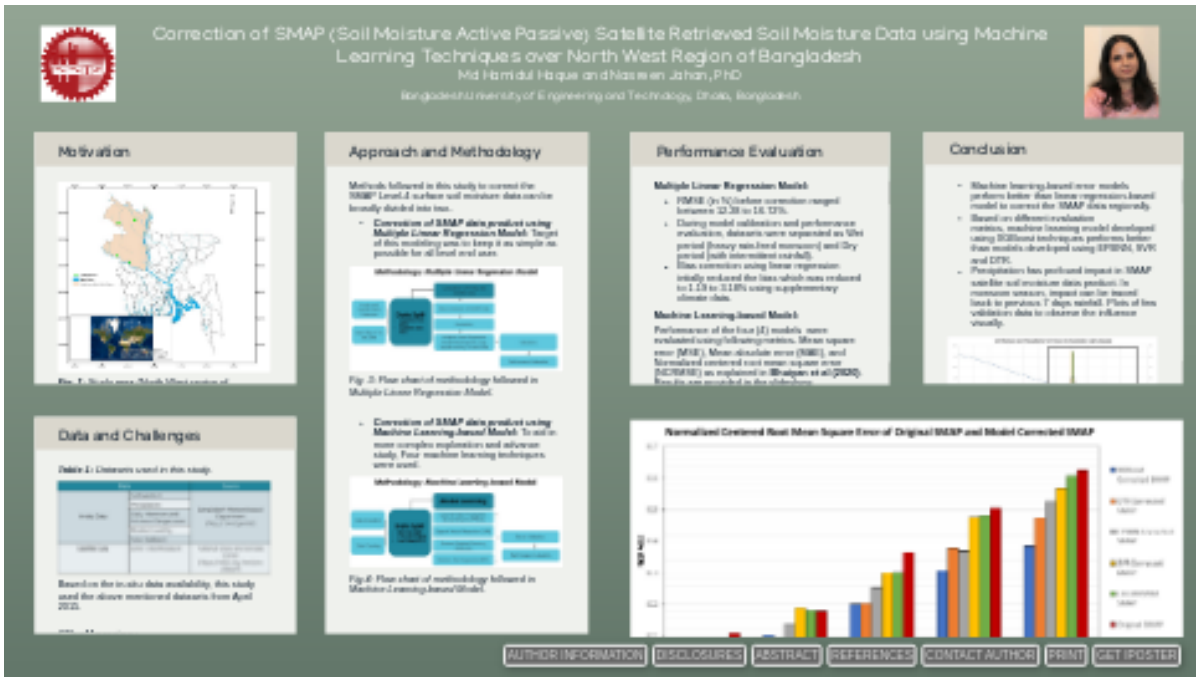
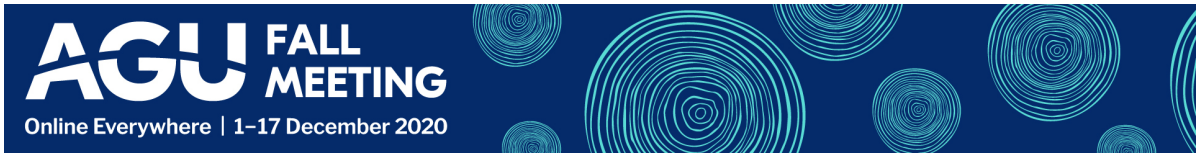
Correction of SMAP (Soil Moisture Active Passive) Satellite Retrieved Soil Moisture Data using Machine Learning Techniques over North West Region of Bangladesh



Md Hamidul Haque and Nasreen Jahan, PhD

Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

PRESENTED AT:



MOTIVATION

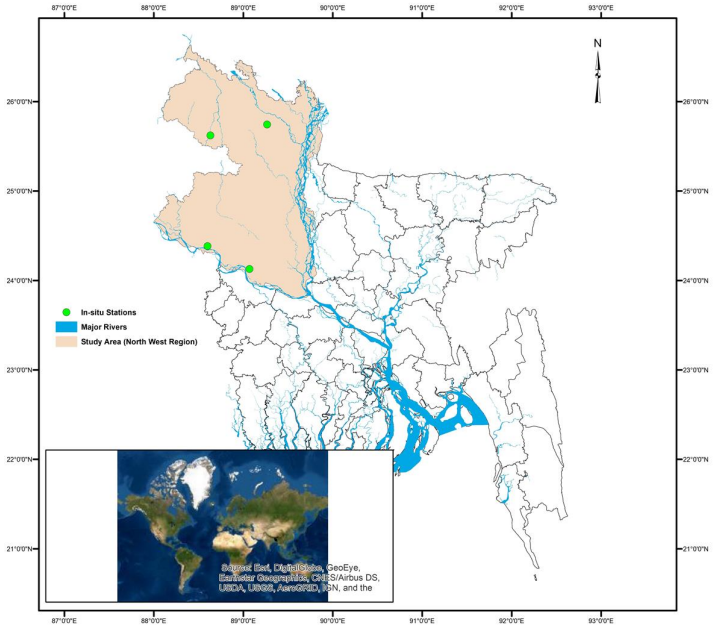


***Fig. 1**: Study area (North-West region of Bangladesh)*

- Bangladesh has only nine agro-meteorological stations that traditionally measure soil moisture four times a month using gravimetric method
- It is insufficient to represent large scale spatial and temporal distribution of soil moisture.
- Therefore, use of soil moisture information in irrigation scheduling, early drought prediction, flood prediction and hydrological modeling is mostly absent.
- North West region of Bangladesh is very vulnerable to different extreme natural events like drought, frequent flood and riverbank erosion.

Goals

- The main objective of this study was to evaluate SMAP (Soil Moisture Active Passive) satellite-based Level-4 surface soil moisture product over the North West Region of Bangladesh and improve it regionally using supplementary climate data.
- Our goal was to promote use of satellite-based soil moisture data for strengthening scientific management of water resources.

Research Questions

Research questions that we tried to solve in this study are:

- How accurately SMAP product provides soil moisture data compared to in-situ observation in the targeted geo-location?
- Does the heavy rain-feed monsoon of the geo-location (Bangladesh) have any effect on the pattern of the bias?
- Can simple error model using supplementary climate data improve the SMAP product further?
- Linear regression to machine learning models: Using simplest to the complex technique to improve the data product.

DATA AND CHALLENGES

Table 1: *Datasets used in this study.*

Data		Source
In-situ Data	Soil Moisture	Bangladesh Meteorological Department (http:// bmd.gov.bd)
	Precipitation	
	Daily Maximum and Minimum Temperature	
	Relative Humidity	
	Solar Radiance	
Satellite Data	Level-4 Soil Moisture	National Snow and Ice Data Center (https://nsidc.org /versions /data/4)

Based on the in-situ data availability, this study used the above mentioned datasets from April 2015.

Challenges:

Bangladesh Meteorological Department officially provides in-situ data 6 to 12 months after the collection with additional time for processing. Therefore, last one and half years data were not used in this study due to unavailability of the data in the study period. As the soil moisture is collected bi-weekly using gravimetric method, repetitive visits were made to their Agro-meteorological division to obtain as much as data possible. In some cases, we were lucky to get additional data from BMD's different time limited projects in our study area.

APPROACH AND METHODOLOGY

Methods followed in this study to correct the SMAP Level-4 surface soil moisture data can be broadly divided into two-

- ***Correction of SMAP data product using Multiple Linear Regression Model:*** Target of this modeling was to keep it as simple as possible for all level end user.

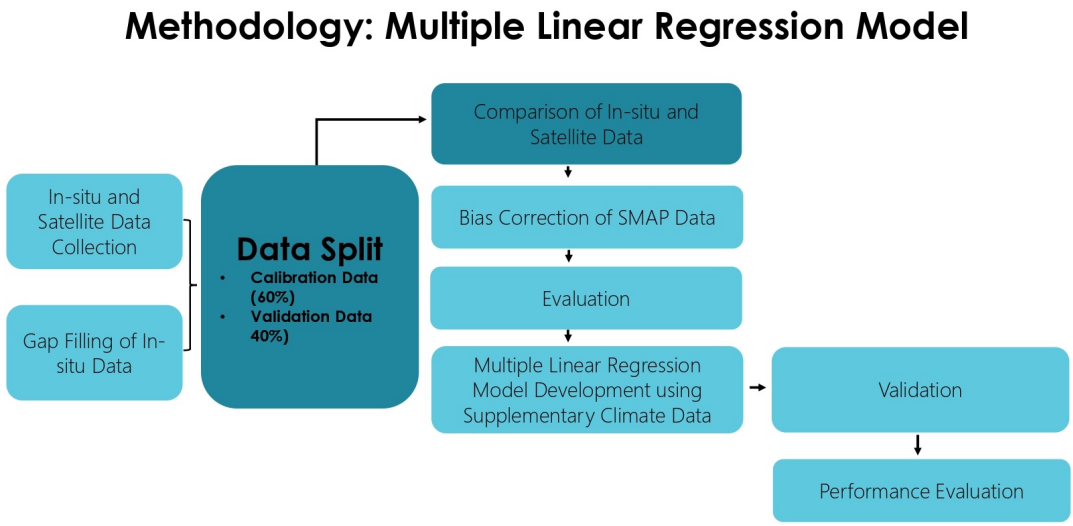


Fig. 3: Flow chart of methodology followed in Multiple Linear Regression Model.

- ***Correction of SMAP data product using Machine Learning-based Model:*** To aid in more complex exploration and advance study, Four machine learning techniques were used.

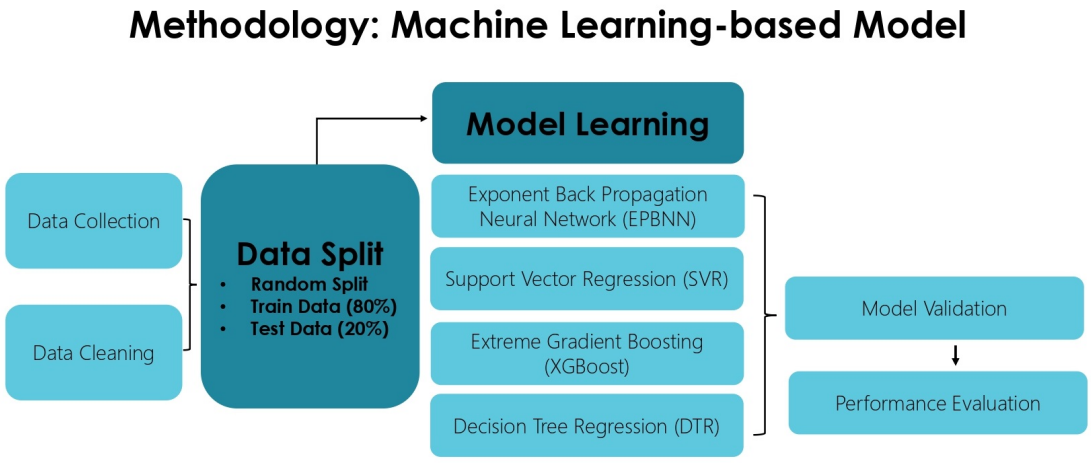


Fig.4: Flow chart of methodology followed in Machine Learning-based Model.

PERFORMANCE EVALUATION

Multiple Linear Regression Model:

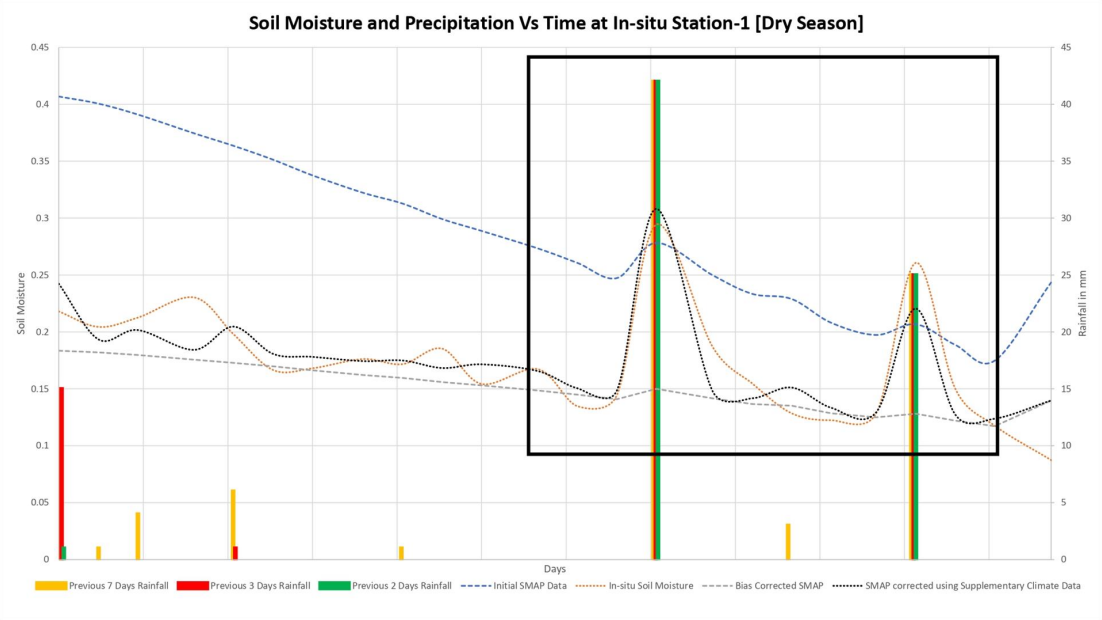
- RMSE (in %) before correction ranged between 12.28 to 16.72%.
- During model calibration and performance evaluation, datasets were separated as Wet period (heavy rain-feed monsoon) and Dry period (with intermittent rainfall).
- Bias correction using linear regression intially reduced the bias which was reduced to 1.19 to 3.18% using supplementary climate data.

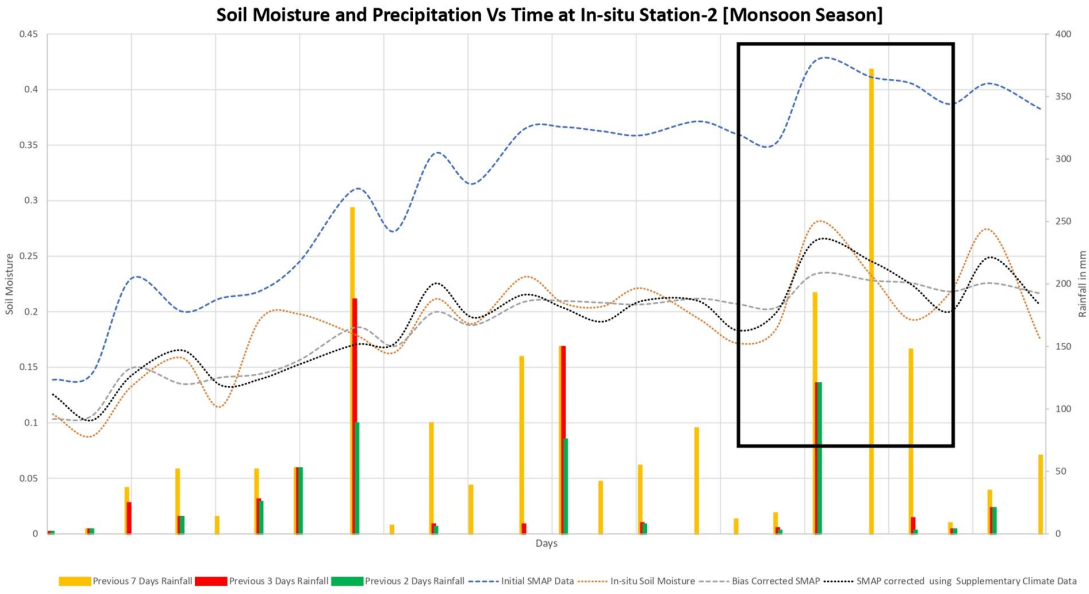
Machine Learning-based Model:

Performance of the four (4) models were evaluated using following metrics- Mean square error (MSE), Mean absolute error (MAE), and Normalized centered root mean square error (NCRMSE) as explained in **Bhuiyan et al (2020)**. Results are provided in the slideshow.

CONCLUSION

- Machine learning-based error models perform better than linear regression-based model to correct the SMAP data regionally. Based on different evaluation metrics, machine learning model developed using XGBoost techniques performs better than models developed using EPBNN, SVR and DTR.
- Machine learning-based error modeling can be used for regional correction of SMAP data.
- Monsson precipitation has profound impact in SMAP satellite soil moisture data product. In monsoon season, impact can be traced back to previous 7 days rainfall. Plots of few validation data to observe the influence visually.





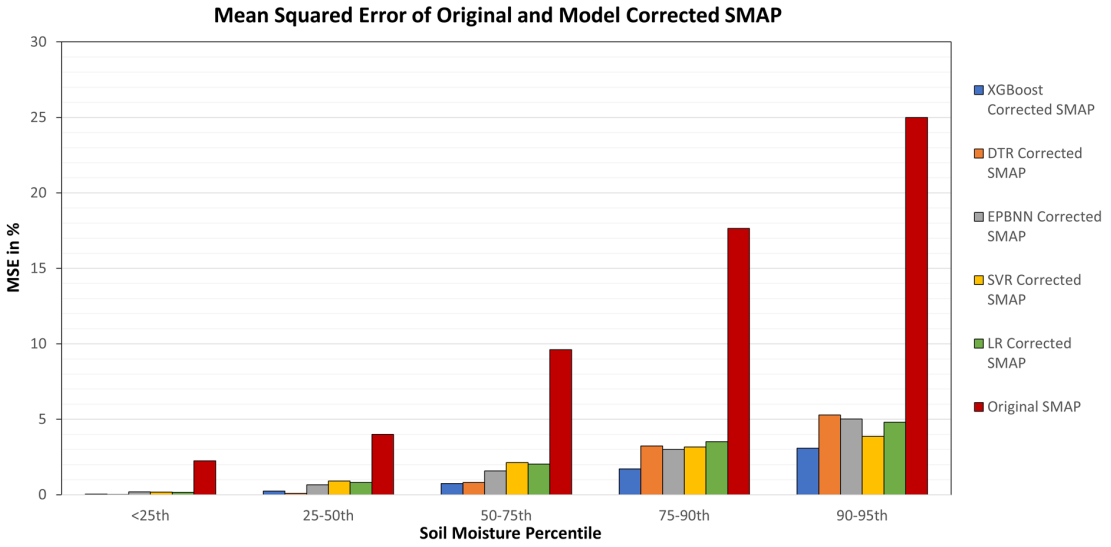
Quick Overview

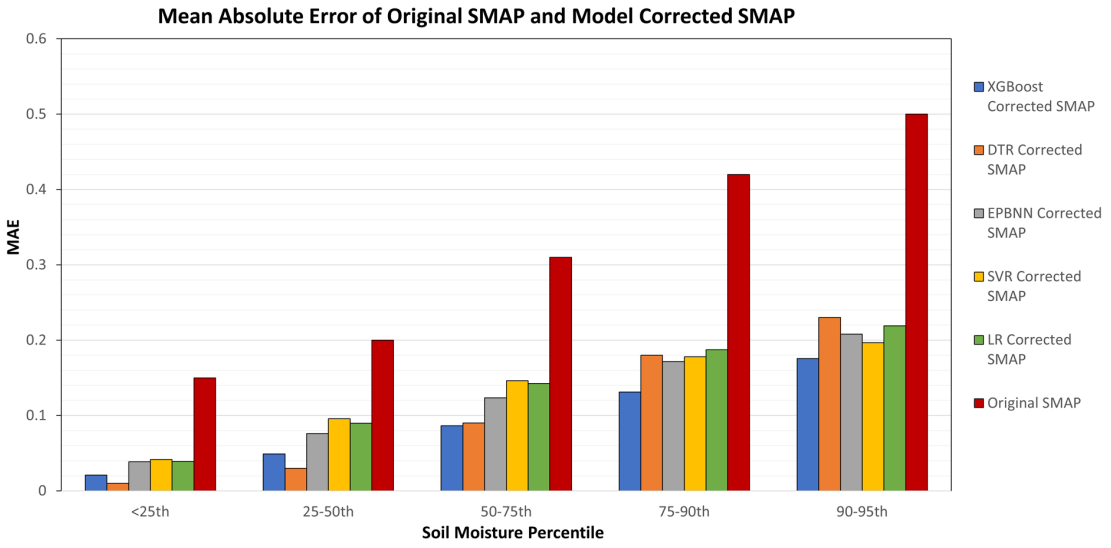
Bangladesh is the 40th largest Economy of the world

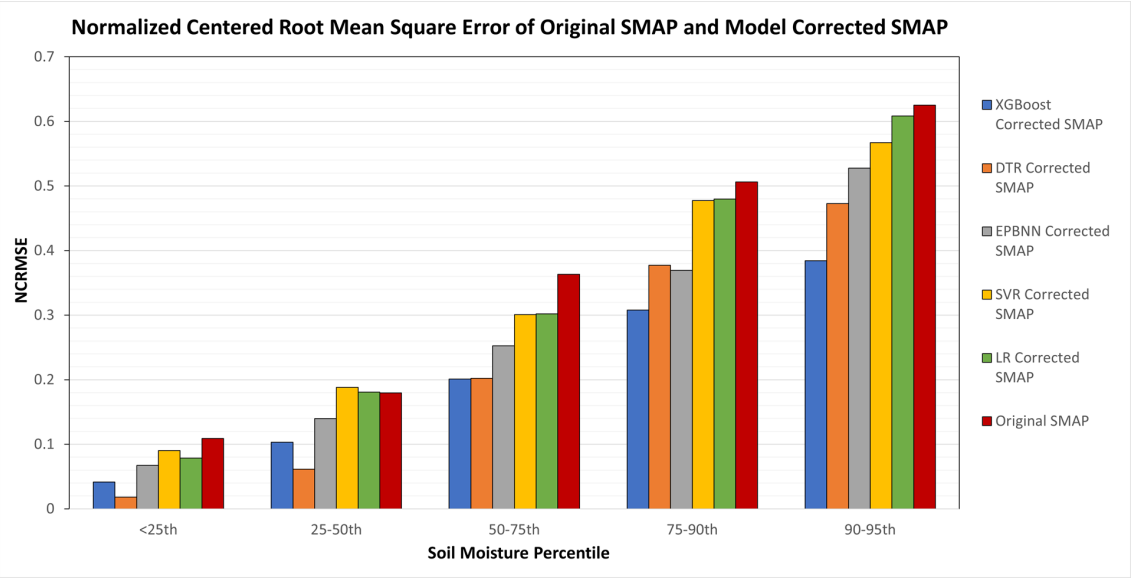
- 14.2% of Total GDP is contributed by **Agriculture** Sector
- 42.7% of Total Labor force works in **Agriculture** Sector
- **Rice** is the main crops
- **Flood** is the most common natural disaster

Source: "CIA – The World Factbook", Central Intelligence Agency and "WELT 2020 | Centre for Economics and Business Research".

Performance Evaluation Metrics






Normalized Centered Root Mean Square Error of Original SMAP and Model Corrected SMAP
NCRMSE
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
<25th
25-50th
50-75th
75-90th
90-95th
Soil Moisture Percentile
XGBoost Corrected SMAP
DTR Corrected SMAP
EPBNN Corrected SMAP
SVR Corrected SMAP
LR Corrected SMAP
Original SMAP

DISCLOSURES

This research received internal undergraduate thesis research funding (~5000 BDT) of BUET for necessary data collection.

Acknowledgement From **Md Hamidul Haque**:

I would like to thank Dr. Mohan Kumar Das, Sr. Scientific Officer, National Oceanographic And Maritime Institute (NOAMI), Dhaka, Bangladesh for aiding in collection of climatic data and helping collect additional soil properties data.

I would like to thank Dr. Jared D Smith of University of Virginia for mentoring me in the last three (3) months through mentoring365 program of AGU which provided me confidence and guidance as the first time participant of AGU Fall meeting.

I would like to thank Dr. Ehsan Bhuiyan of University of Massachusetts, Amherst for his guidance and inspiration.

Lastly, I would like to thank Mushtari Sadia of Bangladesh University of Engineering and Technology (BUET) for her assistance in troubleshooting.

AUTHOR INFORMATION

Md Hamidul Haque

Hamidul is a young water resources engineer with more than three years of professional experience in Hydrology, Hydraulic and Fluvial Geomorphology. He was involved in several consulting projects on hydrologic and hydrodynamic modeling, urban flooding hazard analysis, flood forecasting system development, developing stormwater management model for urban coast, natural channel design, and hazard analysis.

In his undergraduate research thesis, he worked on this current project. His understanding in GIS and remote sensing helped him do several research projects- analyzing 40+ years Landsat imageries of Brahmaputra-Jamuna River of Bangladesh to define braided phenomena across the length, outlining spatial and temporal representation of metal pollution caused by shipbreaking activities in coastal area, assessment of the implementation of the DAP (Detailed Area Plan) using google earth-based land use and land cover analysis. His experience in handling the industry-standard hydrologic and hydrodynamic models has also intrigued him to work on the limitation of the process-based models, and improve it using machine learning techniques. He is currently working with an interdisciplinary research team to experiment on machine learning-based flood forecasting system development using satellite data in Bangladesh-India. Initial results of his study are very notable and accepted to be published in a conference in December 2020. He continuously working to scale it up for the whole basin and improve the techniques further. Hamidul is a recipient of Virtual Berkner Travel Fellowship (2020) to attend the AGU Fall Meeting 2020.

Hamidul can be reached through *hhaque6944@gmail.com* or https://www.linkedin.com/in/hhaque6944 (https://www.linkedin.com/in/hhaque6944).

Nasreen Jahan, PhD

Dr. Nasreen Jahan is currently working as an associate professor at the Department of Water Resources Engineering, Bangladesh Univsersity of Engineering and Technology.

ABSTRACT

Bangladesh, part of Indo-Gangetic-Brahmaputra Plain, is frequently affected by floods and droughts. As the economy of Bangladesh is still agriculture based, effective measurement of soil moisture will not only strengthen the irrigation management but also improve the hydrological modelling and drought prediction. But only nine agro-meteorological stations of Bangladesh measure the soil moisture four times a month which creates a vacuum to scientifically manage her water resources. SMAP (Soil Moisture Active Passive) satellite of NASA provides an unprecedented opportunity for full scale measurement of soil moisture over this region. Field measurements of soil moisture from April 2015 were used to assess the effectiveness of the SMAP's measurement over the North West Region of Bangladesh which suffers from frequent dry spells. Initially the Root mean squared error (RMSE) between the SMAP and observed soil moisture were found to vary between 12.28 to 16.72% for the available stations. The results showed a bias in SMAP data and it was significantly reduced using bias correction. Later multiple linear regression, based on supplementary climate data in addition to SMAP observations, was applied to obtain an improved estimate of soil moisture and the RMSE were reduced to 1.19 to 3.18%. Lastly, different machine learning techniques (i.e. ANN, SVR, XGBoost etc.) were used to reduce the bias further. This study demonstrates a promising potential of using the SMAP data in soil moisture estimation over Bangladesh for its effective water resources management.

REFERENCES

Adamowski, J.; Fung Chan, H.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res. 2012, 48

Bhuiyan, M.A.E.; Yang, F.; Biswas, N.K.; Rahat, S.H.; Neelam, T.J. Machine Learning-Based Error Modeling to Improve GPM IMERG Precipitation Product over the Brahmaputra River Basin. *Forecasting 2020*, 2, 248-266.

Bao, Y.; Xiong, T.; Hu, Z. Multi-step-ahead time series prediction using multiple-output support vector regression. Neurocomputing 2014, 129, 482–493

Chen, Kun-Shan et al. (2016). A Preliminary Evaluation of the SMAP Radiometer Soil Moisture Product over the United States and Europe Using Ground-Based Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, April 2016.

D. Entekhabi et al., “The Soil Moisture Active Passive (SMAP) mission,” in Proc. IEEE, vol. 98, no. 5, pp. 704–716, May 2010.

Haiyan Mo, Jun Wang, Hongli Niu, Exponent Back Propagation Neural Network Forecasting for Financial Cross-Correlation Relationship, Expert Systems With Applications (2016), doi: 10.1016/j.eswa.2015.12.045

Hyndman, Rob J.; Koehler, Anne B. (2006). "Another look at measures of forecast accuracy". *International Journal of Forecasting.* 22 (4): 679688. doi: 10.1016/j.ijforecast.2006.03.001.

Joshua, Willie D. and Rahman, Mostafizur 1983, “Strengthening of Soil Resources and Development Institute”; FAO/UNDP Project BGD/81/023, *Department of Soil Survey*, SRDI, Dhaka.

Koster et al., 2000, “A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure”, *Journal of Geographic Research*, Volume 105, Issue D20, 27 October 2000, Pages 24809–24822

Lobl, E. 2001. “Joint Advanced Microwave Scanning Radiometer (AMSR) Science Team meeting”. *Earth Observer* 13(3): 3-9.

Reichle, R. H., G. De Lannoy, R. D. Koster, W. T. Crow, and J. S. Kimball, 2016: SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 2. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/JJY2V0GJNFRZ.

Ricciardulli, L., Wentz, F.J., April 2016. Remote Sensing Systems ASCAT C-2015 Daily Ocean Vector Winds on 0.25 deg grid, Version 02.1, [indicate subset used]. Santa Rosa, CA: *Remote Sensing Systems*.

Srivastava, K. P., Petropoulos, G.P., Kerr Y.H., 2016 “Satellite Soil Moisture Retrieval, Techniques and Applications”, Elsevier publications.

Zhang, X.; Zhang, T.; Zhou, P.; Shao, Y.; Gao, S. Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. Remote Sens. 2017, 9, 104.