ISSN: 2538-2152; ECOPERSIA 2023;11(3):255-274.

Enhancing Soil Moisture Estimation: Exploring the Synergy of Optical Trapezoid and Deep Learning Models

ARTICLEINFO

Article Type
Original Research

Author
Golnaz Zuravand, *M.Sc.*[1]
Vahid Moosavi, *Ph.D.*[2*]
Seyed Rashid Fallah Shamsi, *Ph.D.*[3]

How to cite this article
Zuravand G., Moosavi V., Fallah Shamsi SR. TEnhancing Soil Moisture Estimation: Exploring the Synergy of Optical Trapezoid and Deep Learning Models. ECOPERSIA 2023;11(3): 255-274

DOR:
20.1001.1.23222700.2023.11.3.7.1

[1] M.Sc. student, Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran.
[2] Assistant Professor, Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran.
[3] Ph.D., Associate Professor, Department of Natural Resource and Environment, Shiraz University, Iran and Visiting Scientist, Institute of Geographical Science and Natural Resources Research-Chinese Academy of Sciences, Beijing, P.R.China

**** Correspondence***
Address: Mazandaran Province, Noor, Emam Khomeini Street, Faculty of Natural Resources, Tarbiat Modares University.
P. O. Box: 8961719311
Phone: +98 9156460501
Fax: +98 1144999121
Email: v.moosavi@modares.ac.ir

Article History
Received: August 3, 2023
Accepted: September 3, 2023
Published: September 20, 2023

ABSTRACT

Aims: This study aimed to propose an effective model for estimating soil moisture by integrating the optical trapezoid method with a deep learning Long Short-Term Memory (LSTM) model. The performance of the proposed model was compared with two other methods, i.e., Partial Least Squares (PLS) regression and Group Method of Data Handling (GMDH) multivariate neural network.

Materials & Methods: This study combined the optical trapezoid method with deep learning models to propose an effective model for soil moisture estimation in the Maragheh watershed. A total of 499 in-situ soil moisture data were collected. Relative moisture content was calculated using the optical trapezoid method and imported into the LSTM model, along with other inputs such as spectral indices and DEM-based derived variables. The performance of the mentioned models was assessed both with and without the optical trapezoid method to evaluate its efficacy on the performance of AI models.

Findings: The results demonstrate that the combined model of deep learning LSTM and the optical trapezoid method achieves satisfactory performance, with an R^2 of 0.95 and a RMSE of 1.7%. The PLS and GMDH methods performed moderately, both without the involvement of the optical trapezoid method and in the combined mode.

Conclusion: This study shows that the optical trapezoid method can improve the performance of deep-learning models in estimating soil moisture. However, considering the significant difference in computational costs among these models, choosing the appropriate model depends on the user's objectives and desired level of accuracy and precision.

Keywords: Deep Learning; Optical Trapezoid Method; Remote Sensing; Sentinel-2; Soil Moisture.

CITATION LINKS

[1] Rouse J.W., Haas R.H., Schell J.A., Deering D.W. ... [2] Fathalolomi S., Vaezi E.R., Alavi Panah K., Ghorbani A. ... [3] Koohbanani H., Yazdani R. Mapping the moisture of surface soil using Landsat 8 imagery (... [4] Tabatabaeenejad A., Burgin M., Duan X., Moghaddam M. P-band radar retrieval of subsurfac ... [5] Sanli F.B., Kurucu Y., Esetlili M.T., Abdikana S. Soil moisture ... [6] Srivastava H.S., Patel P., Sharma Y., Navalgund R.R. Large-area soil moisture estimation ... [7] Silva B.M., Silva S.H.G., Oliveira G.C.d., Peters PHCR, Santos W.J.R.d., Curi N. Soil mo ... [8] Javadi P., Asadi H., Vazife M. Estimation of spatial changes of soil moisture using rand ... [9] Foroughi H., Naseri A.A., Boroomandnasab S., Hamzeh S., Jones S.B. Presenting a new meth ... [10] Shokri S.h., Farrokhian Firoozi A., Babaeian E. Estimation of soil moisture by combining ... [11] Sadeghi M., Babaeian E., Tuller M., Jones S.B. The optical trapezoid model A novel appro ... [12] Sedaghat A., Shabanpoor M., Norozi E.A., Fallah E., Bayat ... [13] Sedaghat A., Shabanpoor M., Norozi E.A., Fallah E., Bayat H. Modeling soil surface moist ... [14] Norozi Aghdam A., Behbahani M.R., Rahimi Khoob E., Aghighi H. Moisture model of the soil ... [15] Norozi Aghdam A., Karami V. Application of remote sensing technology in monitoring and e ... [16] Bagheri K., Bagheri M., Hoseynzade E.A., Parvin M. Estimation of soil moisture using opt ... [17] Zeyliger A.M., Muzalevskiy K.V., Zinchenko E.V., Ermolaeva ... [18] Ramezani Charmhine E.A., Zonemat Kermani M. Investigating the effectiveness of multi-lay ... [19] Avazpoor S., Bakhtiari B., Ghaderi K. ... [20] Golkar E., Ahmadi M.M., Qaderi K., Rahimpour M. Peak velocity of pollutant transport pre ... [21] Godarzi M.R., Godarzi H. Investigating the effectiveness of data group classification me ... [22] Moosavi V., Talebi A., Hadian MR Development of a hybrid wavelet packet forecasting grou ... [23] Adalat MH, Azmi R., Bagherinejad J. An enhanced LSTM. Method to improve the accuracy of ... [24] Jordan M. A. Parallel distributed ... [25] Yu M., Xu F., Xu W., Sun J., Cervone G. Using long-term memory (LSTM) and internet of th ... [26] Liu Z.H., Meng X.D., Wei H.L., Chen L, Lu B.L., Wang Z.H., ... [27] Haddadian H., Karimi E., Esfandiar poor E., Hagh nia Gh. RMSE and NRMSE to calculate the ... [28] Ahmadinezhad Baghban F., Moosavi V. ... [29] Zhou Y., Zhang Y., Wang R., Chen H., Zhao Q., Liu B., Shao Q., Cao L., Sun S. Deep lear ... [30] Yinglan A., Wang G., Hu P., Lai P., Xue B., Fang Q. ... [31] Achieng K.O. Modelling of soil moisture ... [32] Joshi R.C., Ryu D., Lane P.N.J., Sheridan G.J. Seasonal forecast of soil moisture over M ...

Introduction

Soil moisture is a critical variable in hydrological and climatic models, playing a significant role in drought studies [1]. Measuring, modeling, and monitoring spatiotemporal variations in soil moisture are of utmost importance for scientific and practical studies [2]. Accurate determination of the influential factors on soil moisture's spatial and temporal distribution is crucial. Monitoring the spatiotemporal distribution of soil moisture is an effective measure in water and soil management [3], necessitating the generation of high-quality soil moisture maps with precision and accuracy to enhance understanding of regional water and climate conditions [4]. Soil moisture content exhibits substantial variability across different times and locations [5]. Various methods are employed for soil moisture measurement, including direct and indirect approaches. Field observations offer high-quality soil moisture mapping in small areas; however, their large-scale implementation is costly and time-consuming due to limited weather stations, particularly in mountainous and inaccessible regions [6,7,8].

The triangle or trapezoidal method is widely used as an indirect soil moisture modeling and estimation approach. This method relies on the relationship between vegetation cover index and land surface temperature [9,10]. Numerous studies have demonstrated the high capability of the trapezoidal method in estimating soil moisture [11]. However, a significant challenge arises when this method is applied to satellite sensors lacking a thermal band, such as Sentinel, as they cannot calculate land surface temperature. To address this issue, Sadeghi et al. [11] proposed the optical trapezoidal method, which utilizes the shortwave infrared band as a substitute for land surface temperature in calculating relative soil saturation. Mathematical models also have a high application in simulating soil moisture [12]. Artificial intelligence models are highly capable of establishing the relationship between independent and dependent variables. Researchers have reported that Machine learning models perform well in predicting and mapping soil properties. Sedaghat et al. [13] and Norozi Aghdam et al. [14] studied the relationship between soil moisture field measurements and variables such as weather data, delayed NDVI, and MODIS satellite imagery. These researches showed a good correlation between surface soil moisture and 15-day delayed NDVI during the growing season. Norozi Aghdam et al., [15] used Landsat 8 and Sentinel-1 satellite data to calculate various regression relationships between indicators and ground points. After comparing the results, they introduced two multivariate regression models for estimating soil moisture, which showed satisfactory performance. Bagheri et al. [16] presented an algorithm for estimating surface soil moisture using Sentinel-2 images. This study investigated the relationship between soil spectral indices and surface soil moisture using machine learning techniques. The estimation of volumetric soil moisture using the random forest method had higher accuracy than the regression method. It was shown that spectral indices are good predictors for soil moisture estimation. Sadeghi et al. [11] proposed a new optical triangle model to overcome the inherent limitations of triangle-based models, considering the absence of thermal bands in images such as Sentinel. It operates based on the linear physical relationship between soil moisture and reflectance in the shortwave infrared wavelength. They estimated surface soil moisture using this Walnut Gulch and Little Washita watersheds model and compared it with observed data. The results indicated the satisfactory performance of this model in estimating surface soil moisture. Foroughi et al. [9] proposed a novel method for esti-

mating soil moisture based on a new definition of soil moisture isopleths. The proposed model was compared with several common models using LANDSAT-8 satellite images in sugarcane fields in Khuzestan province. For validation, soil moisture was measured at 22 points and 5 depths. The results showed that the proposed model had better agreement with field observations and improved accuracy and precision. In Sedaghat et al. [12] investigation, soil surface moisture was estimated using soil variables and spectral indices derived from the Sentinel-2 satellite sensor. Two distinct methods were employed to ascertain their performance, namely Artificial Neural Network (ANN) and Support Vector Machine (SVM) regression. The findings indicated that ANN outperformed SVM in terms of accuracy. Moreover, the soil's color index (CI) demonstrated superior efficacy compared to other spectral indices when employed in conjunction with the ANN method for estimating soil moisture. Shokri et al. [10] investigated various models derived from combining remote sensing variables and soil physical properties to estimate soil moisture in agricultural fields and industrial areas in Amir Kabir, Khuzestan. The outcomes revealed that integrating these variables substantially enhanced the accuracy of soil moisture estimation. Furthermore, the proposed models facilitated the estimation of spatial and temporal variations in soil moisture. Zeyliger et al. [17] measured surface soil moisture near Vodnyy village in the Volgard region of Russia and estimated soil surface moisture using Sentinel-1 data. Their findings highlighted the potential of artificial intelligence methods and Sentinel-1 imagery in soil moisture estimation.While several studies have explored soil moisture modeling using remote sensing and artificial intelligence methods, few have focused on combining deep learning techniques with optical sensors. To address this gap, this study aims to use the ability of optical trapezoid and deep learning LSTM models synergically in soil moisture modeling in the Maragheh Watershed. The performance of the combined model was also compared with those of the PLS and GMDH models. Considering the differences in computational costs among these methods, evaluating their performance can provide valuable insights for soil moisture studies using remote sensing and modeling techniques.

Materials & Methods

Study Area

This research was conducted in the Maragheh watershed in the East Azerbaijan province. The watershed covers an area of 1,100 square kilometers and encompasses the major cities of Maragheh and Bonab. Its highest elevation is 3,696 m, while the lowest point measures 1,480 m. The average minimum annual temperature is 7.82 degrees Celsius, and the maximum reaches 18.53 degrees Celsius. The region receives an annual precipitation of 340mm. Figure (1) visually represents the study area's location within the country and the province. Various land uses/covers, such as agriculture, rangeland, and barren lands, are present in the area.

Data Collection

In order to synchronize the collection of in-situ soil moisture data with satellite passes, a portable TDR (Time Domain Reflectometry) device was used to gather data at 499 specific points with 8cm depth. The soil type is predominantly clay loam. Due to logistical constraints, it was only feasible to collect some 499 data samples in a single day. Therefore, the data collection process was spread over four days, coinciding with the imaging schedule of Sentinel 2 sensors. This approach ensured that the in-situ measurements aligned with the satellite observations, enabling accurate comparisons and analysis of soil moisture levels. After collect-

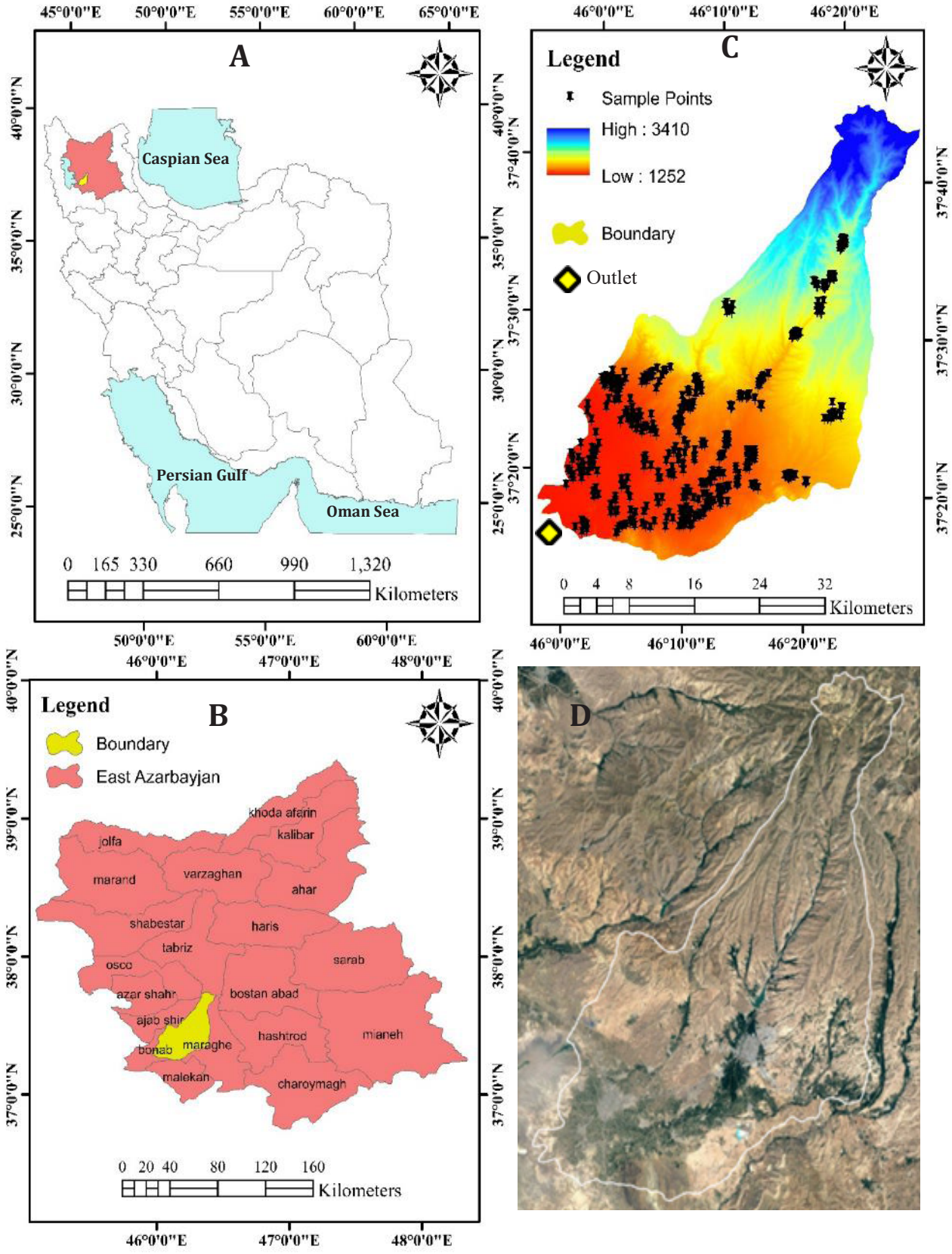

Figure 1) Study area, a) Iran map, b) East Azarbayjan, c) in-situ SM sampling locations, d) satellite map of the study area.

ing in-situ data, the collected data for each day were overlaid onto the corresponding database explicitly provided for that particular day. An imaginary square with an area of 100 square meters was used to record 5 SM values for each sample. Out of these 5 values, 4 were taken from the corners of the square, while the remaining one was obtained from the center, and the average of these 5 points was considered as the SM value for that sample. Sentinel data were collected on four different dates: June 12, 2021; June 17, 2021; June 22, 2021; June 27, 2021; and July 7, 2021. These dates aligned with the collection of in-situ data samples.

Research Methodology

This study utilized Sentinel-2 satellite images obtained from the Copernicus database to model soil moisture. The images were downloaded on specific sampling dates and underwent necessary preprocessing. During this stage, the NDVI and NDWI indices were calculated from the images, along with the TWI, slope, and aspect layers derived from a digital elevation model with a spatial resolution of 10 meters. The optical trapezoid method and artificial intelligence models were employed to estimate soil moisture. Initially, the STR index was computed based on the NDVI index and the SWIR band, which had been converted to reflectance during preprocessing (Equation 1). STR stands for shortwave infrared transformed reflectance. It refers to the reflectance values of an object or surface that have been transformed or processed using the Shortwave Infrared (SWIR) spectrum of light. Based on the linear relationship between soil moisture content and vegetation, the STR-NDVI space formed a trapezoid shape (Figure 2). By plotting dry and wet edge lines using Equations 2 and 3, the relative saturation of the soil was determined using Equation 4.

where and represent the intercept and slope of the dry edge, and represent the wet edge intercept and slope, and W represents the relative saturation of the soil. Finally, all the created layers were stacked together to form an initial database, including the multi-spectral bands of Sentinel-2 and other input variables.

$$STR = \frac{(1-R_{SWIR})^2}{2R_{SWIR}} \quad \text{Eq. (1)}$$

$$STR_d = i_d + s_d NDVI \quad \text{Eq. (2)}$$

$$STR_w = i_w + s_w NDVI \quad \text{Eq. (3)}$$

$$W = \frac{\theta-\theta_d}{\theta_w-\theta_d} = \frac{STR-STR_d}{STR_w-STR_d} \quad \text{Eq. (4)}$$

Figure 1 (c) displays the spatial distribution of the sampling points. Despite facing challenges such as difficult access and limited availability, particularly in the upstream regions, diligent efforts were made to ensure an appropriate distribution of sample collection sites. Subsequently, the point vector layer representing the sampling points for each day was superimposed onto the Sentinel-2 satellite images captured on the same day. The digital number (DN) values, derived index values, and relative moisture content of these locations were extracted from the database above. These extracted datasets were then compiled and utilized as input data for the models, while the soil moisture data served as the output variable. Subsequently, the data from remotely sensed layers were extracted for every ground sampling point, and a final database was prepared containing the input variables along with the corresponding observed soil moisture for the 499 points.

The study involved three models for soil moisture estimation. The available data were divided into training and testing sets, with a 70/30 ratio. The same data was used for all models. The training data was used to calibrate the models and optimize their pa-

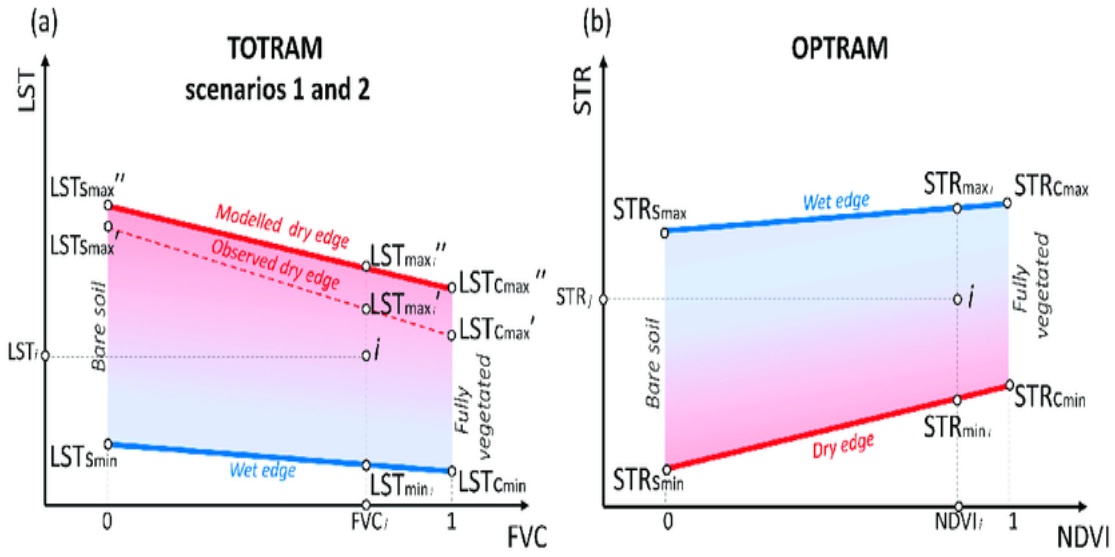

Figure 2) Schematic illustration of the trapezoidal method [11].

rameters, while the testing data was used to evaluate their performance. Statistical measures such as RMSE, R^2, and NRMSE were used to assess the accuracy of the soil moisture estimation. Since the testing data was not used in the training process, it provided a suitable basis for comparing the models' performance. Therefore, only the results from the testing phase were reported and used for comparison purposes. To evaluate the efficacy of using the trapezoidal method with AI models in estimating soil moisture, two scenarios were compared: one including the relative soil saturation obtained from the trapezoidal method as an input variable and another without this inclusion. The research employed three methods, namely PLS regression, GMDH, and LSTM deep learning model, to model and map soil moisture. Partial Least Squares regression is a widely used data-driven model and statistical technique for investigating relationships between variables. It establishes a relationship between a dependent variable and one or more independent variables, generating a fitting function and corresponding equation for analysis. Partial Least Squares regression is particularly effective in handling multicollinearity issues. This method applies the least squares solution to a set of orthogonal components, which are iterative linear combinations of the independent variables. The iterations aim to maximize the covariance between the transformed independent variables and the dependent variables [18, 19].

Another method used in this study is the GMDH model, an artificial neural network that incorporates self-organizing principles inspired by the human mind. This model excels in addressing complex problems with multiple dimensions [20]. When constructing complex models for intricate regression systems, the GMDH algorithm is a heuristic method that offers advantages over traditional modeling approaches [21, 22]. Following is the GMDH algorithm used for soil moisture estimation:

1- The GMDH algorithm is initialized by setting the maximum number of layers and nodes per layer.

2- Layer formation begins with a single layer consisting of all input features. The output of each node in this layer is calculated using an

appropriate activation function.
3- The best-performing models from the previous layer are selected based on an evaluation criterion such as the least error or highest correlation coefficient. These models become candidates for the next layer.
4- Layer augmentation involves forming new layers by adding one additional node at a time to the previous layer's models. Each new node is created by combining the selected models from the previous layer in various ways, such as linear or polynomial combinations.
5- The error of each candidate model in the new layer is then calculated using a suitable error metric like Mean Squared Error.
6- Again, the best-performing models from the current layer are selected based on the error criteria. These models become candidates for the next layer.
7- Steps 4 to 6 are repeated iteratively until the desired number of layers or predefined stopping criteria, such as a minimum improvement in error, is met.
This study employed LSTM, a powerful deep-learning technique, for modeling soil moisture. LSTM is a specific type of recurrent neural network (RNN) that connects its cell outputs and previous layers, allowing information to flow back into itself. This recurrent loop enables the network to use previously obtained information for subsequent computations [23]. However, one challenge faced by RNNs is their limited ability to capture long-term dependencies. To overcome this limitation, LSTM was introduced by Hochreiter and Schmidhuber in 1997 [24]. LSTM solves the problem of vanishing or exploding gradients often encountered with RNNs by regulating the hidden state of the LSTM through input and output gates [25]. In this approach, new data can be added to the cell state through the input, while the output gate controls the output data of the cell, and the temporary storage manages the information stored within the cell state [26].
Following is the algorithm of LSTM used in this study to estimate soil moisture.
1- Initialization: the LSTM parameters including input weights, forget weights, output weights, and cell state weights. Initialize bias terms for the input gate, forget gate, output gate, and cell state were Initialized.
2- Input Processing: For each time step t and input sequence of soil moisture values, the input was passed through the LSTM network.
3- LSTM Cell Operations:
1.a. The input gate activation was calculated by applying a sigmoid function to the sum of the weighted inputs, biases, and soil moisture values.
2.b. The forget gate activation was calculated by applying a sigmoid function to the sum of the weighted inputs, biases, and soil moisture values.
3.c. The output gate activation was calculated by applying a sigmoid function to the sum of the weighted inputs, biases, and soil moisture values.
4.d. Candidate cell state was calculated by applying a hyperbolic tangent function to the sum of the weighted inputs, biases, and soil moisture values.
4- Update Cell State: The cell state was updated by multiplying the forget gate activation with the previous cell state and adding the product of the input gate activation and the candidate cell state.
5- Hidden State Calculation: The hidden state was calculated by applying a hyperbolic tangent function to the updated cell state multiplied by the output gate activation.
6- Output: The final hidden state was considered for estimating the soil moisture for the next time step or as input for subsequent layers.
7- Backpropagation: The error was backpropagated through time by computing the gradients of the loss concerning the parameters, and they were updated using an opti-

mization algorithm like gradient descent or Adam.

8- Training Loop: Steps 2-7 were repeated for multiple epochs until the model converged, adjusting the weights and biases to minimize the difference between the predicted soil moisture values and the actual soil moisture measurements.

Evaluation of Model Performance

Two sets of experiments were conducted to assess the effectiveness of combining the trapezoid model with AI models for soil moisture estimation. The first set involved running the models without incorporating the relative saturation of the soil obtained from the trapezoidal method, while the second set included this parameter. A comparison was made between the results obtained from these two sets. Following the estimation of soil moisture values, statistical metrics such as Root Mean Square Error (RMSE), Normalized RMSE (NRMSE), and R-squared (R^2) were employed to evaluate the performance of the models, as outlined in equations 5 to 7. NRMSE, similar to RMSE, provides a measure of the model's effective error. However, it normalizes this measure by considering the range of numbers used in the standardized model [27].

$$RMSE = \sqrt{Mean(o_i - e_i)^2} \quad \text{Eq. (5)}$$

$$NRMSE = \frac{RMSE}{Mean(o)} \quad \text{Eq. (6)}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(o_i - \bar{o})(e_i - \bar{e})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}\sqrt{\sum_{i=1}^{n}(e_i - \bar{e})^2}} \right)^2 \quad \text{Eq. (7)}$$

where e represents estimated, and o represents observed values.

Findings

Figure 3 presents the results obtained from soil moisture modeling through three different approaches, i.e., PLS, GMDH, and LSTM deep learning model. The results are based on test data analysis, comprising 30% of the overall dataset that was not utilized during the training phase. A closer examination of the figure reveals that the PLS neural network model (part A of the graph) exhibited unsatisfactory performance in predicting soil moisture. This is indicated by the NRMSR and RMSE coefficients, which were approximately 30% and 3.3%, respectively, suggesting the poor accuracy of this model. The simple structure of the PLS neural network could not comprehend and uncover intricate relationships between the independent and dependent variables. As can be observed in part III of section A in the graph, the predictions made by this model varied from an underestimation of over 6% to an overestimation of more than 9%.

Furthermore, it also needed to estimate extreme values (high soil moisture) accurately. The scatter plot in this model indicates a coefficient of determination of approximately 84%, which corresponds to a correlation of about 70%. The results in Section B of the figure demonstrate that the GMDH model outperformed the PLS model. This was evident from the lower NRMSR and RMSE coefficients, which were approximately 25% and 2.8% respectively. Despite minor improvements, the error range remained similar between the two models. It is worth noting that the GMDH model had a more complex structure and incurred higher computational costs than the PLS model. However, contrary to expectations, the GMDH model needed to exhibit superiority over the PLS model in adequately identifying the relationships between variables in the soil moisture process. However, section C of Figure 3 shows the results of the LSTM deep learning model without the intervention of the hysteresis method. These results indicate better performance of this model compared to the previous two models. The NRMSR and RMSE coefficients

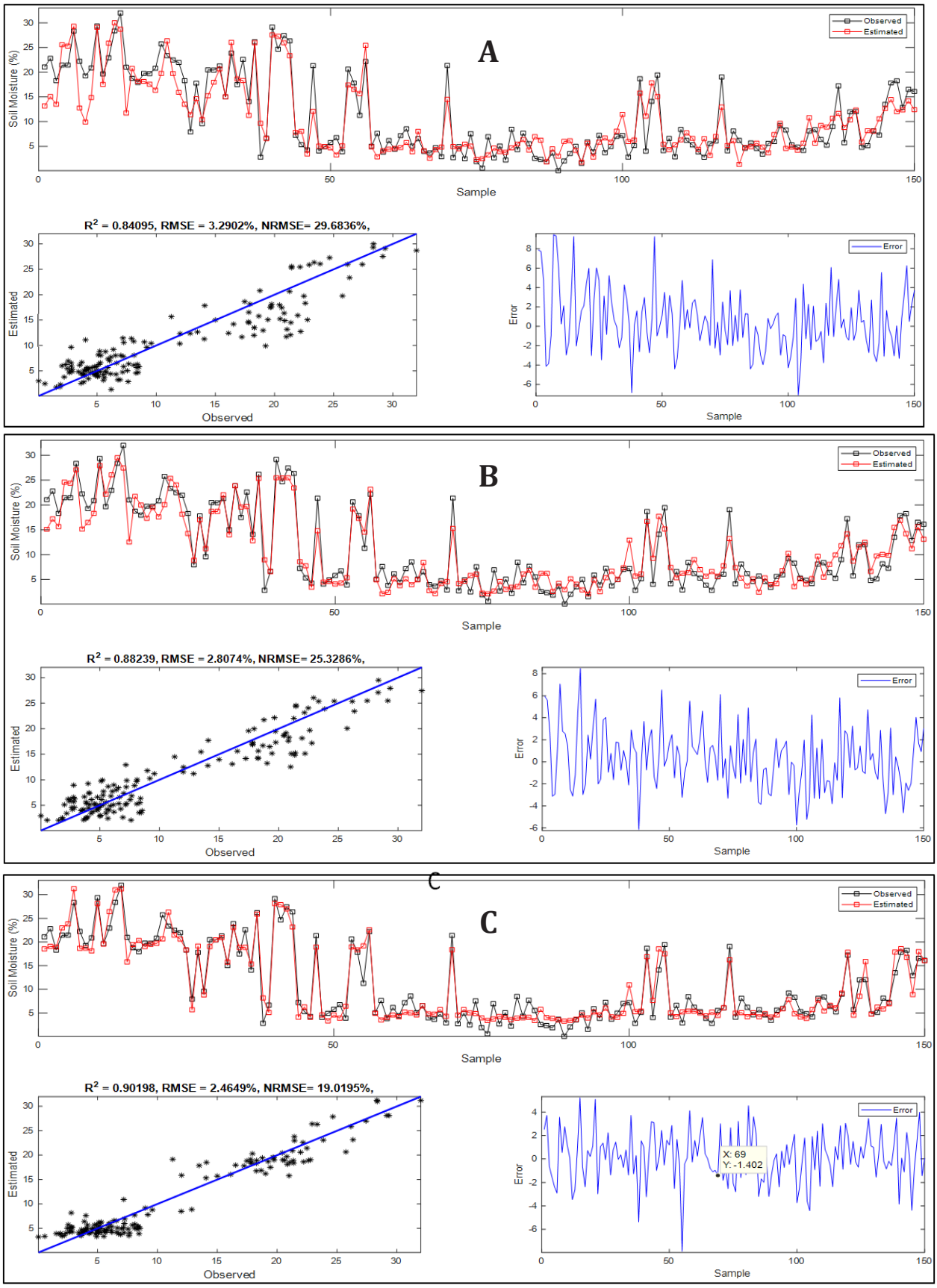

Figure 3) The results of soil moisture modeling using A) PLS, B) GMDH, and C) LSTM.

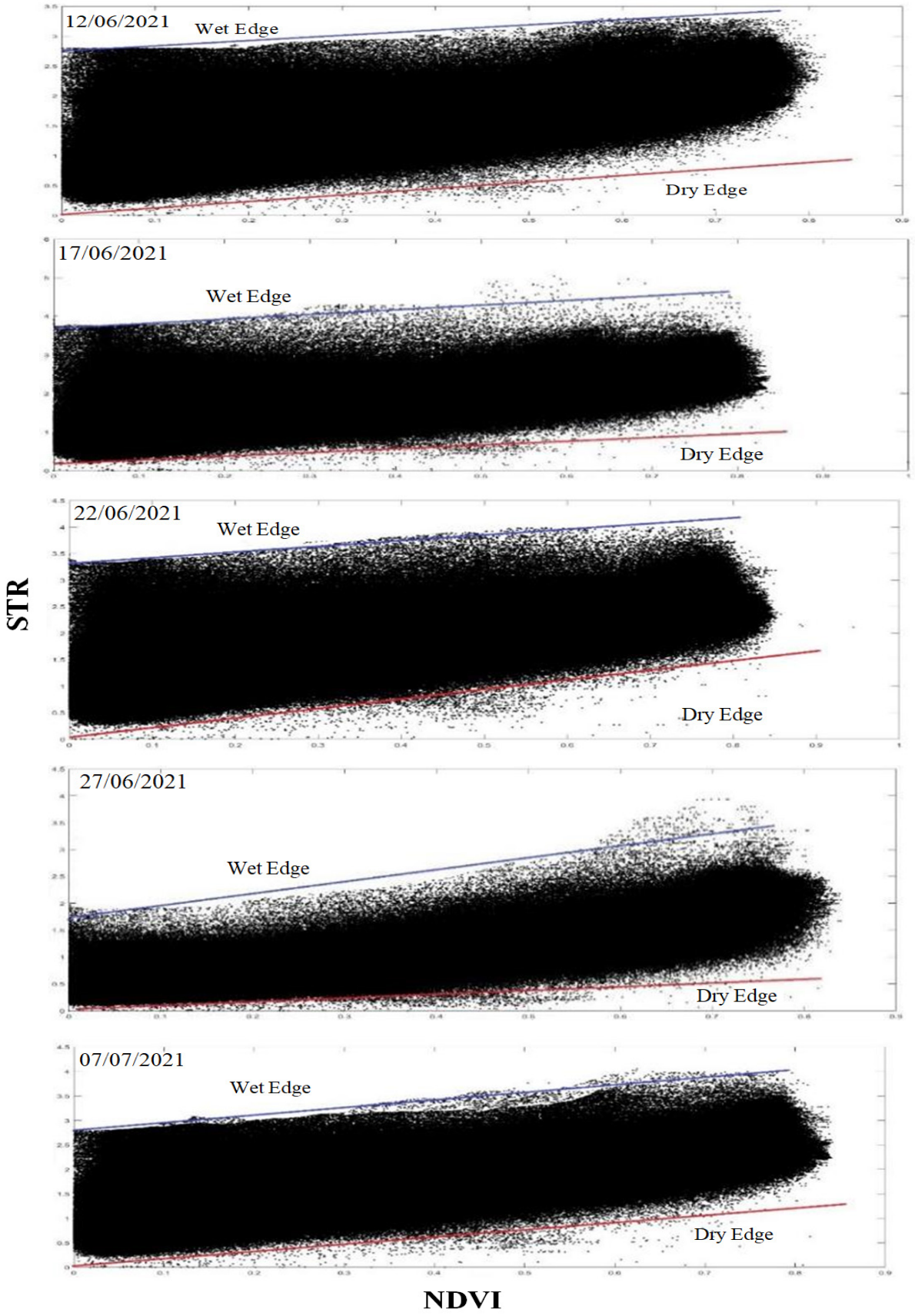

Figure 4) Trapezoidal space obtained from Sentinel-2 image.

for this model were approximately 19% and 2.4%, respectively. This performance is suitable and acceptable for estimating soil moisture, although it still has considerable uncertainties. The error range of the model has also significantly decreased compared to the other two models, and except for limited cases, it fluctuates around zero with changes between -5 and +5. The coefficient of determination of 90% in this model also supports its relatively good performance. The PLS model required approximately 1.5 hours for calibration, while the calibration process took approximately 10 hours for the GMDH model and 18 hours for the LSTM model. One possible reason for the significant difference in calibration times is the complexity of the models. PLS, a simpler model, requires less time to calibrate due to its reduced number of parameters and computations involved. On the other hand, both GMDH and LSTM models are more intricate and involve training a more significant number of parameters, which leads to a longer calibration time. The training algorithms employed by each model can also impact the calibration time. PLS typically uses simpler and faster optimization techniques, resulting in a quicker calibration. On the contrary, GMDH and LSTM models often require more computationally intensive and time-consuming training algorithms, leading to longer calibration times. Overall, the calibration time variation can be attributed to the complexity of the models and the training algorithms utilized, with simpler models and faster optimization techniques resulting in shorter calibration times.

Figure 4 represents a trapezoidal space with dry and wet edges obtained through the trapezoidal method. The relative saturation level was calculated after plotting this trapezoidal space and calculating the equations of the dry and wet edges. This value was then added as a new layer to the existing database and included as a new variable in the models. In the next step, the desired models were executed by considering the new variable as input alongside other variables used in the previous stage, and soil moisture values were estimated.

Figure 5 illustrates the results of soil moisture modeling using three methods, i.e., PLS, GMDH, and LSTM deep learning model, in conjunction with the optical trapezoid method for test data (30% of the data not used in the training phase). As observed in this figure, the performance of the PLS model did not significantly improve. The values of RMSE and NRMSR coefficients have changed from approximately 3.3% and 30% to 3.1% and 27.9%, respectively, which is not a considerable improvement. However, it should be noted that the range of error value variations has decreased to some extent. In this combined model, the error values range between -4 and +4, while in the simple model, they range from -9 to +6. The performance of the GMDH model has also improved to some extent. The NRMSR and RMSE coefficients have decreased from approximately 25% and 2.8% to around 21% and 2.4%, respectively. Although this improvement is more significant than the PLS method, it still needs to be significant. Nevertheless, the coefficient of determination (R-squared) has increased from 0.88 in the absence of the trapezoidal method to 0.92 in the combined approach. Part C of Figure 5 shows the results of soil moisture modeling using the combined optical trapezoid-deep learning method. The values of NRMSR and RMSE coefficients have decreased from 19% and 2.4% in the absence of the trapezoidal method to 15.4% and 1.7% in the combined approach, respectively. The range of error values is between -4 and +4, and most error values are centered around zero. The coefficient of determination has also reached a satisfac-

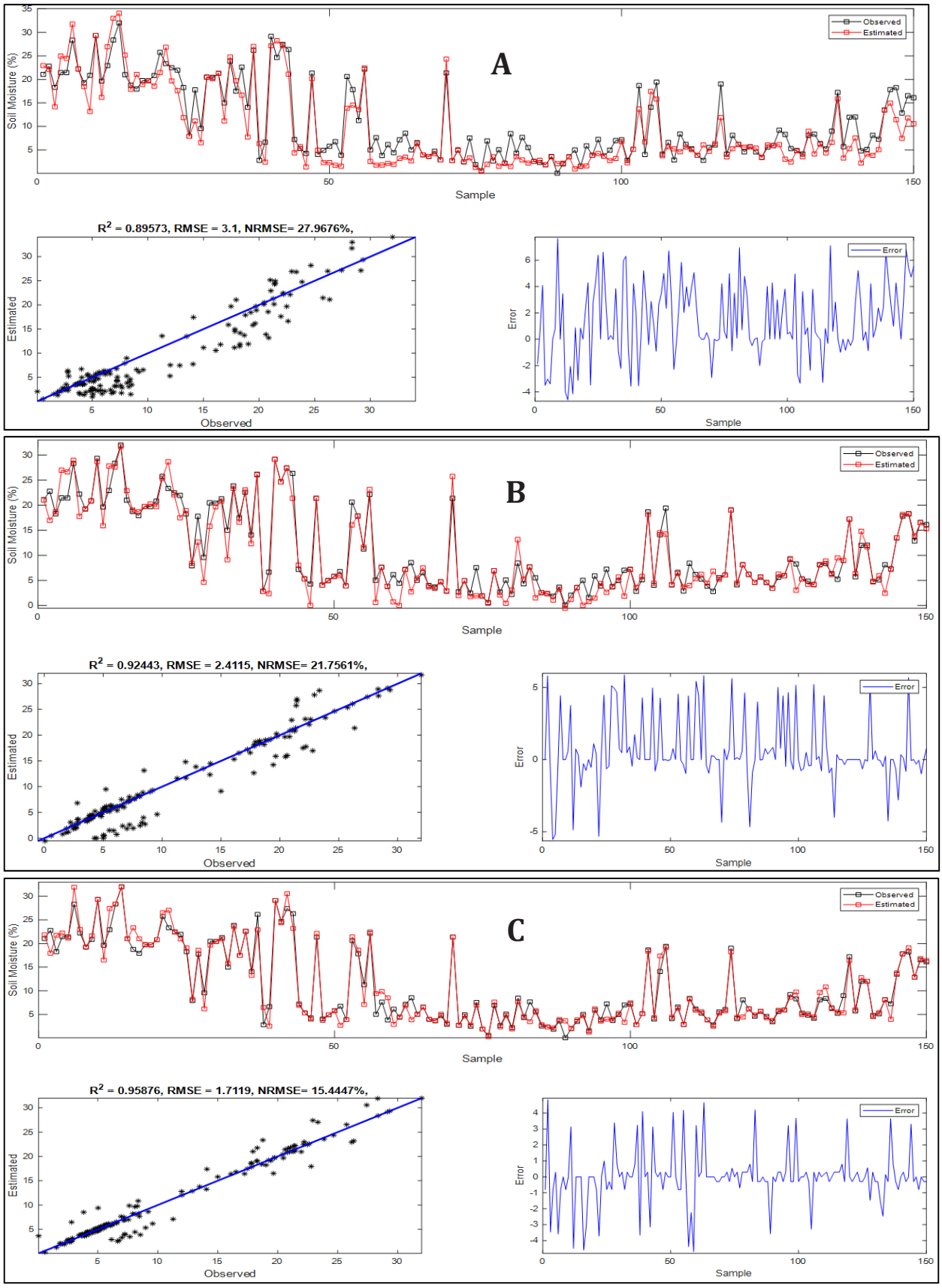

Figure 5) The results of soil moisture modeling by combining the optical trapezoidal method with A) PLS, B) GMDH, and C) LSTM.

tory value of 0.95.

Figure 6 depicts soil moisture maps generated by PLS, GMDH, and LSTM models with 10m spatial resolution. Notably, the soil moisture map produced by the PLS method must exhibit better capability. It demonstrates abrupt changes from high to low soil moisture values and vice versa, particularly in areas with vegetation cover like agricultural lands. These sudden fluctuations do not align logically with the impact of land use on soil moisture or the variations expected based on land use/land cover. Essentially, this method partitions the region into two distinct sections characterized by high and low moisture levels. Although the GMDH method also displays some degree of this partitioning effect, it is less pronounced than the PLS-generated map. The soil moisture values obtained through the GMDH method demonstrate better consistency with the regional logic. On the other hand, the soil moisture map derived from the LSTM method appears much more reasonable. In conclusion, considering both computational costs and the quality of soil moisture mapping, the LSTM method emerges as a superior choice, producing more logical and accurate results compared to PLS and GMDH.

Figure 7 illustrates the SMAP 9km and downscaled 1km products in the region. It is evident from this figure that the 9km product has a very low spatial resolution, particularly when used in small catchments and for assessing the high spatial variability of soil moisture. The SMAP measurements provide an average value for each cell, which may not adequately capture the fine-scale variability within that area. Consequently, this limitation can hinder the applicability of SMAP data for local-scale applications. In contrast, the maps generated using the proposed method, particularly those created by the LSTM model, effectively capture the influence of different land use/covers on the spatial pattern of soil moisture. The 9km SMAP product fails to achieve this due to its low spatial resolution. There are several other limitations related to SMAP products. In areas with dense vegetation, such as forests or dense crops, the SMAP instrument's signal can be heavily attenuated or distorted, which impacts the accuracy of soil moisture estimates in vegetated regions. Additionally, surface roughness resulting from topography or human activities can also affect the accuracy of SMAP soil moisture measurements. Irregular surfaces can scatter and reflect the SMAP signal in complex ways, leading to errors in the soil moisture retrieval process. On the other hand, the 1km product contains several "No Data" areas, particularly near water bodies. A large portion of the area (part B of the figure) needs more data because it is adjacent to Urmia Lake. These limitations greatly restrict the usability of these products, particularly in local applications. This emphasizes the necessity of developing models that can estimate soil moisture based on high-resolution satellite data.

Table 1 presents the final validation outcomes obtained from 50 additional soil moisture samples, affirming the previous findings. The results demonstrate the LSTM deep learning model's superior performance compared to the other two models. Consequently, the GMDH model is the second-most favorable alternative based on prioritization criteria.

Table 1) The results of the final validation of the produced maps

NRMSE (%)	RMSE (%)	R^2	Model
23.70	4.02	0.76	PLS
18.74	3.18	0.84	GMDH
15.21	2.60	0.89	LSTM

Discussion

The poor performance observed might be attributed to the limitations of the Partial

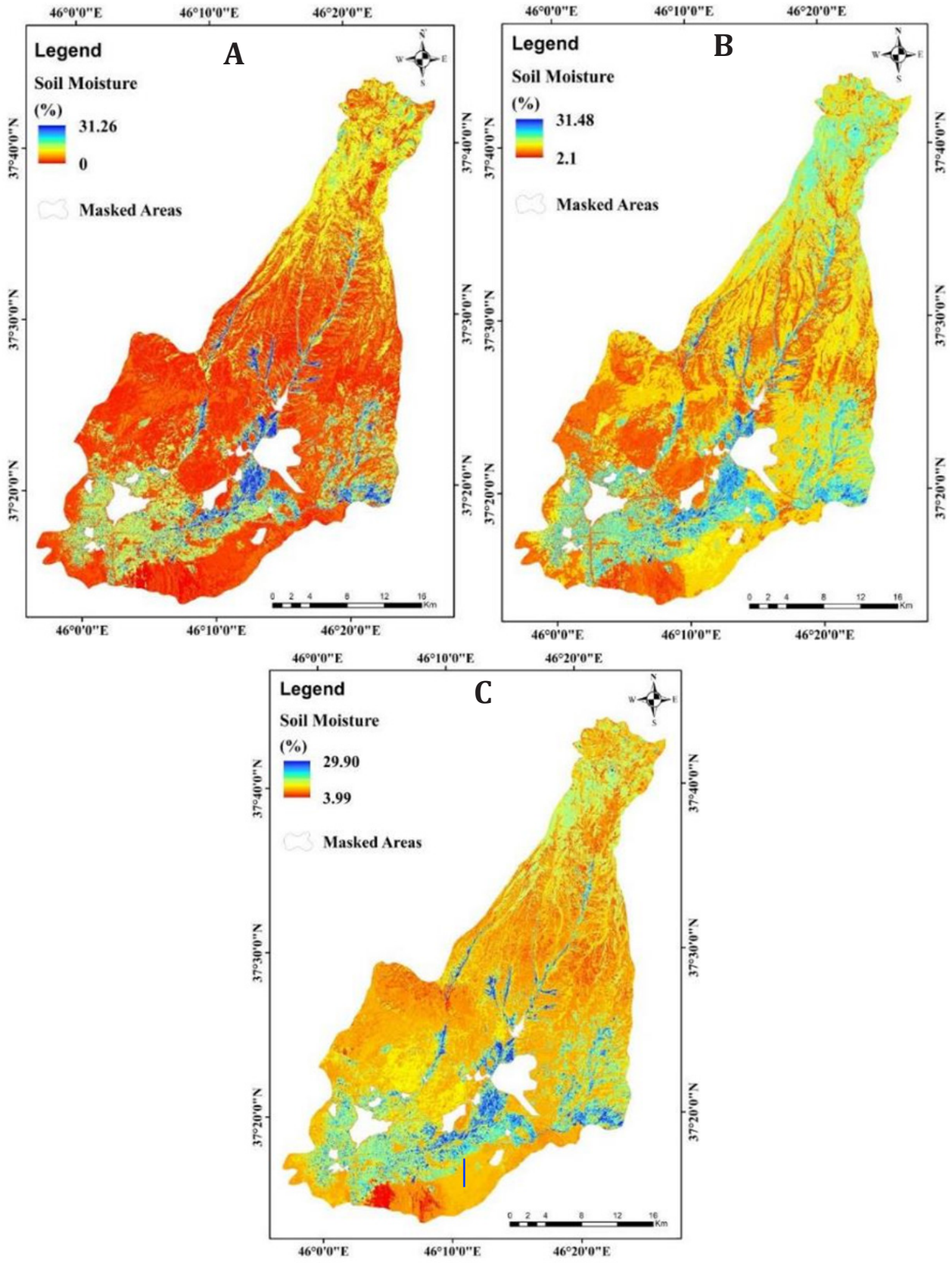

Figure 6) Soil moisture map obtained from A) PLS model, B) GMDH model, and C) LSTM model.

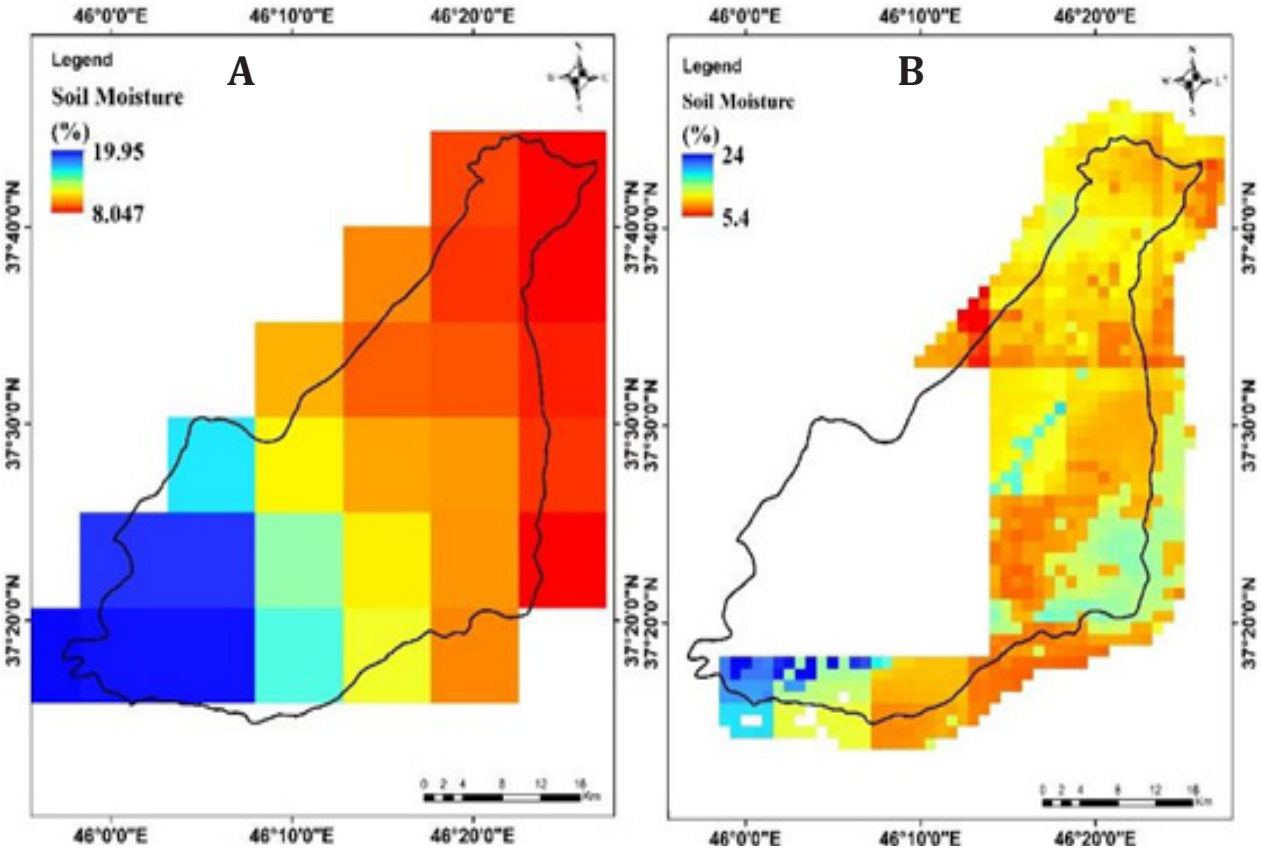

Figure 7) SMAP Products A) 9km product, B) 1km product.

Least Squares (PLS) model. PLS is susceptible to the influence of outliers within the dataset, which could distort the results due to their strong impact on the model. Additionally, PLS encounters challenges when dealing with highly correlated predictor variables, a condition known as multicollinearity. In such instances, accurately determining the individual effects of these predictors becomes difficult for PLS, resulting in less dependable parameter estimates. Furthermore, if the number of predictors is comparatively large compared to the sample size, PLS may be prone to overfitting. Overfitting occurs when the model captures random fluctuations or noise in the data instead of true underlying patterns.

By incorporating relative saturation from the optical trapezoid method as an input, the LSTM model provided more relevant information about the soil moisture dynamics. This additional feature may have contained valuable insights or patterns not captured by the other inputs alone. Including relative saturation can also enrich the feature representation of the LSTM model. The optical trapezoid method captures the relationship between reflectance and soil moisture, which may introduce novel patterns or correlations that were previously missing. This can lead to better model performance and improved accuracy in estimating soil moisture. The optical trapezoid method might provide spatial context information that complements the other inputs used in the LSTM model. Soil moisture levels can exhibit spatial variability, and incorporating a method that accounts for this variability can improve the model's ability to capture localized patterns and variations in soil moisture content. The optical trapezoid method may help reduce uncertainty in soil moisture estimation. By providing an addi-

tional source of information, the model can make more informed predictions, reducing potential errors and improving the overall accuracy of the soil moisture estimates.

The study results show that incorporating the Optical Trapezoid data has improved the performance measures (R^2, RMSE, and NRMSE) for all three models (PLS, GMDH, and LSTM). The PLS model, including Optical Trapezoid, increased R^2 from 0.84 to 0.89, indicating a better fit of the model to the data. Similarly, the RMSE decreased from 3.29 to 3.1, suggesting a reduction in the average prediction error. The NRMSE also decreased from 29.68 to 27.96, indicating a better-normalized model performance. For the GMDH model, there were also improvements when incorporating Optical Trapezoid. The R^2 increased from 0.88 to 0.92, indicating a better fit of the model. The RMSE decreased from 2.8 to 2.41, indicating a smaller average prediction error. The NRMSE decreased from 25.32 to 21.75, suggesting a better normalized performance. The LSTM model also showed enhancements with the inclusion of Optical Trapezoid. The R^2 increased from 0.90 to 0.95, indicating a better fit. The RMSE decreased from 2.46 to 1.7, suggesting a lower average prediction error. The NRMSE decreased from 19.01 to 15.44, indicating an improved normalized performance. Overall, including Optical Trapezoid data has consistently improved the performance of all three models in estimating soil moisture. These improvements can be seen in the increase in R^2, decrease in RMSE, and decrease in NRMSE when comparing the models without and with Optical Trapezoid data. The percentage changes in the performance measures can be calculated to calculate the relative enhancements. For PLS, the relative enhancements can be calculated as follows:

Relative Enhancement of R^2 =
(0.89 - 0.84)/0.84 * 100 = 5.95%
Relative Enhancement of RMSE =
(3.29 - 3.1)/3.29 * 100 = 5.78%
Relative Enhancement of NRMSE =
(29.68 - 27.96)/29.68 * 100 = 5.79%

Similarly, the relative enhancements for GMDH and LSTM models can be calculated as follows:

For GMDH:
Relative Enhancement of R^2 =
(0.92 - 0.88)/0.88 * 100 = 4.55%
Relative Enhancement of RMSE =
(2.8 - 2.41)/2.8 * 100 = 13.93%
Relative Enhancement of NRMSE =
(25.32 - 21.75)/25.32 * 100 = 14.12%

For LSTM:
Relative Enhancement of R^2 =
(0.95 - 0.90)/0.90 * 100 = 5.56%
Relative Enhancement of RMSE =
(2.46 - 1.7)/2.46 * 100 = 31.01%
Relative Enhancement of NRMSE =
(19.01 - 15.44)/19.01 * 100 = 18.77%

These values represent the percentage improvements achieved by incorporating Optical Trapezoid data for each model. These results make the enhancements achieved by including Optical Trapezoid data worthwhile. The models consistently showed improvements in their performance, with the largest enhancement seen in the LSTM model for RMSE. Incorporating Optical Trapezoid data is beneficial for improving these models' accuracy and predictive capabilities.

Another crucial aspect to consider is the disparity in computational cost between different models. Simpler techniques like Partial Least Squares (PLS) entail significantly lower computational requirements than complex methods such as deep learning models. Thus, users should select an appropriate model based on their desired objective, expected accuracy, and precision. Different models require varying amounts of time to train on a given dataset. Simpler techniques like PLS typically have faster training times since they involve solving relatively straight-

forward mathematical equations. On the other hand, complex models like deep learning architectures often involve training large neural networks with numerous parameters, which can be computationally intensive and time-consuming. The memory requirements of a model depend on the size and complexity of the model architecture and the size of the input data. Deep learning models, especially those with many layers and parameters, tend to have higher memory requirements compared to simpler models. This is because they must store intermediate values and gradients during training. Once trained, a model must make predictions or inferences on new, unseen data. The time taken by a model to perform these predictions is known as inference time. Complex models, such as deep learning models, require more computational resources and take longer for inference than simpler models. The computational cost of a model can also be influenced by the hardware infrastructure available. Some models, especially deep learning models, can use specialized hardware accelerators like GPUs (Graphics Processing Units) or TPUs (Tensor Processing Units) to speed up the computations. The following inferences can be made by comparing the computational costs of the mentioned models. PLS is a relatively fast model to train since it involves solving linear equations. It typically has lower computational requirements compared to more complex models like GMDH and LSTM. It does not have high memory requirements as it primarily deals with matrix operations and does not involve storing large amounts of parameters or intermediate values. Once trained, PLS predictions are generally fast since they involve simple matrix multiplications. Inference with PLS is often efficient. GMDH can have moderate to high training times depending on the model architecture's complexity and the dataset's size. GMDH involves iteratively adding and removing layers/nodes, which increases computation. It has higher memory requirements than PLS, especially if the model architecture is significant or the dataset is substantial. It needs to store intermediate values and coefficients during the training process. Inference time with GMDH can vary depending on the complexity of the model. If the model architecture is relatively shallow, inference times can be reasonable. However, deeper architectures may result in longer inference times. LSTM models are computationally expensive to train, especially when working with large datasets or complex architectures. LSTM involves iterative forward and backward passes through the recurrent neural network, which can be time-consuming. This model typically has higher memory requirements compared to PLS and GMDH due to the large number of parameters, recurrent connections, and the need to store hidden states and gradients during training. Inference with LSTM models can be relatively slow compared to PLS and GMDH. The sequential nature of LSTMs and the need to process input sequences step-by-step contribute to longer inference times.

Including a time series of Soil Moisture (SM) maps in this study can offer valuable insights into the region's temporal variability of soil moisture. The proposed methodology allows us to generate a time series of SM data by extrapolating patterns and relationships observed from four days of in-situ soil moisture data. Although the model was trained on specific dates, it can still provide estimates of soil moisture for different SM states throughout the year. The model has learned the underlying relationships between input features (such as sentinel-2 imagery) and soil moisture. As long as the input features accurately represent soil moisture conditions, the model should provide reasonable estimates throughout the year. However, it is

important to acknowledge potential limitations and uncertainties when applying the model across different SM states. Factors like the availability and quality of sentinel-2 imagery and the model's ability to capture the full range of soil moisture variability can influence its performance. External factors, such as changes in land cover or extreme weather events, may also impact the model's accuracy. Further validation and calibration of the model using additional in-situ soil moisture data across a broader range of SM states and periods is recommended to address these concerns. Additionally, continuous soil moisture monitoring through ground-based or remote sensing techniques can enhance the accuracy and reliability of the estimated time series of SM data. One of the limitations is the need for distributed gages of soil moisture data, particularly in countries like Iran. Additionally, gathering in-situ soil moisture data is time-consuming and labor-intensive. Manual collection of measurements at various points requires deploying and maintaining soil moisture sensors in the field, which can be demanding, especially in larger areas or regions with limited accessibility. Therefore, developing models that can use limited data and provide reasonable estimates of soil moisture is valuable. However, exploring the impact of different temporal conditions on the models' performance in predicting soil moisture is recommended for future work.

Conclusion

Based on research findings, the combination of the optical trapezoid method and deep learning models has demonstrated acceptable performance in estimating soil moisture levels. The relative moisture content derived from the optical trapezoid method is a highly suitable input for deep learning models, significantly enhancing their performance. However, it is worth noting that this input has yet to substantially impact the performance of other models, particularly the partial least squares (PLS) model. This may be attributed to the inherent simplicity of the PLS model's structure, which limits its capability to improve beyond a certain extent. Therefore, incorporating auxiliary data has yet to lead to a significant enhancement in its performance. Combining the optical trapezoid method with the group method of data handling (GMDH) model yields average performance results. This approach may be suitable if the user does not require extremely high accuracy and precision and a relative estimation of soil moisture variations is sufficient. Notably, the GMDH model offers the advantage of lower computational cost compared to deep learning models. In contrast, the deep learning long short-term memory (LSTM) model, despite its higher computational cost and more extended calibration time requirements, has exhibited significantly superior accuracy and precision in estimating soil moisture compared to the other two models. This can be attributed to the complex and multi-layered structure inherent in deep learning methods, which allows for comprehensive analysis and representation of underlying patterns in the data. Additionally, the LSTM model's high feature extraction capability from raw input data contributes to its exceptional performance in estimating surface soil moisture.

In conclusion, combining the optical trapezoid method with deep-learning LSTM models presents promising outcomes for soil moisture estimation. It surpasses the other models in terms of accuracy and precision, albeit at the expense of increased computational costs and longer calibration time. Nevertheless, the GMDH model remains a viable alternative when computational efficiency takes precedence over achieving high accuracy. The results follow those of Ahmadinee zhad Baghban and Moosavi [28], Zhou et al. [29],

Yinglan et al. [30], Achieng [31], and Joshi et al. [32], which proved the ability of deep learning in modeling natural processes like soil moisture modeling.

Ethical Permission

All necessary ethical permissions were obtained for the execution of this study.

Conflicts of Interest

The authors declare no conflict of interest.

Funding/Supports

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

V. Moosavi: Conceptualization; Formal analysis; Investigation; Project administration; Supervision; Software; Writing - original draft.

G. Zuravand: Data curation; Formal analysis; Validation; Visualization; Software; Writing - original draft.

SR. Fallah Shamsi: Conceptualization; Formal analysis; review & editing.

References

1. Rouse J.W., Haas R.H., Schell J.A., Deering D.W. Monitoring vegetation systems in the great plains with ERTS third earth resources technology satellite-1 symposium. NASA Spec 1974; 351(1): 309-317.
2. Fathalolomi S., Vaezi E.R., Alavi Panah K., Ghorbani A. Modeling the effect of biophysical properties and surface topography on the spatial distribution of soil moisture in summer (case study: Balkhali Chai watershed). Echohydrology.2020;7(3):563-581.
3. Koohbanani H., Yazdani R. Mapping the moisture of surface soil using Landsat 8 imagery (case study: suburb of Semnan city). Geogr. Environ. Sustain. 2019; 8(3): 65-77
4. Tabatabaeenejad A., Burgin M., Duan X., Moghaddam M. P-band radar retrieval of subsurface soil moisture profile as a second-order polynomial: First AirMOSS results. IEEE. 2014; 53(2): 645-658.
5. Sanli F.B., Kurucu Y., Esetlili M.T., Abdikana S. Soil moisture estimation from RADARSAT-1, ASAR, and PALSAR data in agricultural fields of Menemen plain of western turkey. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Beijing 2008;75-81.
6. Srivastava H.S., Patel P., Sharma Y., Navalgund R.R. Large-area soil moisture estimation using multi-incidence-angle RADARSAT-1 SAR data. IEEE. T. Geosci.Remote.2009;47(8):2528-2535.
7. Silva B.M., Silva S.H.G., Oliveira G.C.d., Peters PHCR, Santos W.J.R.d., Curi N. Soil moisture assessed by digital mapping techniques and its field validation. Cienc.Agrotecnologia.2014;38(2):140-148.
8. Javadi P., Asadi H., Vazife M. Estimation of spatial changes of soil moisture using random forest method and environmental characteristics obtained from satellite images in Marghab basin of Khuzestan. Iran J. Soil Water Res. 2022;52(11):2859-2874.
9. Foroughi H., Naseri A.A., Boroomandnasab S., Hamzeh S., Jones S.B. Presenting a new method for soil-moisture estimation using optical remotely-sensed imagery. Iran. J. Soil Water Res. 2019; 50(3): 641-652.
10. Shokri S.h., Farrokhian Firoozi A., Babaeian E. Estimation of soil moisture by combining physical and hydraulic characteristics of soil with remote sensing optical data using machine learning method. Iran. J. Soil Water Res. 2022; 53(7): 1575-1591
11. Sadeghi M., Babaeian E., Tuller M., Jones S.B. The optical trapezoid model A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. Remote Sens. Environ. 2017; 198(1): 52-68.
12. Sedaghat A., Shabanpoor M., Norozi E.A., Fallah E., Bayat H. Using spectral indices in estimating soil surface moisture based on machine learning algorithm. Iran. J. Soil Water Res. 2021; 52(12): 3001-3018.
13. Sedaghat A., Shabanpoor M., Norozi E.A., Fallah E., Bayat H. Modeling soil surface moisture using machine learning models and Sentinel-2 satellite data. The 17th Iranian Congress of Soil Science and the 4th National Conference on Water Management in the Farm, Wise Soil Revival and Wise Water Governance 2021.
14. Norozi Aghdam A., Behbahani M.R., Rahimi Khoob E., Aghighi H. Moisture model of the soil surface layer using From the index NDVI (a case study: pastures of Razavi Khorasan province). Environ. Mag. 2008; 34(48): 127-136.
15. Norozi Aghdam A., Karami V. Application of remote sensing technology in monitoring and evaluation of irrigation networks drainage in use. The twelfth conference of the National Committee Irrigation and Drainage of Iran 2009;
16. Bagheri K., Bagheri M., Hoseynzade E.A., Parvin M. Estimation of soil moisture using optical, thermal, and radar remote sensing (case study: lands

south of Tehran). Iran. J. Water. Sci. Eng. 2019; 13(47): 63-74.
17. Zeyliger A.M., Muzalevskiy K.V., Zinchenko E.V., Ermolaeva O.S. Field test of the surface soil moisture mapping using Sentinel-1 radar data. Sci. Total Environ. 2022; 807(2): 121-151.
18. Ramezani Charmhine E.A., Zonemat Kermani M. Investigating the effectiveness of multi-layer perceptron neural network support vector regression and variable linear regression methods to predict the level of underground water (study area: Shahrekord plain). J. Water. Manag. Res 2017; 8(15): 1-12.
19. Avazpoor S., Bakhtiari B., Ghaderi K. Investigating the effectiveness of neural network and multivariate regression methods in estimating total solar radiation in several representative stations of arid and semi-arid climates. Iran J. Soil Water Res. 2019; 50(8): 1855-1869.
20. Golkar E., Ahmadi M.M., Qaderi K., Rahimpour M. Peak velocity of pollutant transport prediction in rivers using Group Method Data Handling (GMDH) and intelligent hybrid method (GMDH-HS). J. Water Waste 2018; 30(1): 64-76.
21. Godarzi M.R., Godarzi H. Investigating the effectiveness of data group classification method and wavelet transformation in runoff forecasting (study area: Ghareso watershed). Sci. Res. J. Irrig. Water Eng. 2020; 10(4): 67-81.
22. Moosavi V., Talebi A., Hadian MR Development of a hybrid wavelet packet forecasting group method of data handling (WPGMDH) model for runoff. Water Resour. Manag 2017; 31(1): 43–59
23. Adalat MH, Azmi R., Bagherinejad J. An enhanced LSTM. Method to improve the accuracy of the business process prediction. Indus. Manag. Perspec 2020; 10(39): 71-97.
24. Jordan M. A. Parallel distributed processing approach. Adv. Psychol. 1997; 121(1): 471-495.
25. Yu M., Xu F., Xu W., Sun J., Cervone G. Using long-term memory (LSTM) and internet of things (IoT) for localized surface temperature forecasting in an urban environment. IEEE. 2021; 9(1): 137406-137418.
26. Liu Z.H., Meng X.D., Wei H.L., Chen L, Lu B.L., Wang Z.H., Chen L. A regularized LSTM method for predicting remaining useful life of rolling bearings. Int. J. Automat.Comput. 2021; 18(4): 581-593.
27. Haddadian H., Karimi E., Esfandiar poor E., Hagh nia Gh. RMSE and NRMSE to calculate the effective value of the error of models with different training data sets. The 14th Congress of Soil Sciences of Iran Vali Asr University of Rafsanjan 2015; 16-18 September.
28. Ahmadinezhad Baghban F., Moosavi V. Convolue tional neural networks (CNN)-signal processing combination for daily runoff forecasting. ECOPa ERSIA 2022; 10(3): 231-243.
29. Zhou Y., Zhang Y., Wang R., Chen H., Zhao Q., Liu B., Shao Q., Cao L., Sun S. Deep learning for daily spatiotemporally continuity of satellite surface soil moisture over eastern China in summer. J.Hydrol. 2023; 619: 129308.
30. Yinglan A., Wang G., Hu P., Lai P., Xue B., Fang Q. Root-zone soil moisture estimation based on remote sensing data and deep learning. Environ. Res 2022; 212: 113278.
31. Achieng K.O. Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. Comput.Geosci 2019; 133: 104320.
32. Joshi R.C., Ryu D., Lane P.N.J., Sheridan G.J. Seasonal forecast of soil moisture over Mediterranean-climate forest catchments using a machine learning approach.J.Hydrol.2023;619:12930.