

RS-WaterQuality Mapper Manual

The RS-WaterQuality Mapper is implemented as a plugin for QGIS 3.x using Python and the QGIS Application Programming Interface (API). The architecture of the plugin follows a modular, sequential workflow that guides users systematically through the necessary steps of a water quality analysis, from image data input, processing, to final map products. It provides an intuitive graphical user interface (GUI) within QGIS (Figure 1). Once installed, it appears as a dockable panel with multiple tabs corresponding to the main steps of analysis (Fig. 1): Generate Water Mask, Atmospheric Correction with ACOLITE, Calculate Spectral Indices from Image, Calculate Trophic State Index, Matching Pixel Values with Water Sample Points, Empirical Modeling, Spectral Clustering of Water Sample Points, Ensemble Component Model Calibration and Selection, Spectral-Space Partition and Ensemble Model Generation, Integrated Ensemble Modeling. The toolbox is designed around a logical, sequential workflow that guides the user through each necessary step of water quality analysis, from raw satellite data to final map product. Figure 2 outlines the workflow for water quality remote sensing model development and mapping.

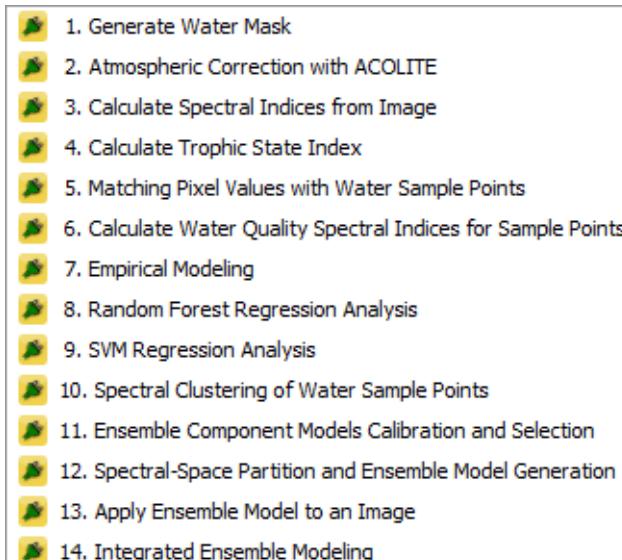
- 
- 1. Generate Water Mask
 - 2. Atmospheric Correction with ACOLITE
 - 3. Calculate Spectral Indices from Image
 - 4. Calculate Trophic State Index
 - 5. Matching Pixel Values with Water Sample Points
 - 6. Calculate Water Quality Spectral Indices for Sample Points
 - 7. Empirical Modeling
 - 8. Random Forest Regression Analysis
 - 9. SVM Regression Analysis
 - 10. Spectral Clustering of Water Sample Points
 - 11. Ensemble Component Models Calibration and Selection
 - 12. Spectral-Space Partition and Ensemble Model Generation
 - 13. Apply Ensemble Model to an Image
 - 14. Integrated Ensemble Modeling

Figure 1. User Interface of core processing modules in the toolbox

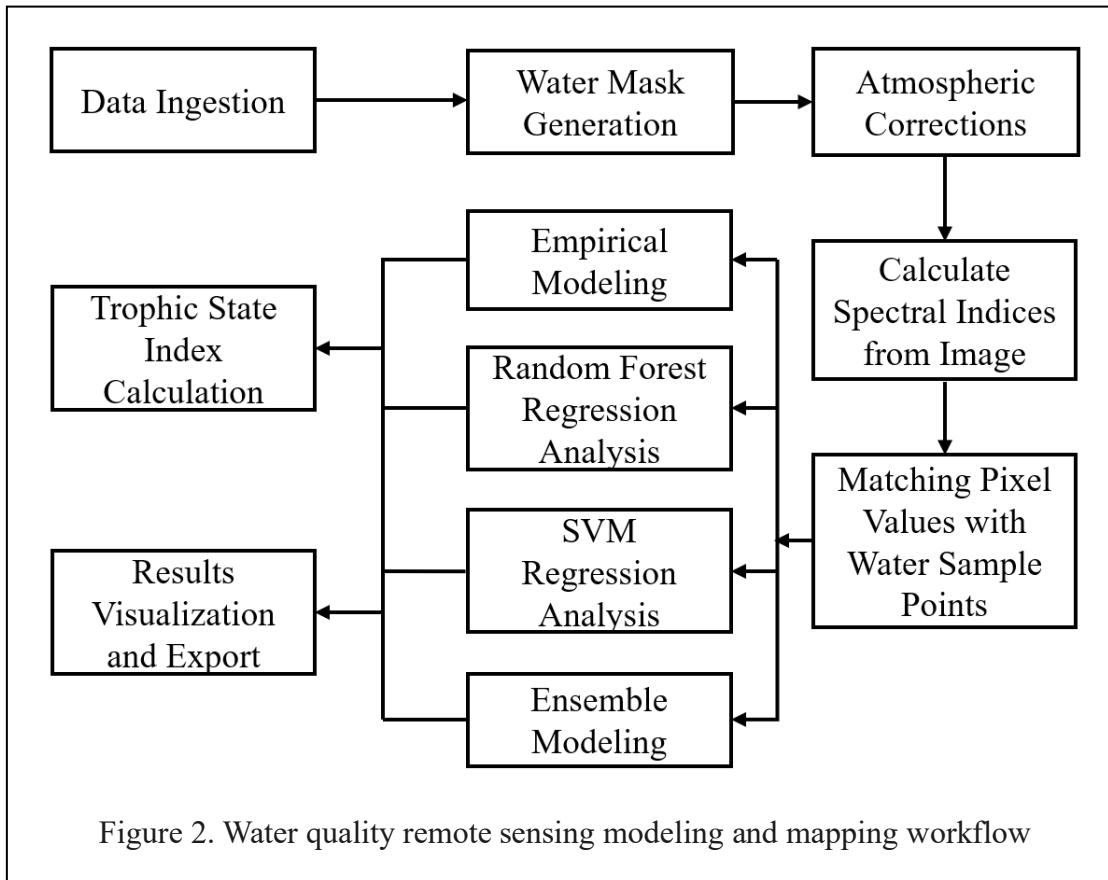


Figure 2. Water quality remote sensing modeling and mapping workflow

1. Install the plugin in QGIS

In QGIS, go to Menu->Plugin, Manage and Install plugins...

Select "Install from zip"

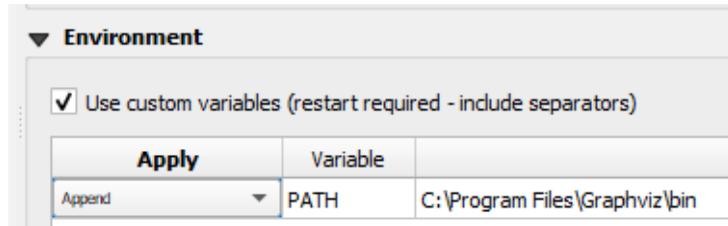
Find the zip file provided as the plugin

Click "Install Plugin"

Click yes to install dependencies for the package. Ignore the error message.

Download and install graphviz from <https://graphviz.org/download/>.

In QIS, go to Settings->Options->System->Environment, check “Use custom variable (restart required – include separators)”, append a new variable PATH, set its value to “C:\Program Files\Graphviz\bin” :



Restart QGIS, go to Menu->Plugins, Manage and Install plugins...., Select “Installed”, check the box before “**RS-WaterQuality Mapper**”.

The Plugin can be applied to the following Satellites: Landsat 5, 7, 8, 9, and Sentinel-2. The Plugin can be used to derive the following water quality parameters from multi-spectral satellite imagery: Chlorophyll, turbidity, CDOM.

2. Input Data preparation

1) The plug-in takes Shapefile as input. CSV table should be imported as Shapefile before it can be used by the plug-in. To add a delimited text file (like a CSV) to QGIS as a point layer and save as Shapefile please follow the steps below:

Open the Data Source Manager: Go to Layer -> Add Layer -> Add Delimited Text Layer or click the "Open Data Source Manager" icon and select the Delimited Text tab.

Select the file: Click the "..." Browse button to locate your delimited text file (e.g., a CSV).

Specify Layer Name: Provide a name for the layer, which will appear in the Layers panel.

Configure File Format: If using a CSV file, select "CSV (comma separated values)".

Define Geometry: Choose "Point coordinates" as the geometry type. Select the column containing the Longitude as the X field and the column containing the Latitude as the Y field.

Set Geometry CRS (Coordinate Reference System): Select the Coordinate Reference System (CRS) that matches your data. If your coordinates are in decimal degrees (unprojected), choose WGS 84 (EPSG:4326). If you have projected coordinates (e.g., UTM zone), select the correct CRS for your data.

Add the Layer: Click the "Add" button, and your points will be displayed in the QGIS map canvas.

Save as Shapefile: save the delimited text layer to a permanent shapefile.

The CSV file should have columns corresponding to the water quality parameters to estimate, i.e., the column corresponding to turbidity should be named **Turbidity**, the column corresponding to the chlorophyll should be named **Chlorophyll**, and the column corresponding to CDOM should be named **CDOM**. A sample dataset (TestData57.csv) is provided.

	A	B	C	D	E	F
1	Site	Turbidity	Chlorophyll	CDOM	Lat	Lon
2	TB14-a	16.91462	4.523076923	48.01154	32.52021	-87.87505
3	TB14-b	16.92333	4.522666667	48.01467	32.52022	-87.87505
4	TB14-c	16.92097	4.627419355	48.12903	32.52023	-87.87505
5	TB11-a	15.77167	5.47	49.63667	32.51756	-87.86944
6	TB11-b	15.79961	5.400784314	49.60235	32.51757	-87.86944
7	TB12-a	16.0375	2.505	46.4375	32.51833	-87.86981
8	TB12-b	16.20286	2.654285714	46.81714	32.51834	-87.86982
9	TB12-c	16.68909	3.179090909	47.58545	32.51835	-87.86983
10	TB12-d	16.81769	3.306153846	48.08731	32.51836	-87.86986
11	TB12-e	16.75	2.995384615	47.85615	32.51837	-87.86989

Because Shapefile field names are limited to a maximum of 10 characters. After importing to shapefile, the “Chlorophyll” field name will be truncated to “Chlorophyl”.

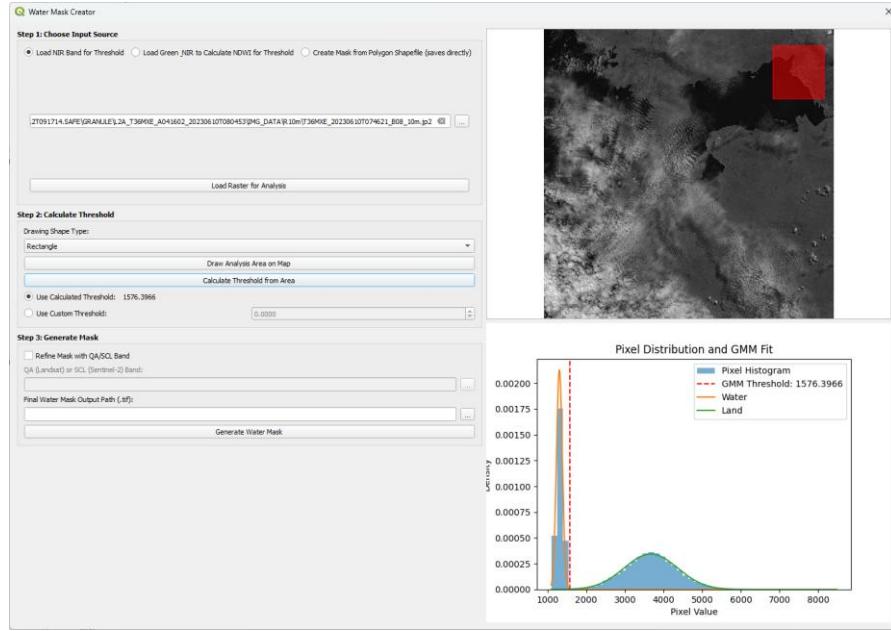
2) If you plan to use the ACOLITE atmospheric correction function available in the plug in, please download Landsat or Sentinel -2 **Level 1** data product. If not, please download **Level 2** surface reflectance product instead. If you download level 2 data product, please stack individual images into a multiband tiff file. For Landsat 5 &7, it should include Band 1 to Band 4 in order. For Landsat 8 &9, it should include Band 1 to Band 5 in order. For Sentinel-2, it should include Band 1 to Band 8 and Band 8A in order. Sample Landsat 8 dataset composite.tif is provided in the SampleDataSet folder.

3. Run the plugin: Go to QGIS menu->Plugins, RS-WaterQuality Mapper

The following section will introduce each tool in the plug in.

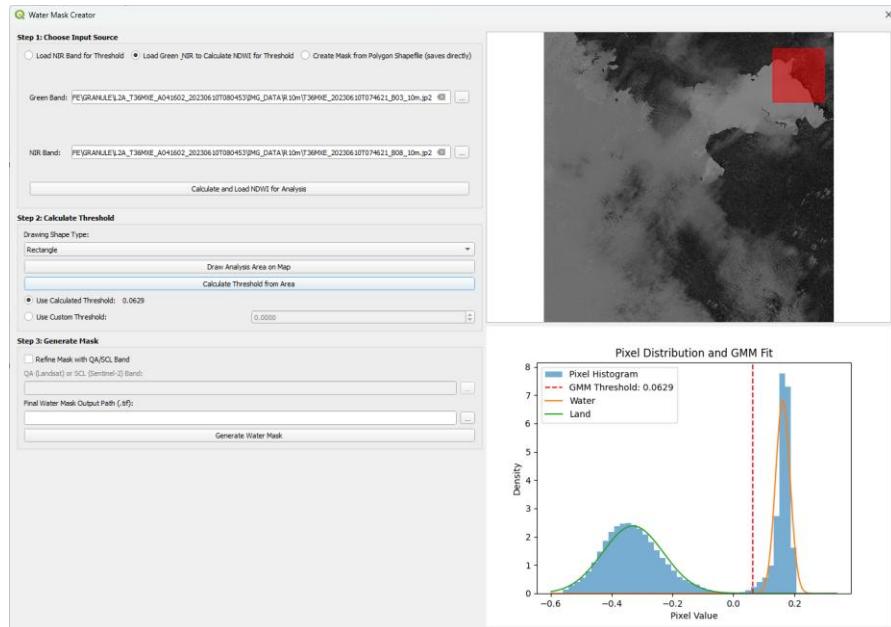
1) Water Mask Generation

There are three options:



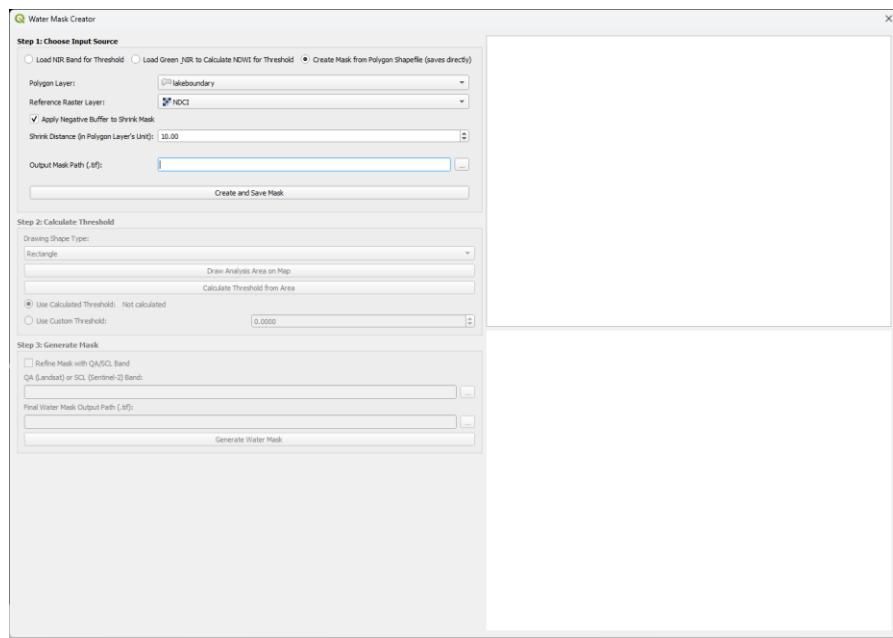
Option 1) Using NIR band to threshold

Once NIR band is loaded, click “Draw analysis area on Map”. Draw a rectangle or polygon on map. If using polygon option, double click to finish drawing. Click “Calculate threshold from area”. You can use the calculated threshold or use your custom threshold to generate the water mask. You can also load the QA band from Landsat or SCL band from sentinel -2 to refine the water mask in step 3.



Option 2: Use Green and NIR band to calculate NDWI for threshold.

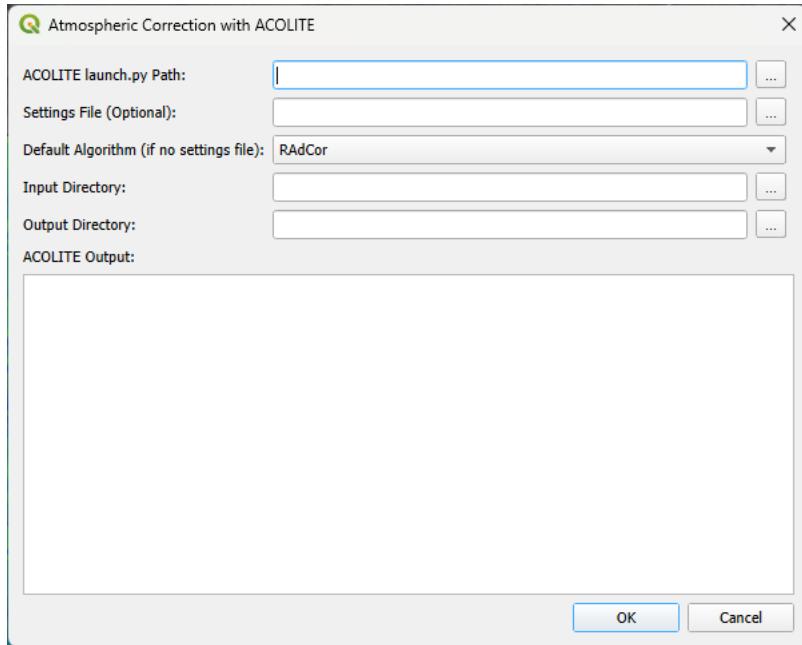
Once loaded required bands (can be level 1 or level 2 product), click on “Calculate and load NDWI for Analysis”. Draw a rectangle or polygon on map. If using polygon option, double click to finish drawing. Click “Calculate threshold from area”. You can use the calculated threshold or use your custom threshold to generate the water mask. You can also load the QA band from Landsat or SCL band from sentinel -2 to refine the water mask in step 3.



Option 3: Create water mask from polygon Shapefile

If you have a water boundary shapefile, you can convert it to a water mask tiff image using this option. To do this, a reference raster layer is needed. The reference raster layer could be any of your input band from Landsat or sentinel 2 data. The pixel size, coordinate system and extent of the water mask will be the same as the reference raster layer. There is an option to apply a negative buffer to shrink the polygon before converting it to a water mask to eliminate mixed pixels.

2) Atmospheric Correction with ACOLITE



Input

- ACOLITE launch.py path. ACOLITE can be downloaded from <https://github.com/acolite/acolite>. launch.py is within the unzipped folder of ACOLITE.
- Setting File (Optional). If no Setting File is provided by user, the default setting will be loaded as follows:

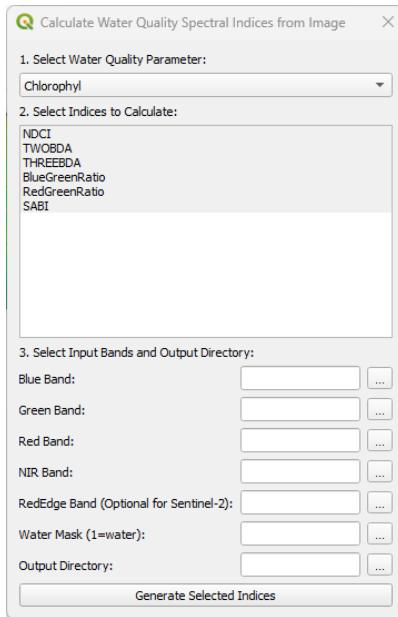

```

polygon=None
atmospheric_correction=True
atmospheric_correction_method=radcor
tact_run=False
l2w_parameters=rhos_*
l2r_export_geotiff=True
l1r_export_geotiff=False
l2w_export_geotiff=False
rgb_rhos=True
map_l2w=False
      
```
- Input directory: this is the directory where you save unzipped Landsat or Sentinel -2 Level 1 product.
- Output directory

Output

- The output includes a level 2 reflectance file in NetCDF format (file name ending with L2R.nc) and individual ***_L2R_rhos_***.tif files corresponding to each band. These two will be used in the later steps.

3) Calculate Water Quality Spectral Indices from Image



Input

- Water Quality Parameter: choose from the drop list
- Spectral Indices: choose from the list. Press control key and use mouse to select more than one.

Indices for Chlorophyll include NDCI, TWOBDA, THREEBDA, BlueGreenRatio, RedGreenRatio, and SABI.

$$\text{NDCI} = (\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$$

$$\text{TWOBDA} = \text{NIR}/\text{Red}$$

$$\text{THREEBDA} = (\text{Blue}-\text{Red})/\text{Green}$$

$$\text{BlueGreenRatio} = \text{Blue}/\text{Green}$$

$$\text{RedGreenRatio} = \text{Red}/\text{Green}$$

$$\text{SABI} = (\text{NIR}-\text{Red})/(\text{Blue}+\text{Green})$$

Indices for Turbidity include RedBlueRatio, NDTI, GreenBlueRatio, and RedPlusNIR.

$$\text{RedBlueRatio} = \text{Red}/\text{Blue}$$

$$\text{NDTI} = (\text{Red}-\text{Green})/(\text{Red}+\text{Green})$$

$$\text{GreenBlueRatio} = \text{Green}/\text{Blue}$$

$$\text{RedPlusNIR} = \text{Red}+\text{NIR}$$

Indices for CDOM include GreenBlueRatio, RedBlueRatio, RedMinusGreen, and GreenRedRatio.

$$\text{GreenBlueRatio} = \text{Green}/\text{Blue}$$

$$\text{RedBlueRatio} = \text{Red}/\text{Blue}$$

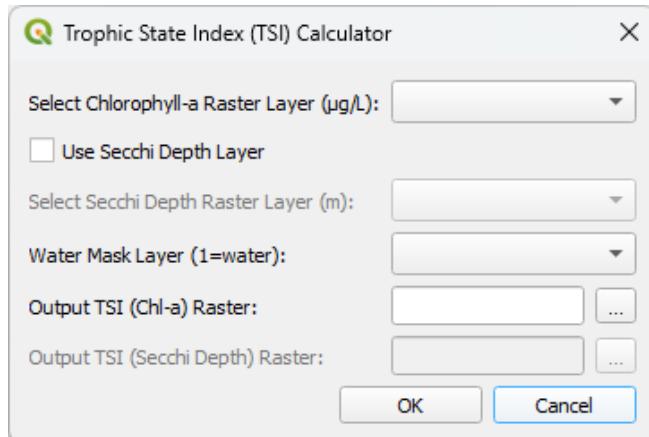
$$\text{RedMinusGreen} = \text{Red}-\text{Green}$$

$$\text{GreenRedRatio} = \text{Green}/\text{Red}$$

- Specify the file corresponding to Blue, Green, Red, NIR, RedEdge (optional) bands and Water Mask.
- Output directory for saving Spectral Indices.

Indices will be automatically loaded to the current QGIS project after they are generated.

4) Trophic State Index (TSI) Calculator



Input

- Chl-a layer in ug/L unit. The layer should be in the project already. If not, load it to the project before running the tool.
- Users can also load an optional Secchi Depth layer to compute TSI(Secchi Depth) by checking the “Use Secchi Depth Layer” box. The toolbox calculates the trophic state for each pixel using the formula:

$$TSI(\text{Chl-a}) = 9.81 * \ln(\text{Chl} [\mu\text{g/L}]) + 30.6$$

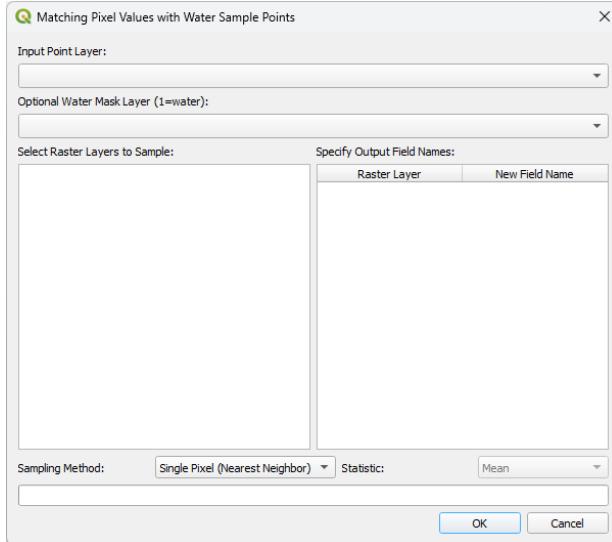
$$TSI(\text{Secchi Depth}) = 60 - 14.41 * \ln(\text{Secchi [m]})$$

- Water Mask

Output

- TSI(Chl-a) layer and optional TSI(Secchi Depth) layer (if Secchi Depth layer was provided).

5) Matching Pixel Values with Water Sample Points



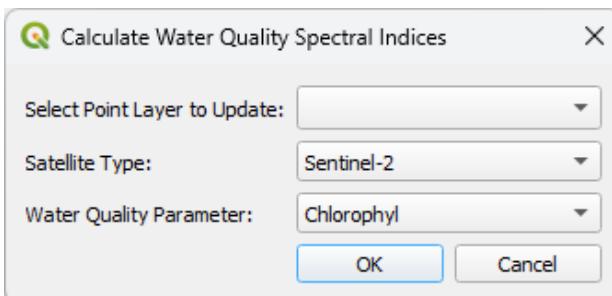
Input

- Point layer in the project that has attribute of water quality parameters (Turbidity, CDOM, or Chlorophyl).
- Optional water mask. Only points within the water mask (pixel value =1) will get values from the selected raster layers if water mask layer is selected.
- Select Raster Layer to Sample. You select multiple layer by pressing Control key and click the layers one by one. Once the layers are selected, change the New Field Name to B1, B2, B3... for the corresponding surface reflectance layer TIFF files. For example, B1 for Landsat 8 Band 1 443nm or Band 8A for Sentinel-2 865 nm. Required fields: B1-B4 for Landsat 5-7, B1-B5 for Landsat 8 and 9. B1-B8 and B8A for Sentinel-2.
- Sampling Method: Choose from options of “Single Pixel”, “3*3 Window” or “5*5 Window”. If “3*3 Window” or “5*5 Window” options are chosen, choose Mean or Median statistics. The extracted value will be the mean or median of pixels within the window.

Output

- No output. Changes will be applied to the input point layer. New fields will be added to the attribute table. Open the attribute table of the point layer to check.

6) Calculate Water Quality Spectral Indices for the Sample Points



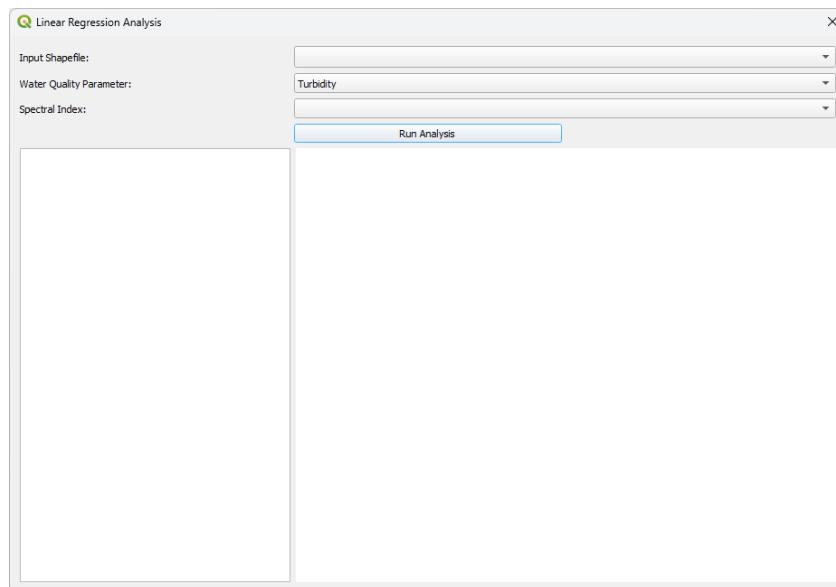
Input

- Point layer to be update. This layer should be the same layer from tool “**Matching Pixel Values with Water Sample Points**” after that tool has been applied to the layer.
- Satellite Type: Choose from list of Landsat 5/7, 8/9 and Sentinel-2.
- Water Quality Parameter: Choose from Turbidity, Chlorophyl, and CDOM. The following new fields for Turbidity will be added to attribute tables: RedBlueRatio, NDTI, GreenBlueRatio, and RedPlusNIR. The following new fields for Chlorophyll will be added to attribute tables: NDCI, TWOBDA, THREEBDA, BlueGreenRatio, RedGreenRatio, and SABI. The following new fields for CDOM will be added to attribute tables: GreenBlueRatio, RedBlueRatio, RedMinusGreen, and GreenRedRatio.

Output

- No output. Changes will be applied to the input point layer. New fields will be added to the attribute table. Open the attribute table of the point layer to check.

7) Empirical Modelling

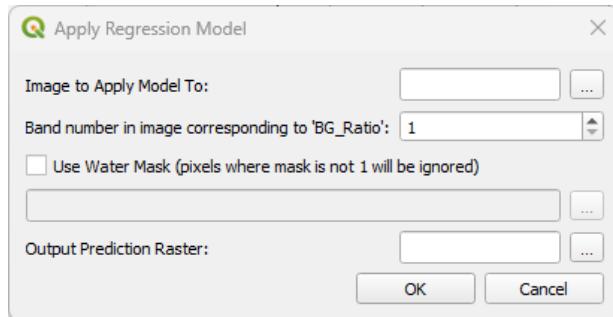


Input

- Input Shapefile: This layer should be the same layer from tool “**Calculate Water Quality Spectral Indices for the Sample Points**”.
- Water Quality Parameter: choose from the drop list.
- Spectral Index: choose from the drop list. The list includes all spectral indices pertaining to selected water quality parameter available from the attribute table of the input layer. Click Run Analysis. You can switch spectral index and run the analysis multiple times.

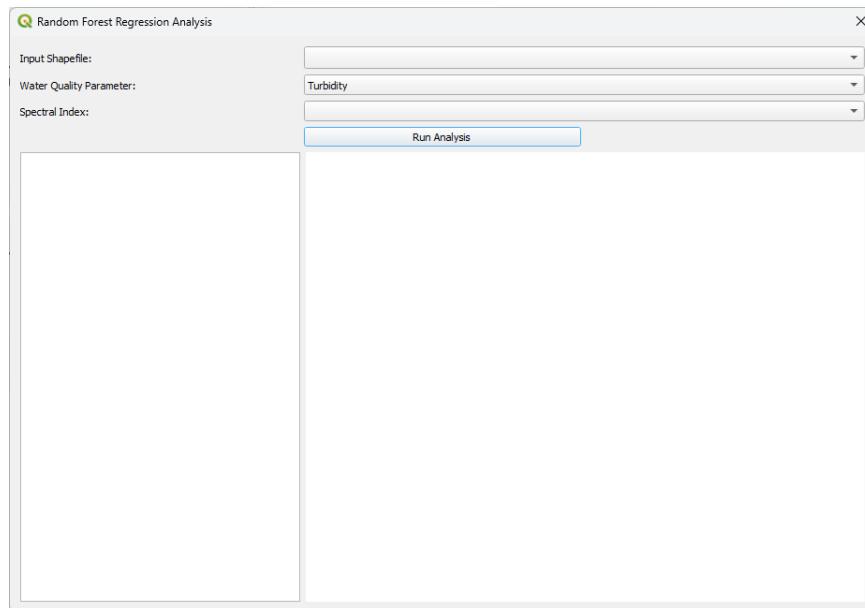
Output

- After clicking on Run Analysis, the lower left side will show the Regression Equation and Performance Metrics including R-squared, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The lower right side will show the scatterplot of actual value vs predicted value.
- After clicking on **Run Analysis**, the **Apply Model to Image** button will appear. Clicking it will get a new window **Apply Regression Model**:



Specify file name and location of the image you want to apply the model to (e.g., the corresponding spectral index from tool “**Calculate Water Quality Spectral Indices from Image**”). Specify output file name and location and use water mask as needed.

8) Random Forest Regression Analysis



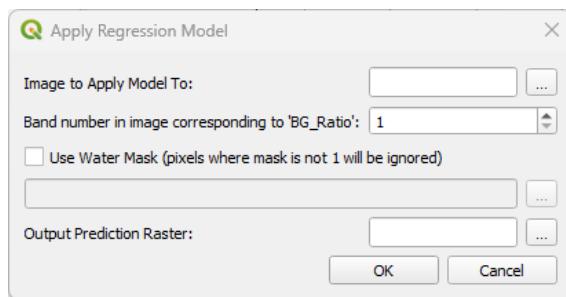
Input

- Input Shapefile: This layer should be the same layer from tool “**Calculate Water Quality Spectral Indices for the Sample Points**”.
- Water Quality Parameter: choose from the drop list.

- Spectral Index: choose from the drop list. The list includes all spectral indices pertaining to selected water quality parameter available from the attribute table of the input layer. Click Run Analysis. You can switch spectral index and run the analysis multiple times.

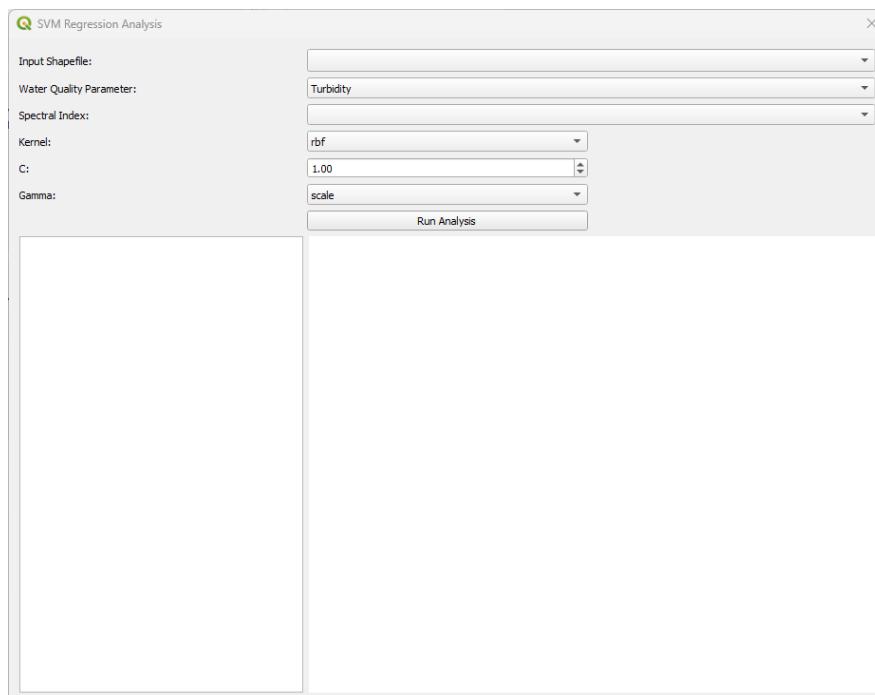
Output

- After clicking on Run Analysis, the lower left side will show Model Specifications ($n_{\text{estimators}}$) and Performance Metrics including R-squared, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The lower right side will show the scatterplot of actual value vs predicted value.
- After clicking on **Run Analysis**, the **Apply Model to Image** button will appear. Clicking it will get a new window **Apply Regression Model**:



Specify file name and location of the image you want to apply the model to (e.g., the corresponding spectral index from tool “**Calculate Water Quality Spectral Indices from Image**”). Specify output file name and location and use water mask as needed.

9) SVM Regression Analysis

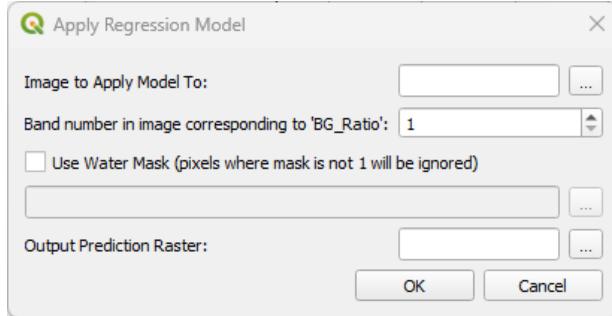


Input

- Input Shapefile: This layer should be the same layer from tool “**Calculate Water Quality Spectral Indices for the Sample Points**”.
- Water Quality Parameter: choose from the drop list.
- Spectral Index: choose from the drop list. The list includes all spectral indices pertaining to selected water quality parameter available from the attribute table of the input layer. Click Run Analysis. You can switch spectral index and run the analysis multiple times.
- Kernel: The kernel function transforms the input data into a higher-dimensional space to make it easier to find a linear boundary (or hyperplane) for regression. It defines how the model measures similarity between data points, enabling SVR to handle non-linear relationships. Radial Basis Function (RBF): Most common, uses a Gaussian function to handle non-linear data (`kernel='rbf'`). Linear: No transformation, suitable for linearly separable data (`kernel='linear'`). Polynomial: Maps data to a polynomial space (`kernel='poly'`).
- C: The C parameter controls the trade-off between achieving a low error on the training data and minimizing the complexity of the model (margin maximization). High C: Prioritizes fitting the training data closely, potentially leading to overfitting (less margin, more sensitive to outliers). Low C: Emphasizes a larger margin, allowing some errors (more robust to outliers, risks underfitting). Typical Range: Values like 0.1, 1, 10, 100 are tested (often tuned via cross-validation).
- Gamma: The gamma parameter defines how far the influence of a single training example reaches in the kernel function (primarily for RBF, polynomial, and sigmoid kernels). High Gamma: Points need to be very close to influence each other, leading to a more complex model that may overfit. Low Gamma: Points farther away influence each other, creating a smoother decision boundary that may underfit. 'auto': gamma is set to $1 / n_features$, where $n_features$ is the number of features in the input data. Provides a simple heuristic for gamma based only on the number of features, ignoring data variance. 'scale': gamma is calculated as $1 / (n_features * X.var())$, where $n_features$ is the number of features in the input data X , and $X.var()$ is the variance of the input data. Automatically adjusts gamma based on the data's feature count and variance, making it data-dependent and more robust to variations in dataset scale.

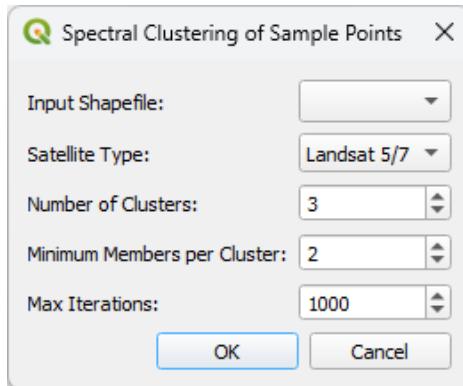
Output

- After clicking on Run Analysis, the lower left side will show Model Specification, Number of support vectors and Performance Metrics including R-squared, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The lower right side will show the scatterplot of actual value vs predicted value.
- After clicking on **Run Analysis**, the **Apply Model to Image** button will appear. Clicking it will get a new window **Apply Regression Model**:



Specify file name and location of the image you want to apply the model to (e.g., the corresponding spectral index from tool “**Calculate Water Quality Spectral Indices from Image**”). Specify output file name and location and use water mask as needed.

10) Spectral Clustering of Sample Points



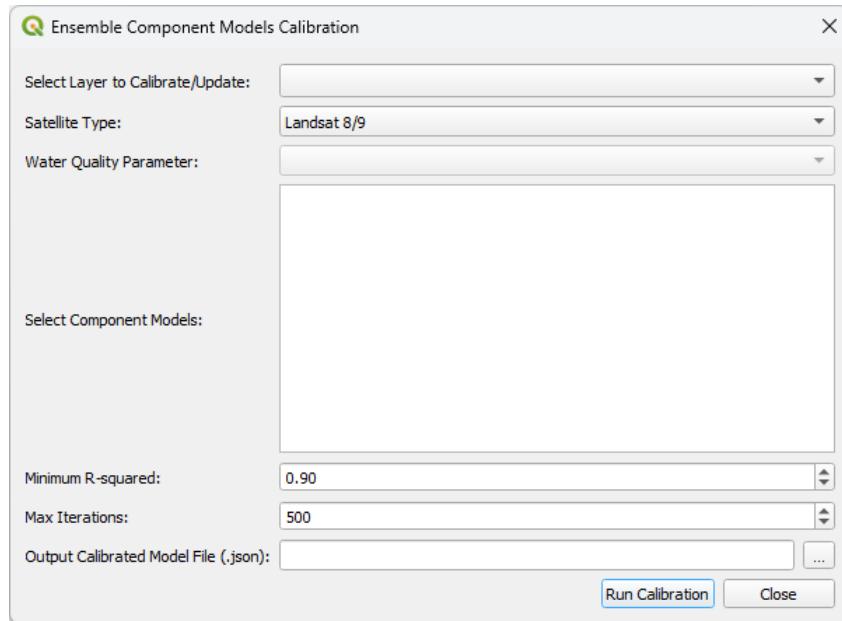
Input

- Input Shapefile: this is the Shapefile that tool “**Calculate Water Quality Spectral Indices for the Sample Points**” has been applied to.
- Satellite: choose from the drop list.
- Cluster Number: how many clusters will the records in the table be classified to using ISODATA clustering algorithm, default value is 3.
- Minimum member in each Cluster: default value is 2.
- Max iteration: maximum iteration for the ISODATA clustering algorithm, default value is 1000

Output

- No output. Changes will be applied to the input point layer. New field “cluster” will be added to the attribute table. Open the attribute table of the point layer to check.

11) Ensemble Component Models Calibration and Selection



Input

- Select Layer to Calibrate/Update: this is the Shapefile that tool “**Spectral Clustering of Sample Points**” has been applied to.
- Satellite Type: choose from the drop list.
- Water Quality Parameter: choose from the drop list. It will only show the Water Quality Parameter available in the attribute table of the input layer.
- Select Component Models: this will show the available spectral indexes in the attribute table of the input layer pertaining to the water quality parameters selected. You can select multiple models by pressing Control key when selecting.
- Minimum R2: the iteration will stop when the best linear regression model for each cluster has R2 higher than this value.
- Max Iteration: maximum iteration for the water sample data points to be reassigned to different best model

Output

- Output Calibration Model File (.json): specified the output location and file name for the Jason file to save ensemble component models to be reused in “**Apply Ensemble Model to an Image**” tool later.
- A scatterplot will show up at the bottom of the window when calibration is done. It shows one of the three reasons the calibration is stopped: min R-square achieved, max iteration reached, or the model itself converges (model assignment is not changing for any data point). You can change the R2 and max iteration to run the model multiple times and check the scatter plot until getting satisfying results.

12) Spectral Space Partition and Formation of Ensemble Model



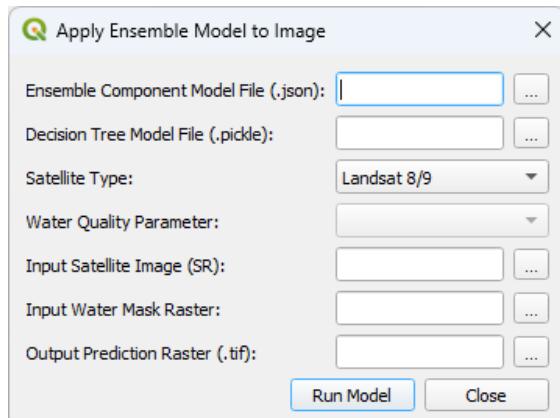
Input

- Input Calibrated Shapefile: this is the Shapefile that tool “**Ensemble Component Models Calibration and Selection**” has been applied to.
- Satellite Type: choose from the drop list.
- Water Quality Parameter: choose from the drop list. It will only show the Water Quality Parameter available in the attribute table of the input layer.
- Minimum Samples per Node: minimum samples in each node in the decision tree for spectral space partition, recommended value: 1 to 5. Default value 2.
- Max Tree Depth: maximum depth of the decision tree, recommended value: 4 to 10. Default value 10.

Output

- Output Decision Tree Image: output PNG file to save the decision tree as a graph
- Output Decision Tree Model (.pickle): output pickle file to save the decision tree to be reused in **Apply Ensemble Model to an Image** tool.
- After running the tool, the decision tree PNG image will be loaded at the button of window.

13) Apply Ensemble Model to an Image



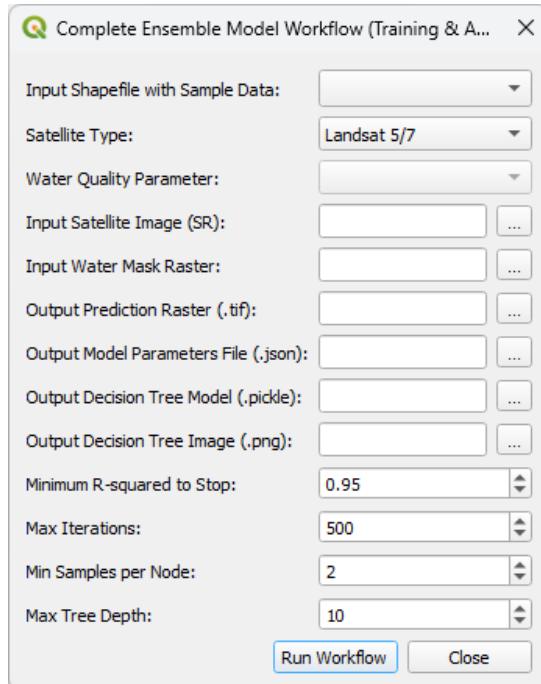
Input

- Ensemble Component Models json File: this is the output json file from the **Ensemble Component Models Calibration and Selection** tool.
- Decision Tree Pickle File: this is the Pickle file from the **Spectral Space Partition and Formation of Ensemble Model** tool.
- Satellite: choose from the drop list.
- Water Quality Parameter: choose from the drop list. It will only show the available water quality parameter in the json file.
- Input Satellite Image (SR): this is the level 2 surface reflectance image. It could be the output NetCDF format file (name ending with L2R.nc) from ACOLITE or multiband composite of individual ***_L2R_rhos_ * *.tif files corresponding to each band from ACOLITE output. For Landsat 5 & 7, it should include Band 1 to Band 4 in order. For Landsat 8 & 9, it should include Band 1 to Band 5 in order. For Sentinel-2, it should include Band 1 to Band 8 and Band 8A in order. Sample Landsat 8 dataset Landsat8_5Bands.tif is provided.
- Water Mask Image

Output

- Output Prediction Raster (.tif): folder and file name for output for output Water Quality Prediction image TIFF file. The output raster will be loaded automatically to the project when it is saved.

14) Integrated Ensemble Modelling



Input

- Input Shapefile with Sample Data: this is the Shapefile that has already been applied **Matching Pixel Values with Water Sample Points** tool. The attribute table should have spectral indices corresponding to water quality parameters.
- Satellite Type: choose from the drop list.
- Water Quality Parameter: choose from the drop list. It will only show the water quality parameters available in the attribute table of input Shapefile.
- Input Satellite Image (SR): this is the level 2 surface reflectance image. It could be the output NetCDF format file (name ending with L2R.nc) from ACOLITE or multiband composite of individual ***_L2R_rhos_*_.tif files corresponding to each band from ACOLITE output. For Landsat 5 & 7, it should include Band 1 to Band 4 in order. For Landsat 8 & 9, it should include Band 1 to Band 5 in order. For Sentinel-2, it should include Band 1 to Band 8 and Band 8A in order. Sample Landsat 8 dataset composite.tif is provided in SampleDataSet folder.
- Water Mask Image: prepared by tool **Water Mask Generation**.
- Minimum R2: the iteration will stop when the best linear regression model for each cluster has R2 higher than this value.
- Max Iteration: maximum iteration for the water sample data points to be reassigned to different best model.
- Minimum Samples per Node: minimum samples in each node in the decision tree for spectral space partition, recommended value: 1 to 5. Default value 2.
- Max Tree Depth: maximum depth of the decision tree, recommended value: 4 to 10. Default value 10.

Output

- Output Prediction Raster: folder and file name for output Water Quality Prediction image TIFF file. The output raster will be loaded automatically to the project when it is saved.
- Output Model Parameters File (.json): specified the output location and file name for the Jason file to save ensemble component models.
- Output Decision Tree Model: folder and file name for the output pickle file to save the decision tree.
- Output Decision Tree Image: folder and file name for the output PNG file to save the decision tree as a graph.