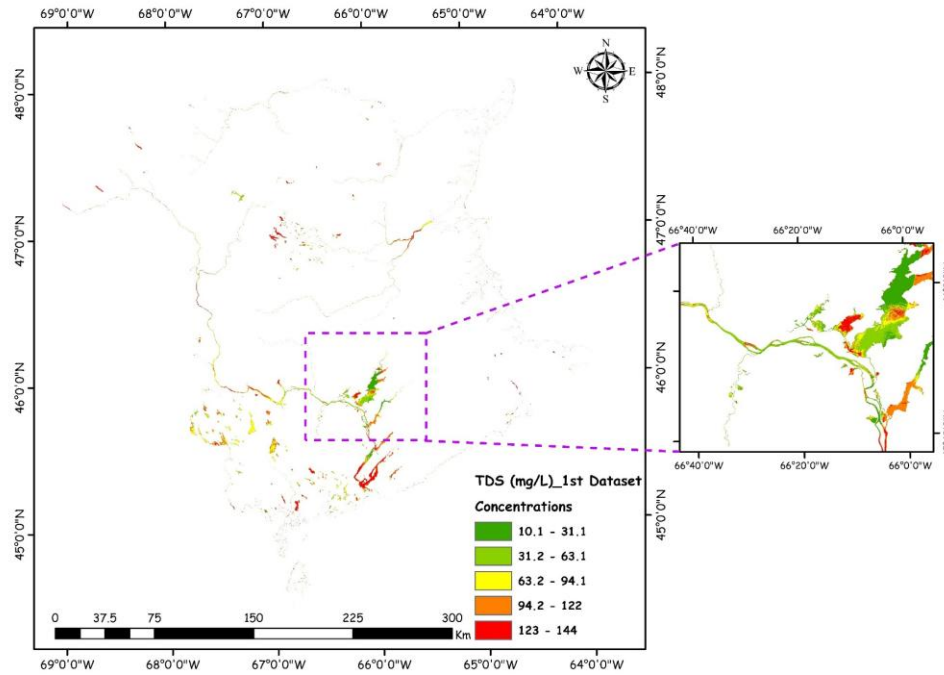


Figure 13: A spatial distribution map for TDS using the developed Landsat8-based-BPNN for the 1st dataset of images.

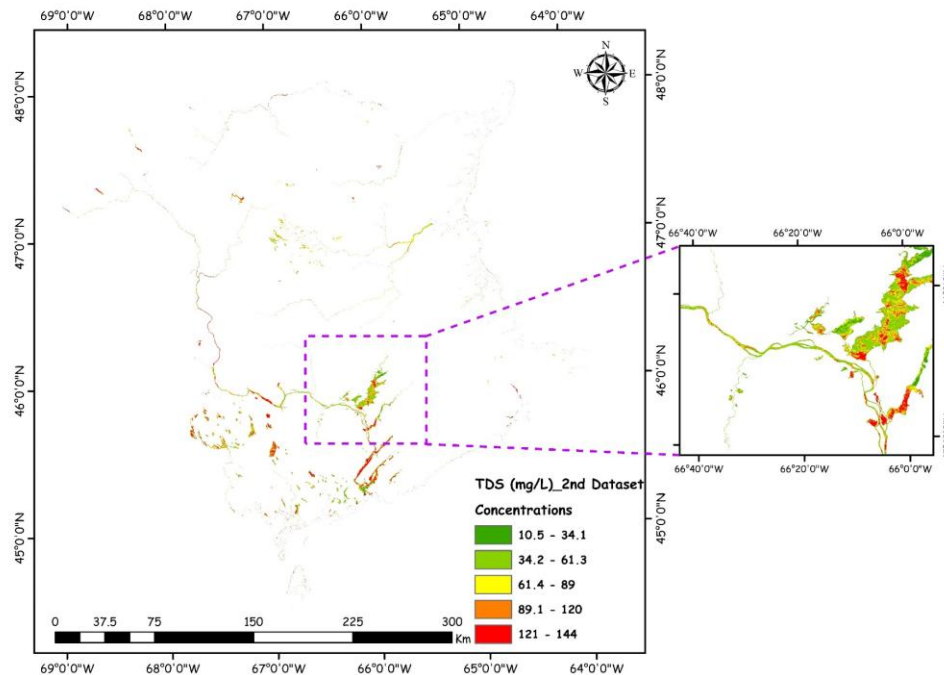


Figure 14: A spatial distribution map for TDS using the developed Landsat8-based-BPNN for the 2nd dataset of images.

7. Conclusion

This study integrated remote sensing imagery with machine learning algorithms to develop a Landsat8-based BPNN for accurately mapping SWQPs, including TOC and TDS, in New Brunswick, Canada. The methodology achieved strong correlations across watersheds in the study area. Identifying the optimal number of epochs revealed that while increased epochs improved R^2 values and reliability, they also resulted in longer processing times and higher computational costs. However, the Landsat8-based BPNN models are not universally applicable to other water bodies without adaptation. Each water surface has distinct environmental, climatic, and geographical characteristics. For instance, proximity to industrial or agricultural drainage systems or differences in rainfall patterns necessitate modifications to the model architecture and design to ensure accuracy and contextual relevance.

Acknowledgment

The authors wish to extend sincere gratitude to the USGS Landsat Archive Center for providing the Landsat8 Level 1T imagery that was crucial for our study.

References

- [1] M. H. Gholizadeh, A. M. Melesse, and L. Reddi, "A comprehensive review on water quality parameters estimation using remote sensing techniques," *Sensors (Switzerland)*, vol. 16, no. 8, 2016, doi: 10.3390/s16081298.
- [2] B. K. Purandara, B. S. Jamadar, T. Chandramohan, M. K. Jose, and B. Venkatesh, "Water Quality Assessment of a Lentic Water Body Using Remote Sensing: A Case Study," pp. 371–380, 2018, doi: 10.1007/978-981-10-5792-2_30.
- [3] T. Lillesand, R. Kiefer, and J. Chipman, "Remote sensing and image interpretation," *Aust J Geod Photogramm Surv*, vol. 39, pp. 73–77, 1994, Accessed: Jun. 24, 2023. [Online]. Available: <https://www.wiley.com/en-it/Remote+Sensing+and+Image+Interpretation%2C+7th+Edition-p-9781118343289>
- [4] E. Sharaf El Din and Y. Zhang, "Neural Network Modelling of the Saint John River Sediments and Dissolved Oxygen Content from Landsat OLI Imagery," 2017.

- [5] Y. H. Zweiri, J. F. Whidborne, and L. D. Seneviratne, "A three-term backpropagation algorithm," *Neurocomputing*, vol. 50, pp. 305–318, 2003, doi: 10.1016/S0925-2312(02)00569-6.
- [6] "New Brunswick Map & Satellite Image | Roads, Lakes, Rivers, Cities." Accessed: Jun. 05, 2024. [Online]. Available: <https://geology.com/canada/new-brunswick.shtml>
- [7] R. P. Gupta, "Remote Sensing Geology Third Edition," 2018.
- [8] T. Perkins et al., "Retrieval of atmospheric properties from hyper and multispectral imagery with the FLAASH atmospheric correction algorithm," in *Remote Sensing of Clouds and the Atmosphere X*, SPIE, Oct. 2005, p. 59790E. doi: 10.1117/12.626526.
- [9] G. W. Felde et al., "Analysis of Hyperion Data with the FLAASH Atmospheric Correction Algorithm," 2003.
- [10] David. Schaefer, 31st Applied Imagery Pattern Recognition Workshop : proceedings : Washington, DC : October 16-18, 2002. IEEE Computer Society, 2002.
- [11] J. Yuan and Z. Niu, "Evaluation of Atmospheric Correction Using FLAASH," 2008.
- [12] L. S. Bernstein, "Quick atmospheric correction code: algorithm description and recent upgrades," *Optical Engineering*, vol. 51, no. 11, p. 111719, Jul. 2012, doi: 10.1117/1.oe.51.11.111719.
- [13] A. S. Mahiny and B. J. Turner, "A Comparison of Four Common Atmospheric Correction Methods," *Photogramm Eng Remote Sensing*, vol. 73, no. 4, pp. 361–368, 2007, doi: 10.14358/PERS.73.4.361.
- [14] L. S. Bernstein, S. M. Adler-Golden, X. Emin Jin, B. Gregor, and R. L. Sundberg, "QUICK ATMOSPHERIC CORRECTION (QUAC) CODE FOR VNffi-SWIR SPECTRAL IMAGERY: ALGORITHM DETAILS," *WHISPERS*, 2012.
- [15] L. S. Bernstein, S. M. Adler-Golden, R. L. Sundberg, R. Y. Levine, T. C. Perkins, and A. Berk, "A New Method for Atmospheric Correction and Aerosol Optical Property Retrieval for VIS-SWIR Multi-and Hyperspectral Imaging Sensors: QUAC (QUick Atmospheric Correction)," 2005.

- [16] L. Yan and J. Li, “A Case Study of Dark-objects Subtraction based Atmospheric Correction Methods for GF-1 Satellite Images,” 2018 Fifth International Workshop on Earth Observation and Remote Sensing Applications (EORSA), pp. 1–4, Dec. 2018, doi: 10.1109/EORSA.2018.8598541.
- [17] P. Wicaksono and M. Hafizt, “Dark target effectiveness for dark-object subtraction atmospheric correction method on mangrove above-ground carbon stock mapping,” IET Image Process, vol. 12, no. 4, pp. 582–587, Apr. 2018, doi: 10.1049/iet-ipr.2017.0295.
- [18] “ATCOR - Ground Reflectance Workflow - CATALYST.” Accessed: Oct. 10, 2024. [Online]. Available: <https://catalyst.earth/tutorial/atcor-ground-reflectance-workflow/>
- [19] C. Raab, B. Barrett, F. Cawkwell, and S. Green, “Evaluation of multi-temporal and multi-sensor atmospheric correction strategies for land-cover accounting and monitoring in Ireland,” Remote Sensing Letters, vol. 6, no. 10, pp. 784–793, Oct. 2015, doi: 10.1080/2150704X.2015.1076950.
- [20] B. Pflug and M. Main-Knorn, “Validation of atmospheric correction algorithm ATCOR,” in Remote Sensing of Clouds and the Atmosphere XIX; and Optics in Atmospheric Propagation and Adaptive Systems XVII, SPIE, Oct. 2014, p. 92420W. doi: 10.1117/12.2067435.
- [21] “Standard Methods for the Examination of Water and Wastewater,” 2005.
- [22] T. Kim and T. Adali, “Approximation by Fully Complex Multilayer Perceptrons,” Neural Comput, vol. 15, no. 7, pp. 1641–1666, Jul. 2003, doi: 10.1162/089976603321891846.
- [23] A. Suliman and Y. Zhang, “A Review on Back-Propagation Neural Networks in the Application of Remote Sensing Image Classification,” Journal of Earth Science and Engineering, vol. 5, pp. 52–65, 2015, doi: 10.17265/2159-581X/2015.
- [24] E. Sharaf El Din, Y. Zhang, and A. Suliman, “Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework,” Int J Remote Sens, vol. 38, no. 4, pp. 1023–1042, Feb. 2017, doi: 10.1080/01431161.2016.1275056.

- [25] P. V. Anusha, C. Anuradha, P. S. R. Chandra Murty, and C. S. Kiran, "Detecting Outliers in High Dimensional Data Sets using Z-Score Methodology," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 48–53, Nov. 2019, doi: 10.35940/IJITEE.A3910.119119.
- [26] M. chakouri, R. Lhissou, A. El Harti, S. Maimouni, and Z. Adiri, "Assessment of the image-based atmospheric correction of multispectral satellite images for geological mapping in arid and semi-arid regions," *Remote Sens Appl*, vol. 20, Nov. 2020, doi: 10.1016/j.rsase.2020.100420.

Citation: Karim M. El Zahar, Hafez A. Afify, Essam Sharaf El Din. (2025). Satellite-Driven Assessment of Total Organic Carbon and Dissolved Solids using Artificial Intelligence: Code Development and Implementation. *International Journal of Civil Engineering and Technology (IJCET)*, 16(1), 31-58.

Abstract Link: https://iaeme.com/Home/article_id/IJCET_16_01_003

Article Link:

https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_16_ISSUE_1/IJCET_16_01_003.pdf

Copyright: © 2025 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Creative Commons license: Creative Commons license: CC BY 4.0



✉ editor@iaeme.com