**BROSS Technical Background**

*Written by Claire Beveridge of University of Washington SASWE Research Group.*

*Reviewed by Dr. Robin Biswas?*

*Originally published on April 14, 2020. Last modified on April 14, 2020.*

1. **Introduction**

This manual provides technical background on the BROSS (Bangladesh remote sensing of suspended sediment). BROSS was developed for the Bangladesh Water Development Board (BWDB) by the University of Washington SASWE Research Group. BROSS provides SSC predictions for the major rivers of Bangladesh using satellite remote sensing imagery in the visible (red, green, blue) and near-infrared (NIR) spectra. It is intended that BROSS will help BWDB with sediment monitoring and management in major rivers.

In the text that follows, Section 2 provides background theory on satellite remote sensing of sediment. Section 3 describes how BROSS was calibrated. Section 4 details limitations and uncertainty of BROSS outputs.

Information on how to operate BROSS are found in the accompanying User Manual. Further details on the stakeholder engagement in developing BROSS will be found in a forthcoming journal publication.

1. **Background theory on satellite remote sensing of sediment**

# Satellite remote sensing offers a practical response to the sustainable sediment monitoring and management needs. Satellite remote sensing offers extensive spatial coverage, frequent and extensive temporal records, cost-effectiveness, and readily transferable data and methods. Satellite remote sensing has been broadly applied in literature for monitoring suspended sediment concentrations (SSC) of water bodies due to the relationship between SSC and satellite remote sensing visible and near infrared (NIR) bands (e.g., Ritchie et al., 1987; Pavelsky and Smith, 2009; Zhang et al., 2014; Gholizadeh et al., 2016; Yepez et al., 2018). Essentially, higher levels of suspended sediment increase visible and NIR surface reflectance due to the backscattering effect of sediment in the water column (Peterson et al., 2018; Ritchie et al., 2003).

Approaches for estimating SSC from satellite remote sensing are generally either empirical or physics-based (Wackerman et al., 2017). We used two empirical approaches since there is insufficient Bangladesh river sediment data available to properly parametrize physics-based models. One empirical approach was a regression between in situ SSC and remote sensing visible and NIR data. We used this technique because of its simplicity and widespread application. A second approach was the machine learning technique of artificial neural networks (ANN), which is more advanced and has been found to be stronger than traditional regression techniques (Peterson et al., 2018). We used this technique because it is well-suited for the highly complicated nature of Bangladesh rivers.

* 1. **Regression**

Linear regression techniques for estimating SSC have commonly used the visible (red, green, blue) and NIR bands of the Landsat satellite series (TM, ETM+, OLI) to correlate to in situ SSC measurements (Gholizadeh et al., 2016). Generally, regression has been conducted between in situ measurements and a single band or band ratio, with the values in linear or exponential form. The red band (alone or in a ratio) has been used most often. Using band ratios has been more robust than using single bands, particularly when sediment color varies (Pavelsky and Smith, 2009). Regression relationships have typically been exponential, linear, or 2nd order polynomial (higher order polynomials often overfit). Exponential relationships have often been strongest, particularly for high SSC (>50 mg/L) (Pavelsky and Smith, 2009; Wackerman et al., 2017).

The only previous study we found that used regression to predict SSC from surface reflectance in Bangaldesh inland waters was by Islam et al. (2000). They used four data points from Bahadurabad and Hardinge Bridge in situ monitoring points that were collected in year 1991, and one data point from another river system. They found a linear relationship using the red band to be optimal. However, when we tested this relationship we found that the range of predicted values was low compared to the range of observed SSC values. We aimed to develop a stronger realtionship using additional data collected since year 1991.

* 1. **Artificial Neural Networks (ANN)**

ANN models can capture both linear and non-linear relationships and are generally found to be superior to and more flexible than regression techniques (Peterson et al., 2018; Sudheer et al, 2006). Although relatively few studies have applied ANN techniques to remote sensing of river suspended sediment, the results are promising (Peterson et al., 2018). However, because of their complexity, ANN models can be “black boxes.” It is important that those who are developing ANN models understand the theoretical foundation and ANN modeling approaches (see Maier and Dandy, 2000; Loy, XXXX [online book]).

ANNs consist of several layers with interconnected nodes and are highly customizable. The most common form of an ANN model is the “feedforward neural network” in which model information is passed forward through nodes, not looped. Information is first passed to the model input layer and then through a series of hidden layers. Each layer has a specified dimensionality, taking in a set of weighted inputs and generating outputs based on a specified activation function.

Developing a predictive ANN model consists of several stages and decisions. Following the logic used in developing regression models, ANN inputs should be those that have a strong correlation with SSC. Hence, individual bands and band ratios can be inputs along with other band combinations (e.g., total visible reflectance; red/green + NIR). Other spatial (e.g., sample monitoring station) and temporal (e.g., sample month) information can also be inputs.

1. **Data**
   1. **In situ data**

There are four in situ monitoring points along the main stem rivers for which suspended sediment and streamflow discharge data is collected by BWDB (Figure 1): Bahadurabad (Brahmaputra-Jamuna river), Hardinge Bridge (Ganges-Padma river), Baruria (Padma river) and Mawa (Padma river). Typically, suspended sediment data is typically collected using a Binkley silt sampler. A river cross-section is divided into approximately 25 sub-sections such that the streamflow in each section is not higher than 10% of the total cross-section streamflow. Then, suspended sediment samples are collected in each sub-section at alternating depths of 0.2D and 0.8D, where D is the total sub-section depth. The total suspended sediment discharge is then determined using all sub-section samples (personal communication, Dr. Robin Biswas, September 19, 2019).

We were provided with the complete suspended sediment discharge, Qs (units of kilograms per second [kg/s]), and streamflow discharge, Q (units of cubic meters per second, [m3/s]), datasets for these four stations. The suspended sediment discharge data was provided as volumes of “sand” and “fine” sediment, and we combined these values to get the total suspended sediment discharge. We computed the suspended sediment concentration (SSC) in units of milligrams per liter (mg/L) using the equation:

x [1]

There was not a Q measurement for every Qs measurement and vice versa. A summary of the SSC data is shown in Table 1. We analyzed the SSC data for quality control and excluded SSC values which were anomalously high or low.

***TABLE 1: Summary of in SSC data***



* 1. **Satellite data**

Satellite data was downloaded and processed using Google Earth Engine (GEE), a cloud-based remote sensing platform (Gorelick et al., 2017). We used GEE for its multiple advantages including high-performance cloud-computing, free availability, and strong user-interface capabilities. Throughout the entire application computation processes, Google’s cloud infrastructure powers all data processing, stores all satellite datasets and data outputs, and immediately displays processed data on an interactive map. Users can also export data in various file formats. GEE and all the datasets it contains are freely available for non-commercial use. GEE programming is conducted in Javascript, however customized applications can readily be developed and deployed so that non-programmers can easily operate them with a simple graphical user interface.

BROSS integrates data from multiple satellites, which are: Landsat 5 (TM), 7 (ETM+), and 8 (OLI); Sentinel-2; and MODIS (MOD09GA). Each of these satellites provide data in the visible (red, green, blue) and NIR wavelengths. However, the band wavelength ranges vary between satellites. These satellites also have different spatial resolution, temporal resolution, and temporal availability, as presented in Table 2. Although each Landsat satellite has a 16-day revisit interval, there is an 8-day offset between any two satellites that have overlapping operational periods. Landsat 7 has limited availability since 2003, when a failure of the Scan Line Corrector occurred (Chander et al., 2009). The calibration of Landsat 8 is significantly improved compared to Landsat 5 and 7 (Mishara et al., 2016).

***TABLE 2: Satellite data used in BROSS***



Because of the differences between the satellites’ band wavelength ranges and spatial resolutions, it was necessary for separate regression and ANN models to be developed. We developed separate models for three distinct satellite groups:

1. Landsat 5 and 7: These have the same spatial resolution and nearly identical band ranges (except for a ~1 μm difference in the NIR (band 4) and shortwave infrared (band 7) ranges).
2. Landsat 8 and Sentinel-2: Although their data properties are notably different, the data were corrected for differences so that they could be combined using algorithms developed by Zhang et al. (2018). There was also an insufficient number of Sentinel-2 data points to develop a reliable algorithm for this satellite alone.
3. MODIS

We used data satellite data with the highest quality rating for Landsat satellites (Tier 1), however this tiered rating was not an option for the other satellites. We used surface reflectance products from GEE that had already been atmospherically corrected. Although the Sentinel-2 satellite was launched in 2014, the surface reflectance records are available in GEE starting in 2017. We harmonized the Sentinel-2 data with the Landsat 8 data using the equations provided in Zhang et al., (2018) (Table 3 in referenced publication). The surface water pixels were mapped for Landsat and Sentinel-2 by applying the Dynamic Surface Water Extent algorithm (Jones, 2015). For MODIS, we mapped surface water pixels used the index ranges for permanent water that were determined for Bangladesh rivers by Kwak et al. (2013).

Pixels with clouds and cloud shadows (except for Sentinel-2 for the latter) were masked out of scenes to the extent feasible by combining multiple methods. One method was using the quality assessment (QA) bands of the surface reflectance datasets. Each dataset has QA bands for clouds, and the Landsat and MODIS datasets also have QA bands for cloud shadows. Another method was masking out pixels for which the sum of the red and blue bands was above a threshold (0.4 for Landsat, 0.6 for Sentinel-2, 0.9 for MODIS). Each threshold was determined through manual calibration. The final methods was masking out pixels for which any surface reflectance band (red, blue, green, or NIR) was greater than 0.4. We note however that these methods for masking out clouds and cloud shadows are not perfect and that cloud impacts remain a limitation of BROSS, particularly in the wet season.

Developing the calibration dataset required having the surface reflectance time series at each of the four in situ monitoring stations. We collected the red, blue, green, and NIR data spatially-averaged over a stream reach roughly three times as long as the stream width at each station (Figure 1). We retained samples with at least 30% of pixels classified as being free of clouds and cloud shadows over the sample reach.

1. **Calibration**

We extensively explored options for calibrating the empirical relationship between SSC and surface reflectance to be used in BROSS. We selected options which had the strongest performance and made sense physically. The options we explored and selected are the following:

* **In situ monitoring stations:** These were tested individually and combined. **We ultimately decided to use combined data from the Baruria and Mawa monitoring stations.** These stations are along the Padma river (i.e., junction of Ganges-Padma and Brahmaputra-Jamuna) and therefore represent sediment properties of both upstream mainstems rivers. As we learned from BWDB, data from Baruria is most reliable. However, using Baruria data alone made the calibration dataset size undesirably small, so we added data from Mawa. Including data from the other monitoring stations weakened the results, likely because the in situ and satellite data for these points was lower quality.
* **Satellites:** There were tested individually (e.g., MODIS) and combined (e.g., all Landsat and Sentinel-2 combined). **We ultimately decided to calibrate separately for three groups: (1) Landsat 8 and Sentinel-2; (2) Landsat 5 and 7; and (3) MODIS**. These groupings are discussed above (Section 3.2).
* **Satellite data inputs:** These were tested using single surface reflectance bands (red, green, blue, NIR), band ratios (red/green, red/blue, red/NIR, green/NIR), and other band combinations that have been found to be significant in literature ((red+green)/2, red/green + NIR, Peterson et al. 2018). **The inputs varied between regression and ANN and for each satellite group.**
* **Data overlap:** These were tested with in situ and satellite data collected on the same day, within ±1 day, and within ±2 days. **We ultimately decided to calibrate with in situ and satellite data collected within ±2 days.**
* **Seasons:** These were tested for seasons individually (e.g., wet, dry) and combined. **We ultimately decided to calibrate with the seasons combined.**
* **SSC ranges:** These were tested for SSC ranges separately (e.g., low: ≤100 mg/L and high: >100 mg/L) and combined. **We ultimately decided to calibrate with the SSC ranges combined.**
  1. **Regression**

We tested both the SSC and reflectance values in linear and exponential forms with exponential, linear, and 2nd order models. Values in linear form and in exponential models were the strongest performing. We randomly divided the dataset as 70% calibration points and 30% testing points. This random division was conducted 30 times for each satellite group. We selected the highest performing empirical equation and calibration/testing grouping for the 30 iterations. The code for this calibration is found at [INSERT GITHUB LINK].

* 1. **ANN**

We used the Keras package (<https://keras.io/>) in Python to develop and implement the ANN model. For all models, we used:

* Feedforward model architecture
* Sequential models, which are linear stacks of layers.
* Densely connected neural network layers
* Rectified linear unit activation functions, which are ideal for model non-linearity
* Scaled input data to have a zero mean and unit variance
* Optimization conducted using “adam” algorithm
* Optimization based on the mean-square-error metric

The top performing ANN model architecture and inputs slightly varied for each model, as shown in Table 3. We randomly divided the dataset as 70% calibration points and 30% testing points. This random division was conducted 10 times for each satellite group. We selected the highest performing ANN model for the 10 iterations. The codes for the calibrations are found at [INSERT GITHUB LINK].

***TABLE 3: Summary of ANN models’ architecture***



1. **Results**

Plots comparing the in situ SSC observations with regression and ANN models (training and testing data) are shown in Figure 2. The performance metrics for the regression and ANN models testing datasets (e.g., 30% of data that was used to test, not train, the models) are shown in Table 4. The five metrics we used to assess the models are:

* Root mean square error (RMSE): This is a common and strong hydrologic modeling metric that shows the prediction error in the terms of the variable being calculated (Ritter and Munoz-Carpena, 2013), which in our case is SSC (mg/L). RMSE is defined as

[2]

where *O* is the observations, *P* is model predictions, and *N* is the number of samples. The RMSE range is 0 to , and the lower the RMSE the stronger the results are.

* Coefficient of determination (r2): This is another common hydrologic modeling metric which quantifies the degree of collinearity between the observed and predicted SSC. It is defined as

[3]

The r2 range is 0 to 1, where 0 means no correlation and 1 means a perfect fit.

* Anomaly: This is the average deviation between the predicted SSC and the respective long-term observed monthly mean SSC. It is defined as

Anomaly [4]

where *M* is the long-term observed monthly mean SSC (i.e., based all al *O*) for the same month that each *P* is made. It is the same units as the variable being measured (SSC, mg/L). It ranges from to - to to and results are stronger when they are closer to 0. A positive anomaly means the prediction is higher than the observed monthly mean and a negative anomaly means the prediction is lower than the observed monthly mean.

* Anomaly correlation coefficient (ACC): This metric is common in weather forecasting and tests the correlation between the anomalies (see [4]) of observation and predictions. It is defined as

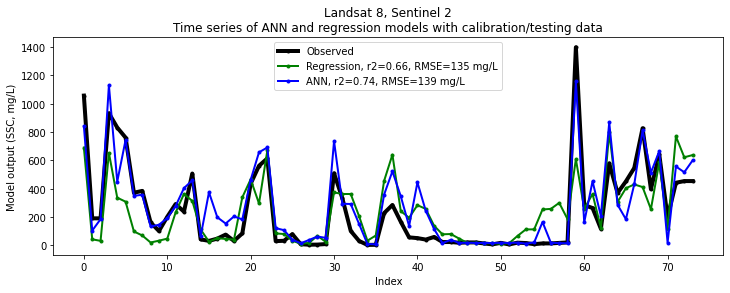
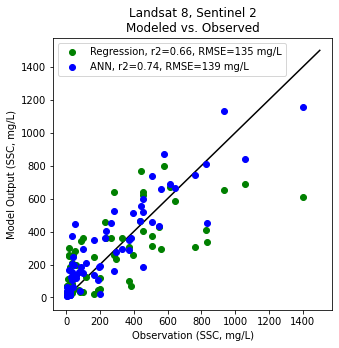
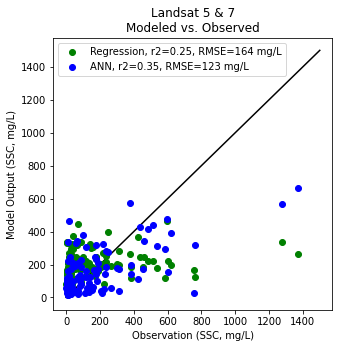
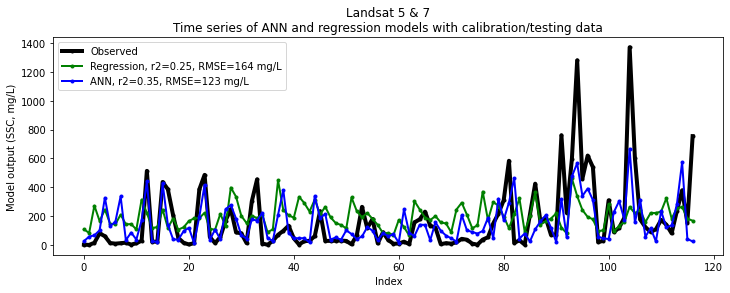
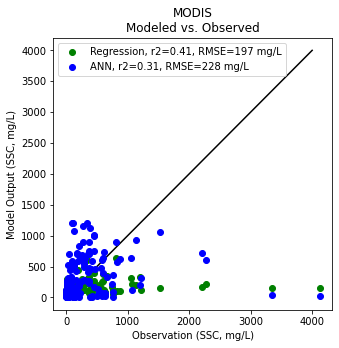
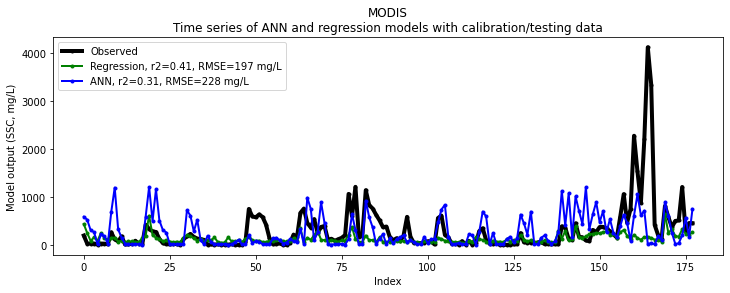
ACC*=*  [5]

The ACC range is -1 to 1, where 1 means perfect coincidence between the anomalies of the prediction and observations, -1 means that the anomalies of the prediction and observations are perfectly reversed, and 0 means no correlation (Japanese Meteorological Agency, 2013)

* Spearman r (*rs*): This is a nonparametric measurement based on the rank of data and is defined as

*rs =* 1- [6]

where *R(Oi)* and *R(Pi)* are *O* and *P* of corresponding ranks. Like ACC, the *rS* range is -1 to 1, where 1 means perfect correlation of the prediction and observations, -1 means that the correlation is perfectly reversed, and 0 means no correlation (Hamed, 2016). Because *rs* is based on ranked data, it captures the non-linearity of data.



***FIGURE 2: Comparison of SSC in situ observations with Regression and ANN models****.*

***TABLE 4: Summary of ANN and Regression model performance.***  *Stronger result for each satellite group is indicated in* ***bold***



As seen in Table 4 and Figure 2, the overall the performance of the ANN models is generally stronger than the regression models for the Landsat 5 + 7 and Landsat 8+ Sentinel-2 groups. For MODIS, the regression metrics are generally stronger or equal. This is likely because of the lower quality of MODIS satellite data due to larger spatial resolution and reduced ability to filter clouds and landscape. With the lower quality of data, the ability of ANN to capture patterns is poor and the more generalized trend captured by regression is stronger. It is also clear that the Landsat 8+Sentinel-2 models are generally stronger than Landsat 5+ Landsat 7 which is likely because of the improved calibration of Landsat 8 compared to previous Landsat satellites as well as the higher resolution of Sentinel-2.

Overall, the results are acceptable although not as strong as results in other studies (e.g., Peterson et al. (2018) ANN testing performance was r2 of 0.78-0.81 and RMSE of 71.7 mg/L-93.1 mg/L). There may be various reasons for this such as the complexity of the river system and the quality and consistency of the in situ and satellite data. The limitations and uncertainties are further discussed in section 6.

1. **Limitations and uncertainty**

Also, since SSC typically increases with depth, the shallow SSC samples likely underestimate the mean cross-section SSC (Walling, 2009).

ANN is typically unable to extrapolate beyond the range of data used for training (Maier and Dandy, 2000 and references)

There are notable limitations and sources of uncertainty in developing and applying empirical techniques for estimating SSC from visible and NIR satellite data. River sediment properties (e.g., color, mineralogy, grain size distribution) can vary across a region and over time. This can limit the spatial and temporal applicability of empirical SSC-reflectance relationships (Pavelsky et al., 2009). Other reflective suspended or dissolved material (e.g., chlorophyll, carotenoids) may also alter river surface reflectance and therefore the validity of calibrated relationships (Wackerman et al., 2017). Another limitation comes from the penetration depth of satellite sensors for surface reflectance of water (top ~1-2 meters). When the river bottom is shallower than the sensor penetration depth, it will scatter the remote sensing reflectance (Volpe et al., 2011). When the river bottom is deeper than the sensor penetration depth, the SSC measured in the surface layer may significantly differ from the depth-integrated SSC. This latter case is likely to occur at high discharges, when bedload and coarser sediment in the lower water column may be a higher proportion of the total load. Thus, SSC predicted from remote sensing cannot be directly used for depth-integrated SSC analyses and modeling. Furthermore, it is not possible to differentiate if increases in remotely-sensed SSC are resulting from suspended sediment increases in the entire water column or mixing between the lower and upper water columns (Markert et al., 2018).

The temporal extent and frequency of remote sensing imagery can also limit its capacity to monitor SSC (e.g., 8- or 16- day revist interval for Landsat; Sentinel-2 available since 2017). Imagery quality may be limited due to cloud cover. This issue is prevalent in Bangladesh due to its monsoonal hydroclimatology. Hence, it is generally appropriate to rely on remote sensing for monitoring background seasonal SSC rather than isolated events (Wackerman et al., 2017). Seasonal SSC from dry, non-cloudy seasons is more reliable than that from wet, cloudy seasons. The spatial resolution of remote sensing imagery (e.g., 30m for Landsat) can also limit the use of satellite remote sensing for sediment. The stream locations where SSC can be monitored must have river channels wide enough that there are sufficient remote sensing pixels of water that do not mix with the river banks. Narrow channel widths are common for streams with low-orders and steep slopes. These conditions are often found in the uplands of mountainous regions which are typically large sources of sediment.

*Uncertainty varies spatially and temporally. For example, braided portions, such as those in the Jamuna river, have highest uncertainty because of the narrow braided channel widths. The channel bottom depth also significantly varies and significantly impacts surface reflectance and sediment transport properties.*

1. **Recommendations for future work**

*Despite these challenges there is immense opportunity. The Bangladesh river system is perhaps the most complicated in the world when it comes to river sediment. Findings from this work have the potential to give meaningful insights to river systems throughout the world. BROSS is not intended to replace any existing approaches to sediment monitoring, but there is substantial opportunity for it to complement and provide another layer of information that can aid in decision-making.*

* *Integrate additional data from future collection*
* *Develop individual algorithms for each river and*
* *Sediment concentration profile*
* *Analyze how the SSC relates to riverbank erosion*

1. **References**

Beveridge et al. (2020)

Hamed, K. H. (2016) The distribution of Spearman’s rho trend statistic for persistent hydrologic data, Hydrological Sciences Journal, 61:1, 214-223, DOI: 10.1080/02626667.2014.968573

Japanese Meteorological Agency, 2013. Appendix A- Verification Indices. Outline of the Operational Numerical Water Prediction at the Japan Meteorological Agency. <https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2013-nwp/pdf/outline2013_Appendix_A.pdf>