

## Article

# Forecasting Multi-Step Soil Moisture with Three-Phase Hybrid Wavelet-Least Absolute Shrinkage Selection Operator-Long Short-Term Memory Network (moDWT-Lasso-LSTM) Model

W. J. M. Lakmini Prarthana Jayasinghe<sup>1</sup>, Ravinesh C. Deo<sup>1,\*</sup>, Nawin Raj<sup>1</sup>, Sujan Ghimire<sup>1</sup>, Zaher Mundher Yaseen<sup>2</sup>, Thong Nguyen-Huy<sup>3,4</sup> and Afshin Ghahramani<sup>5</sup>

- <sup>1</sup> Artificial Intelligence Applications Laboratory, School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia; lakmini.jayasinghe@gmail.com or lakminiprarthanajayasinghe.jayasinghemudiyanselage@unisq.edu.au (W.J.M.L.P.J.); nawin.raj@unisq.edu.au (N.R.); sujan.ghimire@usq.edu.au (S.G.)
- <sup>2</sup> Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; z.yaseen@kfupm.edu.sa
- <sup>3</sup> Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4300, Australia; thong.nguyen-huy@unisq.edu.au
- <sup>4</sup> Faculty of Information Technology, Thanh Do University, Kim Chung, Hoai Duc, Ha Noi 100000, Vietnam; nhthong@thanhdouni.edu.vn
- <sup>5</sup> Department of Environment and Science, Queensland Government, Rockhampton, QLD 4700, Australia; afshin.ghahramani@des.qld.gov.au
- \* Correspondence: ravinesh.deo@unisq.edu.au

**Abstract:** To develop agricultural risk management strategies, the early identification of water deficits during the growing cycle is critical. This research proposes a deep learning hybrid approach for multi-step soil moisture forecasting in the Bundaberg region in Queensland, Australia, with predictions made for 1-day, 14-day, and 30-day, intervals. The model integrates Geospatial Interactive Online Visualization and Analysis Infrastructure (Giovanni) satellite data with ground observations. Due to the periodicity, transience, and trends in soil moisture of the top layer, time series datasets were complex. Hence, the Maximum Overlap Discrete Wavelet Transform (moDWT) method was adopted for data decomposition to identify the best correlated wavelet and scaling coefficients of the predictor variables with the target top layer moisture. The proposed 3-phase hybrid moDWT-Lasso-LSTM model used the Least Absolute Shrinkage and Selection Operator (Lasso) method for feature selection. Optimal hyperparameters were identified using the Hyperopt algorithm with deep learning LSTM method. This proposed model's performances were compared with benchmarked machine learning (ML) models. In total, nine models were developed, including three standalone models (e.g., LSTM), three integrated feature selection models (e.g., Lasso-LSTM), and three hybrid models incorporating wavelet decomposition and feature selection (e.g., moDWT-Lasso-LSTM). Compared to alternative models, the hybrid deep moDWT-Lasso-LSTM produced the superior predictive model across statistical performance metrics. For example, at 1-day forecast, The moDWT-Lasso-LSTM model exhibits the highest accuracy with the highest  $R^2 \approx 0.92469$  and the lowest RMSE  $\approx 0.97808$ , MAE  $\approx 0.76623$ , and SMAPE  $\approx 4.39700\%$ , outperforming other models. The moDWT-Lasso-DNN model follows closely, while the Lasso-ANN and Lasso-DNN models show lower accuracy with higher RMSE and MAE values. The ANN and DNN models have the lowest performance, with higher error metrics and lower R2 values compared to the deep learning models incorporating moDWT and Lasso techniques. This research emphasizes the utility of the advanced complementary ML model, such as the developed moDWT-Lasso-LSTM 3-phase hybrid model, as a robust data-driven tool for early forecasting of soil moisture.

**Keywords:** soil moisture model; wavelet transform; artificial intelligence; hybrid models; deep learning



**Citation:** Jayasinghe, W.J.M.L.P.; Deo, R.C.; Raj, N.; Ghimire, S.; Yaseen, Z.M.; Nguyen-Huy, T.; Ghahramani, A. Forecasting Multi-Step Soil Moisture with Three-Phase Hybrid Wavelet-Least Absolute Shrinkage Selection Operator-Long Short-Term Memory Network (moDWT-Lasso-LSTM) Model. *Water* **2024**, *16*, 3133. <https://doi.org/10.3390/w16213133>

Academic Editor: Dongmei Zhou

Received: 12 August 2024

Revised: 20 September 2024

Accepted: 21 October 2024

Published: 1 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Soil moisture, an integral component of the soil-plant-atmosphere water cycle, refers to the water present in the soil, which is essential for maintaining plant growth [1]. It is a key factor in determining irrigation water requirements [2]. Forecasting soil moisture is invaluable for understanding future soil moisture trends, managing water stress conditions affecting crops, and planning irrigation schedules while conserving limited water resources. The Bundaberg region in Queensland, Australia, the focus of this study, is extensively used for growing commercial crops. Thus, a soil moisture forecasting model will significantly benefit agricultural operations in this region.

Data-driven predictive models have demonstrated superior competency in predictions of soil moisture [3] and other hydro-meteorological variables such as precipitation [4], drought [5,6], evapotranspiration [7] and river flow [8]. For example, Jamei et al. [9] constructed multi-level pre-processing model frameworks using NASA's Soil Moisture Active Passive (SMAP) satellite datasets for multi-step (one and seven days ahead) daily forecasting of Surface Soil Moisture (SSM) in Iran's dry and semi-arid regions. This experiment integrated Boruta Gradient Boosting Decision Tree (Boruta-GBDT) feature selection and Multivariate Variational Mode Decomposition (MVMD) techniques with advanced Machine Learning (ML) models, including Bidirectional Gated Recurrent Unit (Bi-GRU), Cascaded Forward Neural Network (CFNN), Adaptive Boosting (AdaBoost), Genetic Programming (GP), and classical Multilayer Perceptron neural network (MLP). The results indicated that MVMD-Boruta-GBDT-CFNN outperformed all other hybrid models in one and seven days soil moisture forecasting across all tested sites. Basak et al. [10] also employed a data-driven approach to forecast soil moisture. Their study developed and tested two data-driven models based on Naive Accumulative Representation (NAR) and the Additive Exponential Accumulative Representation (AEAR). The proposed NAR and AEAR models have demonstrated their competitiveness in forecasting tasks, performing better than several well-established and cutting-edge benchmark models.

Among data intelligence approaches, deep learning (DL), the latest generation of artificial intelligence systems, has become increasingly popular and performs exceptionally well in both industrial and scientific research [11]. The strength of DL techniques lies in their ability to learn complex nonlinear functions of input data with low-level information, allowing them to successfully capture and extract the detailed features from extensive raw input data sets accumulated over decades. These capabilities make DL techniques particularly valuable for research initiatives. The LSTM algorithm is one of the DL artificial intelligence approaches that is being utilized to forecast various hydrological and environmental variables, such as water quality [12] and rainfall-runoff [13]. Several studies have investigated the feasibility of using LSTM-based model for SM prediction.

For example, in South Louisiana in the United States, ElSaadani et al. [14] found that among the spatial-temporal models tested, the ConvLSTM outperformed other Convolutional Neural Network (CNN) and LSTM-based models in SM prediction. To improve the SM prediction accuracy, Li et al. [15] experimented with a unique residual learning encoder-decoder model (EDT-LSTM), using data from 13 sites across different countries. This model demonstrated improved accuracy in forecasting moisture levels in 5 cm deep surface soil layers for 1, 3, 5, 7 and 10 days ahead. In another study, Suebsombut et al. [16] developed LSTM-based models to forecast SM values in Chiang Mai province, Thailand, showing that the LSTM-based model performs well in predicting SM. Additionally, Zeynoddin and Bonakdari [17] proposed two DL methods, Genetic and Teacher-Learner-based Algorithms (GA and TLA), coupled with LSTM for SM forecasting in Quebec, Canada. Their results showed that TLA-LSTM model was more computationally efficient and therefore a better option than GA-LSTM.

To further enhance forecasting model capabilities, many researchers have recently been developing hybrid models. It is common to combine data pre-processing techniques with forecasting models when designing these hybridized models, as pre-processing methods particularly work well with nonlinear and non-stationary time series data. In artificial

intelligence model hybridization, feature selection is a popular data pre-processing method and a variety of research studies have shown that it enhances the model's performance. The purpose of this process is to reduce the high dimensionality of input data by screening out the most correlated input data sets to the target variable data set as a first step in advanced data-driven model development. For example, Iterative Input Selection (IIS) has been used to forecast streamflow [3], while Boruta-random forest (Boruta) has been applied to forecast evapotranspiration [18] and soil moisture [9]. Additionally, Neighbourhood Component Analysis (NCA) has been used to forecast wind speed [19] and agricultural drought [20].

The Lasso feature selection method, used in this study, has also been employed in hydrological forecasting research. For instance, Alizadeh et al. [21] developed a Support Vector Regression (SVR) based model for monthly stream flow prediction at the Karaj River in Iran, using both Lasso and Particle Swarm Optimization-Artificial Neural Networks (PSO-ANN) feature selection methods to identify the most correlated input variables. The results indicated that Lasso input selection is more accurate than the PSO-ANN algorithm, therefore improving the accuracy of the model forecast. Chu et al. [22] has also employed the Lasso feature selection technique along with Fuzzy C-means (FCM) classification and Deep Belief Networks (DBN) deep learning model (Lasso-FCM-DBN) to forecast streamflow at gauge stations in the Tennessee River catchment, USA. They found that the Lasso-FCM-DBN approach enhanced the performance of streamflow prediction compared to ANN. However, this feature selection technique has not so far been employed with any deep learning approach in soil moisture forecasting model development.

Along with feature selection, wavelet decomposition is a common data pre-processing step in data intelligence model hybridization. Hydrological and water resources time series data are inherently complex due to periodicity, transients, and trends. Wavelet transform algorithms can decompose this complex data into sub-time series data that are more interpretable for data-driven models. As a result, wavelet decomposed data often improve model performance and are widely used in hydrological and water resources-related prediction applications.

The wavelet decomposition methods widely used in recent model hybridization works are Discrete Wavelet Transformation (DWT), Maximum Overlap Discrete Wavelet Transform with Multi-Resolution Analysis (moDWT-MRA), Maximum Overlap Discrete Wavelet Transform (moDWT), and (a) trous (AT) algorithm [23]. For instance, Prasad et al. [3] employed moDWT in their hybrid IIS-moDWT-ANN model designed for forecasting streamflow and it has shown better accuracy than the counterpart single and hybrid benchmark models. Adib et al. [24] in their study for predicting one-day-ahead snow depth (SD) at the North Fork Jocko snow telemetry (SNOTEL) station in Missoula, Montana (US), tested different wavelet transform (WT) approaches including discrete wavelet transform (DWT), maximal overlap discrete wavelet transform (MODWT), and multiresolution-based MODWT (MODWT-MRA) along with autoregressive integrated moving average (ARIMA), and artificial intelligence (AI) models. Their findings showed that hybrid ARIMA-AI models produced more accurate results than standalone ARIMA and AI models, highlighting the wavelet technique's capacity to enhance the model performances.

It is important to note that DWT and moDWT-MRA can introduce errors into forecasts due to boundary condition-related issues and provide results that are not realistically achievable in real-world scenarios. Therefore, these methods are unsuitable for practical applications. However, moDWT and AT wavelet transform algorithms, when applied correctly, can resolve boundary condition related issues [23]. These boundary condition issues, their impact to the model forecast and the remedies to overcome them will be discussed later in detail under the theoretical overview section of this paper.

Despite these constraints, many recent hybrid forecasting model development studies, including the examples mentioned earlier that employed wavelet transform techniques to decompose hydrological and water resources related data, have not adequately considered these limitations. Instead, they have used DWT and moDWT-MRA regardless of their shortcomings. Furthermore, moDWT and AT wavelet transform algorithms, which do

not add errors to model forecasts due to boundary condition issues, are not much used in hydrological prediction as DWT and moDWT-MRA. Therefore, these algorithms still need to be explored further.

In this study, time series data from satellites and ground stations are combined. The methodology section provides detailed information about the types of data collected, and their resolutions and sources. It is well-documented that data from satellite sensor variables can lower the accuracy of hydrological variable predictions [25]. This issue can be minimized by integrating ground-based and satellite-based data, as done in this study. Ghimire et al. [26] used data from Goddard's Online Interactive Visualization and Analysis Infrastructure (Giovanni) combined with reanalysis data from the European Centre for Medium-Range Weather Forecasting (ECMWF) to forecast long-term solar radiation. Similarly, Ahmed et al. (2021b) used a combination of satellite GLDAS data, ground Scientific Information for Landowners (SILO) data, and meteorological indices to predict soil moisture.

Due to the high dimensionality of hydrological time series data extracted in large volumes, this study requires feature selection and wavelet decomposition data pre-processing techniques. Thus, this hybridizing approach used moDWT and Lasso algorithms for wavelet decomposition and feature selection, respectively, along with an LSTM data-driven DL network. This methodology is novel, as there is no existing literature that explains the use of Lasso feature selection and moDWT data decomposition techniques in SM prediction. Additionally, this study has implemented remedial procedures to overcome the boundary condition related issues which are likely to add errors to real-world forecasts. That is also a forward step in prediction studies that use wavelet transform data decomposition procedures. Further, the proposed combination of algorithms, abbreviated as the moDWT-Lasso-LSTM model, has not yet been tested in other geographic locations, thus filling this gap in soil moisture prediction research.

The objectives in this study are threefold:

1. To develop deep learning methods for forecasting soil moisture (SM) at 10 cm depth, integrating moDWT data decomposition methods with Lasso methods as feature selection procedures to produce a prediction model based on LSTM utilizing satellite data from GIOVANNI and ground data from SILO.
2. To employ the hybrid moDWT-Lasso-LSTM model in multi-step SM forecasting, i.e., 1-day ( $t + 1$ ), 14-day ( $t + 14$ ) and 30-day ( $t + 30$ ) SM forecasting.
3. To compare the objective model with benchmark models: LSTM, DNN, and ANN (standalone models), Lasso-LSTM, Lasso-DNN, and Lasso-ANN (2-phase hybrid models) and moDWT-Lasso-DNN and, moDWT-Lasso-ANN (3-phase hybrid models).

The above objectives have been established in this study to design a precise SM forecasting model for short-, medium- and long-term SM predictions and to confirm its comparative advantage. SM, as a major form of water resource, significantly influences agricultural production and consequently affects food security. Similar to other forms of water resources available globally, SM is a limited resource with growing demand due to the expansion of agricultural activities. Under SM depleted conditions, the demand for water storage for irrigation purposes increases, thereby restricting water availability for other purposes like drinking and recreational activities.

Currently, agriculture accounts for an average of 70 percent of global freshwater withdrawals [27]. Accurate SM predictions are crucial for early identification of moisture stress in crops and actual irrigation water requirements in advance. Furthermore, precise SM predictions can help to minimize water wastage in irrigation activities, provide early warnings of crop production fluctuations, and conserve valuable water resources. Given these benefits, this study aims to design a SM forecasting model using LSTM deep learning algorithm, combined with Lasso feature selection and moDWT wavelet transform data decomposition techniques.

Further, this study aims to apply the proposed model to multi-step SM forecasting scenarios, including 1-day ( $t + 1$ ), 14-day ( $t + 14$ ) and 30-day ( $t + 30$ ) multi-step SM

forecasts. This will give an opportunity to observe its usefulness in short-, medium- and long-term forecasting time horizons. A wide range of forecasting time horizons is important for implementing remedial actions against SM stress conditions at different levels. For instance, short-term SM predictions may be important in taking prompt actions against potential sudden crop failures due to moisture stress, while long-term SM predictions are valuable for strategic planning to cope with future drought conditions, conserving water resources, and ensuring stable crop production in the long run. In addition, by comparing the proposed model with competitive rival models, this study aims to recognize the performance improvement without overestimating the proposed model capabilities. The research objectives in this study will help to make further improvement in the precision of SM prediction and thereby adding valuable contribution to the SM prediction studies.

## 2. Theoretical Overview

This section describes the moDWT, Lasso, and LSTM algorithms used in the current study to develop the model. Additionally, benchmark models were used to evaluate the target model's performance. These benchmarks are relatively recent machine learning models with neural networks, similar to LSTM, and were chosen for their advanced capabilities, making them suitable competitive benchmarks for the data-driven forecasting algorithm used in this study.

The theoretical foundation of the single neural layer ANN machine learning model is described in earlier research [28]. In hydrology, ANN is widely used and has proven competent in various prediction tasks. For instance, Prasad et al. [29], developed an ANN-CoM based multi-model ensemble strategy to forecast monthly soil moisture at four farming locations in the Murray-Darling Basin, Australia. The ANN-CoM model, validated against Volterra, Random Forest, M5 tree, and ELM models, demonstrated high competency in capturing the nonlinear dynamics of soil moisture levels. Similarly, Shirsath and Singh [30] constructed ANN and Multiple Linear Regression (MLR) models for pan evaporation estimation, and statistical comparisons revealed that the ANN model outperformed other models.

The DNN algorithm, an improvement over ANN, has been detailed by Le et al. [31] and is progressively used in hydrology. It consists of multiple neural layers and falls under the deep learning subset of machine learning. El Bilali et al. [32] developed an interpretable ML framework to forecast daily pan evaporation using hourly climate datasets, employing DNN along with Extra Tree, XGBoost, and SVR models. The interpretability of these models was evaluated using SHAP, Sobol-based sensitivity analysis, and LIME, showing good consistency with real hydro-climatic processes of evaporation in a semi-arid environment.

### 2.1. Decomposition Method: Maximum Overlap Discrete Wavelet Transform (moDWT)

Maximum Overlap Discrete Wavelet Transform (moDWT) decomposition method decomposes complex time series data, characterized by multiple periodicities, transients, and trends, into high and low frequency sub-time series, termed as wavelet and scaling coefficients. These wavelets and scaling coefficients resulting from moDWT are defined as follows [23]:

$$W_{(j,t)} = \sum_{n=1}^{L-1} h_l X_{j-1,t-2^{j-1}n} \text{mod } N \quad (1)$$

$$V_{(j,t)} = \sum_{l=0}^{L-1} g_l X_{j-1,t-2^{j-1}l} \text{mod } N \quad (2)$$

where  $X$  is a time series input vector with  $N$  values;  $j = 1, 2, \dots, J$ , and  $J$  represents the level of decomposition at the time  $t$ ; the  $j$ th level wavelet ( $W_{(j,t)}$ ) and scaling ( $V_{(j,t)}$ ) filters of moDWT are represented as  $h_l$  and  $g_l$ , respectively, and  $L$  is the  $j$ th level filters' width.

The moDWT addresses issues associated with other data decomposition algorithms such as DWT and moDWT multi resolution analysis (moDWT-MRA). These issues arise when the decomposition process cannot calculate the output values of coefficients accu-

rately (without adding errors) due to the lack of necessary time series observations at specific time points. These errors are termed boundary conditions.

For instance, DWT and moDWT-MRA require future data at certain points to calculate detail and approximation coefficients. While historical time series data can provide this future data, real-world scenarios do not have access to future data, leading to inaccuracies in the calculated coefficients. Consequently, models developed using DWT and moDWT-MRA may fail to accurately forecast soil moisture (SM) in practical implementations.

The moDWT effectively addresses these boundary condition issues by relying solely on current and past time series data for its decomposition outputs, namely wavelet and scaling coefficients, without requiring future data. However, moDWT cannot accurately calculate decomposition outputs for data points at the beginning of a time series because the necessary past data is unavailable. As a result, the wavelet and scaling coefficients calculated for early data points are incorrect and are affected by boundary conditions.

The number of incorrect or boundary-condition-affected coefficients depends on the decomposition level and the wavelet filter used. Higher decomposition levels and longer wavelet filters tend to increase the total number of incorrect coefficients. The total number of incorrect coefficients can be determined using Equation (3) [23].

$$L_J = (2^J - 1)(L - 1) + 1 \quad (3)$$

where  $L_J$  represents the number of wavelet and scaling coefficients affected by the boundary condition for decomposition level  $J$  and a wavelet filter of length  $L$ .

In order to improve model forecasting accuracy, it is necessary to remove all these boundary-condition-affected, incorrect wavelet and scaling coefficients derived at the beginning of the data set. High decomposition levels and lengthy wavelet filters result in more incorrect coefficients that must be removed, reducing the number of correct coefficients available for model training. Therefore, the appropriate selection of decomposition levels and wavelet filters is essential for enhancing model performance. Quilty and Adamowski [23] provides detailed discussions on boundary condition issues arises due to unavailability of future data when employing DWT and moDWT-MRA for data decomposition.

There is no standard rule for determining the optimal decomposition level and wavelet filter type. Increasing the number of boundary-condition-affected coefficients unnecessarily should be avoided, as it leaves an inadequate number of correct coefficients for model training. However, Equation (4) can be used to calculate the maximum decomposition level ( $J$ ) that can be adopted [33].

$$J = \text{int}(\log_2 N) \quad (4)$$

## 2.2. Feature Selection Method: Least Absolute Shrinkage and Selection Operator (Lasso)

In this study, the Lasso algorithm [34] is employed as a feature selection technique after decomposition of input time series variables by the moDWT algorithm. Suppose that the dataset consists of  $p$  input variables and  $N$  observations. Let  $X = [x_1, x_2, \dots, x_p] \in R^{(N \times p)}$  is the input data matrix, in which each column denotes an input variable and  $y = [y_1, y_2, \dots, y_N] \in R^N$  is the response variable where the response value at observation  $j$  is represented by  $y_j$  and  $x_j$  is a vector containing  $p$  characteristics.

The Lasso method therefore follows the work of [35]:

$$\beta = \text{argmin} \|y - X^T \beta\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (5)$$

$$L_1 = \lambda \sum_{j=1}^p |\beta_j| \quad (6)$$

By applying a  $L_1$  penalty for the regression coefficients, the Lasso technique degrades least-squares by shrinking the regression coefficients ( $\beta$ ) to zero. The variables are chosen to be included in the model during this feature selection procedure if their coefficients

after the shrinking step are still non-zero. This process minimizes the prediction error by reducing the complexity of the model.

### 2.3. Data Driven Forecasting Model: Long Short-Term Memory Network (LSTM)

The LSTM is a unique type of Recurrent Neural Network (RNN) [36] in connection with traditional artificial neural networks that can recognize the intrinsic characteristics of time sequence predictors and targets, considering the recurrent patterns and tendencies throughout long periods. Input, output, and forget gates are the main components of the special units, or memory blocks, that the LSTMs use to operate and these memory blocks regulate the flow of information and are continuously updated [37]. The 4 steps calculations are described as follows [38]:

- The forget gate  $f_t$  is used by the LSTM layer to determine which data should either be discarded or retained depending on the most recent hidden layer output  $h_{t-1}$  and the new input  $x_t$ :

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (7)$$

where  $w_f$  stands for weight matrix;  $b_f$  stands for bias vector and  $\sigma(\dots)$  stands for sigmoid logistic function.

- After the information is updated by utilizing a “input gate”  $i_t$ , the LSTM layer determines which signal must be kept in the newly formed cell state  $c_t$  that is denoted as the new candidate cell state  $\hat{C}_t$ :

$$\hat{C}_t = \tanh(w_C[h_{t-1}, x_t]) + b_C \quad (8)$$

$$i_t = \sigma(w_i[h_{t-1}, x_t]) + b_i \quad (9)$$

where the hyperbolic tangent function is denoted by  $\tanh(\dots)$ .

- The “forget gate”  $f_t$  removes unwanted information from the old cell state  $C_{t-1}$  to  $C_t$  and the “input gate”  $i_t$  obtains a new candidate cell state  $\hat{C}_t$ :

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (10)$$

- The cell state  $C_t$  and the “output gate”  $o_t$  are then used to calculate the output  $h_t$ :

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

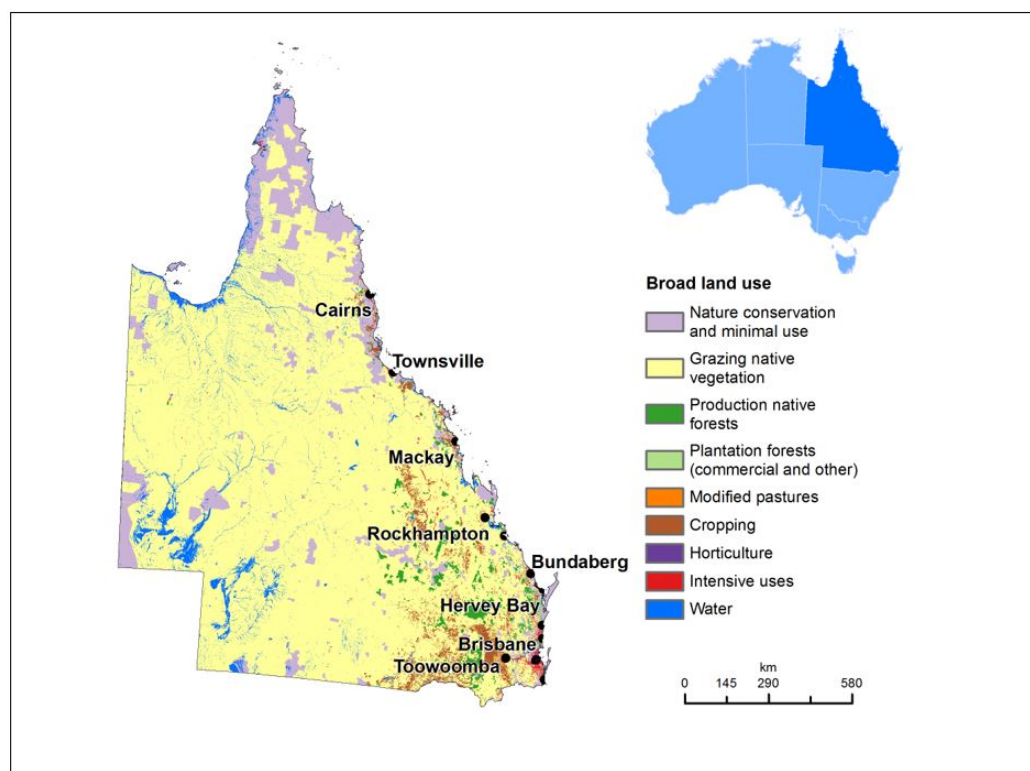
## 3. Materials and Method

This section presents a detailed breakdown of the methodology employed in this study, divided into four key areas. First, the study region is described to establish the geographical and environmental context. Next, the data utilized in the research is presented, followed by the design of the predictive model, which outlines the core framework and approach. Finally, the model permanence evaluation subsection discusses the metrics used to assess the models’ accuracy and effectiveness. Tables A1 and A2 show the lists of symbols and prefixes used in this study, respectively.

### 3.1. Study Region

Figure 1 shows the study area where 1-day, 14-day and 30-day lead time soil moisture forecast model was developed using this 3-phase long short-term memory network, wavelet and Lasso regression moDWT-Lasso-LSTM model. The case study focuses on the Bundaberg region, located at 152.32° E, 24.91° S in Wide-Bay Burnett region of Queensland state, Australia. This region covers 6444 square km and features a subtropical climate with warm, wet summers and mild winters. The average annual temperature is around 20 °C, with an average annual rainfall of 774 mm, most of which falls in the summer.

During the hot summer months from November to March, the average daily maximum temperature exceeds 28 °C, with January being the warmest month, having average maximum and minimum temperatures of 30 °C and 23 °C, respectively. July, the coldest month, sees average minimum and maximum temperatures of 14 °C and 21 °C, respectively. Bundaberg experiences significant seasonal fluctuations in monthly rainfall, peaking in February with an average of 120 mm and dropping to its lowest in September, averaging 28 mm. Humidity levels vary greatly throughout the year, while the average hourly wind speed experiences mild seasonal variation [39,40].



**Figure 1.** Study site geographical location and land use of the region and surrounding areas [41].

According to the Bundaberg Regional Council's population statistics, the region's total resident population reached 100,118 in 2021, with a population density of 15.54 people per square km. The agricultural, forestry and fishing sector is valued at approximately \$1.2 billion, making it the food bowl capital in Australia, contributing 12% of Queensland's total agricultural production. The region's fertile soils, favourable climate, and steady water supply support a wide range of agricultural operations. For instance, this region contributes to producing 50% of Australia's macadamia nuts, representing the largest proportion of the country's macadamias production.

Bundaberg also leads in avocado production, allocating the largest land area for avocado farming in Australia. In addition, this region significantly contributes to the production of mandarin, sweet potato, passion fruit, and pastures. These factors confirm that Bundaberg provides an ideal platform for agricultural industries, making this sector dominant in the region. Developing an accurate forecasting model to predict soil moisture 1, 14, and 30 days in advance is strategically important for the early identification of water deficit and surplus conditions affecting crop production. Furthermore, it will aid in employing precision irrigation practices, thereby conserving valuable water resources for future use and other water-demanding activities. Therefore, the Bundaberg region was selected for this study, which aims to develop a deep learning artificial intelligence model to forecast soil moisture.



### 3.2. Data

To conduct this research, satellite and ground-based daily climatic data of 15 predictive and target variables from 1 January 2005 to 31 December 2020, were collected for the selected study site. This period comprises a total of 5844 data points. Satellite-based data, including the target variable, soil moisture (SM) (0–10 cm depth), were obtained from the Goddard Online Interactive Visualization and Analysis Infrastructure (Giovanni) platforms, including the Global Land Data Assimilation System (GLDAS) and the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS), with a spatial resolution of 0.01 degrees. Giovanni, a web interface developed by NASA, facilitates the analysis of gridded data from various satellites and surface observations, offering simple access to a massive amount of Earth Science remote sensing data [42]. Ground-based data for this study were collected from the Scientific Information for Landowners (SILO) database for the same period. This database is managed by the Queensland Government [43]. Table 1 lists the data sources and predictor variables used in this study along with their corresponding acronyms.

**Table 1.** Data derived from the Satellite-based Goddard Online Interactive Visualization and Analysis Infrastructure (GIOVANNI) Global Land Data Assimilation System (GLDAS) spectrometer satellite and Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) spectrometer with Scientific Information for Land Owners (SILO) used as ground-based predictors to develop the proposed hybrid moDWT-Lasso-LSTM and benchmark models.

Data Source	Name of Predictor Variable	Acronym	Unit	
GIOVANNI-Satellite data	Soil Temperature (0–10 cm depth)	ST0-10	K	
	Soil Temperature (10–40 cm)	ST10-40	K	
	Soil Temperature (40–100 cm)	ST40-100	K	
	Soil Moisture (10–40 cm)	SM10-40	kgm <sup>-2</sup>	
	Soil Moisture (40–100 cm)	SM10-40	kgm <sup>-2</sup>	
	Soil Moisture (100–200 cm)	SM10-40	kgm <sup>-2</sup>	
	GLDAS Model	Ground Water Storage	GWS	mm
SILO-Ground based data	Maximum Temperature	max-temp	°C	
	Minimum Temperature	min-temp	°C	
	Solar radiation	radiation	MJm <sup>-2</sup>	
	Relative humidity at max temp	rh-tmax	%	
	Relative humidity at min temp	rh-tmin	%	
	Mean sea level pressure	mslp	hPa	
	Rainfall	rain	mm	
Reference Evapotranspiration	ET	mm		

The 15 predictor variables were selected based on a correlation matrix and trial runs, where variables were included and excluded to assess their correlation with the target variable. These trials showed that predictor variables with a weaker correlation to the target variable tend to reduce the forecast accuracy of all models. Therefore, to improve forecasting accuracy, only predictor variables with a sufficiently high correlation with the target variable were selected. Additionally, soil moisture data from other layers (i.e., SM10-40, SM40-100, SM100-200) that had a good correlation with the target layer (i.e., SM0-10) were also considered.

### 3.3. Predictive Model Design

#### 3.3.1. Computers and Software

The proposed multi-stage moDWT-Lasso-LSTM model and all other benchmark models were developed using a computer configured with an Intel Core *i7* @ 3.3 GHz processor (Intel Corporation, Santa Clara, CA, USA) and 16 GB of memory loaded with freely downloadable deep learning libraries of Keras 2.0 [44] and TensorFlow 2.0 [45] in Python 3.1. The moDWT data decomposition algorithm and the Lasso feature selection method are run on a MATLAB R2019b and Python platforms, respectively. Furthermore, the study has

adopted “matplotlib” and “seaborn” tools in the Python programming environment to produce graphical illustrations for visualization of the result.

### 3.3.2. Identification of Model Inputs

Row data of all 15 predictor variables are time lagged against row data of target variable, i.e., SM accordance with forecasting lead times  $t + 1$ ,  $t + 14$ , and  $t + 30$  respectively. In case of lagging data for  $t + 1$  SM forecasting, all data are stacked in a way that, data of predictor variables at each time point in the predictor data sequence are always coinciding with 1 day ahead data of the target variable. Similarly, in case of  $t + 14$  and  $t + 30$  SM forecasting, data of predictor variables at each time point are always coinciding with 14 and 30 days ahead target variable data respectively.

### 3.3.3. Data Decomposition Using moDWT for Developing Three Phase Hybrid Models

This study has adapted the moDWT as the data decomposition algorithm to decompose the lagged data of all predictor variables in the development of three-phase hybrid models: the moDWT-Lasso-ANN, the moDWT-Lasso-DNN and the moDWT-Lasso-LSTM. However, the target variable’s data were not decomposed using moDWT, as this does not provide additive reconstruction functions [23].

Given that there are no established rules for determining the optimal decomposition level and wavelet filter type for a decomposition process, this study employs trial-and-error procedures, a common practice in similar research. However, Equation (4), discussed in the theoretical overview section, is used for calculating the maximum decomposition level, which in this research is determined to be 9. According to the Equation (3), such higher decomposition level increases the number of incorrect wavelet and scaling coefficients, exacerbated when combined with longer wavelength wavelet filters. Therefore, this study selects three different decomposition levels-2, 4 and 6-below the maximum level (9) for the trial-and-error process.

In this study, seven commonly used wavelet filters from three different wavelet families were employed, specifically: Haar (wavelet length of 2), *db2*, *db4* and *db6* (wavelet lengths of 4, 8, and 12, respectively) from the Daubechies family, and *fk4*, *fk8*, and *fk14* (wavelet lengths of 4, 8, and 14, respectively) from the Fejer-Korovkin family. This resulted in 21 trials to determine the best combination of decomposition level and wavelet filter for each three-phase hybrid model at a particular lead time. Given the three lead times ( $t + 1$ ,  $t + 14$  &  $t + 30$ ) and three 3-phase hybrid models (moDWT-Lasso-ANN, moDWT-Lasso-DNN, and moDWT-Lasso-LSTM), a total of 189 trials were conducted.

Despite the availability of many wavelet filters from various families, the study limited its scope to these seven filters due to time constraints and to maintain study simplicity. The longest wavelet filter used was *fk14*. Wavelet filters longer than *fk14* were not considered to avoid increasing the number of boundary condition-affected wavelet and scaling coefficients. For instance, using the *fk14* filter with decomposition level six results in 820 boundary condition-affected coefficients, calculated by Equation (3).

To ensure consistency across trials, 820 wavelet and scaling coefficients were removed from the beginning of each dataset used for the standalone and two-phase hybrid models. This standardization ensured that all trials, differentiated by wavelet filters, decomposition levels, and forecasting model combinations, were based on the same dataset.

### 3.3.4. Feature Selection Using Lasso to Develop 2-Phase and 3-Phase Hybrid Models

Feature selection is conducted using the Lasso algorithm to identify the predictor variables most correlated with the target variable for developing the 2-phase hybrid models, such as Lasso-ANN, Lasso-DNN and Lasso-LSTM. For this purpose, undecomposed predictor variable data is used separately for each lead time scenario, and only the undecomposed data of the selected predictor variables are fed into the 2-phase hybrid forecasting models. Additionally, the Lasso algorithm is employed to identify the most correlated wavelet and scaling coefficient data series derived from the original predictor variable

data series during the data decomposition process using moDWT. This step is crucial for developing the three-phase hybrid models (moDWT-Lasso-ANN, moDWT-Lasso-DNN, and moDWT-Lasso-LSTM). The feature selection task is performed separately for each model and lead time scenario.

### 3.3.5. Data Normalization

In this study, the data ranges for each predictor variable in the data sets prepared for the forecasting models vary across all scenarios. Variables with larger data ranges can be unnecessarily favoured in model forecasting over inputs with narrower ranges regardless of their intrinsic relationship. To address this issue, data normalization is carried out using Equation (13) to scale the data within the 0-1 range. During normalization, the training and testing data partitions for a particular model scenario are combined. This approach ensures that the model parameters, trained on the normalized data, can generalize to unseen data effectively.

$$X_n = \frac{(X_{actual} - X_{min})}{(X_{max} - X_{min})} \quad (13)$$

where  $X_{actual}$ ,  $X_{max}$  and  $X_{min}$  denote the input data for actual, maximum, and minimum values of the target (i.e., soil moisture), respectively.

### 3.3.6. Hyperparameter Optimization

To construct the best forecasting model designs, the Hyperopt hyperparameter optimization algorithm, available in the Python Hyperopt library [46], is used to identify the optimal hyperparameters for the target and all other benchmark models for each lead time forecast separately. The training data partitions are used in this process. In comparison to Grid search and Random search, the Hyperopt hyperparameter optimization technique performs better since it can speed up the model training process while improving model accuracy.

The list of hyperparameters and their search space used in hyperparameter optimization processes are provided in Table 2. The optimal hyperparameters identified through the hyperparameter optimization process for designing the target LSTM and all other benchmark model architectures are detailed in Table 3.

**Table 2.** List of hyperparameters and their search space used in hyperparameter optimization process. ReLU and Adam stand for Rectified Linear Units and Adaptive Moment Estimation respectively.

Model	Name of Model Hyperparameters	Search Space for Optimal Hyperparameters
LSTM	LSTM Layer 1	[50, 70, 100, 150]
	LSTM Layer 2	[50, 70, 100, 150]
	LSTM Layer 3	[50, 70, 100, 150]
	Dense Layer	[1]
	Epochs	[100, 200, 500]
	Activation Function	[ReLU]
	Optimizer	[Adam]
	Dropout Ratio	[0.1, 0.2]
	Batch Size	[5,10,20,30]
DNN	Hidden neuron 1	[10, 20, 30]
	Hidden neuron 2	[10, 15, 25]
	Hidden neuron 3	[5, 10, 20]
	Dense Layer	[1]
	Epochs	[30, 50, 100, 200]
	Activation Function	[ReLU]
	Optimizer	[Adam]
	Dropout Ratio	[0.1, 0.2, 0.3, 0.4, 0.5]
	Batch Size	[3, 5, 10]
ANN	Hidden neuron	[10, 20, 30]
	Dense Layer	[1]
	Epochs	[30, 50, 100, 300, 1000, 2000]
	Activation Function	[sigmoid, tanh, ReLU]
	Optimizer	[Adam]
	Dropout Ratio	[0.3, 0.4, 0.5]
	Batch Size	[3,5,10]

**Table 3.** List of optimal hyperparameters selected by hyperparameter optimization process for LSTM, DNN and ANN models designing at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead times.

Time	Model	Layer 1 Neuron	Activation	Dropout	Layer 2 Neurons	Activation	Dropout	Layer 3 Neurons	Activation	Dropout	Batch Size	Epochs
$t + 1$	MoDWT-Lasso-LSTM	50	ReLU	0.1	150	ReLU	0.1	50	ReLU	0.1	20	500
	MoDWT-Lasso-DNN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	MoDWT-Lasso-ANN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	Lasso-LSTM	50	ReLU	0.1	150	ReLU	0.1	50	ReLU	0.1	20	500
	Lasso-DNN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	Lasso-ANN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	LSTM	50	ReLU	0.1	150	ReLU	0.1	50	ReLU	0.1	30	500
	DNN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	ANN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
	$t + 14$	MoDWT-Lasso-LSTM	100	ReLU	0.3	150	ReLU	0.2	100	ReLU	0.1	10
MoDWT-Lasso-DNN		20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
MoDWT-Lasso-ANN		20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
Lasso-LSTM		100	ReLU	0.3	150	ReLU	0.1	50	ReLU	0.1	30	500
Lasso-DNN		20	ReLU	0.4	10	ReLU	0.1	5	ReLU	0.1	10	100
Lasso-ANN		20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
LSTM		50	ReLU	0.1	100	ReLU	0.2	50	ReLU	0.1	10	200
DNN		20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
ANN		30	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	10	100
$t + 30$		MoDWT-Lasso-LSTM	50	ReLU	0.2	100	ReLU	0.2	50	ReLU	0.1	10
	MoDWT-Lasso-DNN	30	ReLU	0.5	20	ReLU	0.2	10	ReLU	0.1	5	300
	MoDWT-Lasso-ANN	20	ReLU	0.3	10	ReLU	0.2	10	ReLU	0.1	5	300
	Lasso-LSTM	50	ReLU	0.2	100	ReLU	0.2	50	ReLU	0.1	10	200
	Lasso-DNN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.1	5	300
	Lasso-ANN	10	ReLU	0.2	10	ReLU	0.1	5	ReLU	0.1	5	300
	LSTM	50	ReLU	0.2	100	ReLU	0.2	50	ReLU	0.1	10	200
	DNN	10	ReLU	0.5	25	ReLU	0.1	5	ReLU	0.3	3	300
	ANN	20	ReLU	0.3	10	ReLU	0.1	5	ReLU	0.3	3	300

### 3.3.7. Data Partitioning and Data Feeding to Models

In this study, for all model scenarios, the first 75% of the respective data is allocated for training purposes, while the remaining 25% is used for testing purposes. This allocation allows both training and the testing data partitions to acquire adequate data for successful model running. Although a total of 5844 data points were initially considered, the number of data points finally utilized at the model running stages for each lead time scenario is reduced due to the above explained data pre-processing steps, including data lagging, decomposition, and removal.

In the case of  $t + 1$  lead time SM forecasting, all models are fed with 5023 data points, while in case of  $t + 14$  and  $t + 30$  lead time SM forecasting, 5010 and 4994 data points are fed to the forecasting models. For  $t + 1$ ,  $t + 14$  and  $t + 30$  forecasting scenarios, 3767, 3757 and 3745 data points are used in the training phase, respectively, as 75% of the total data set is used for training. Thus, in  $t + 1$ ,  $t + 14$  and  $t + 30$  forecasting scenarios, 1256, 1253 and 1249 data points (the last 25%) remain for testing. In  $t + 1$  lead time case, for instance, daily data points from 1 April 2007 to 23 July 2017 are used for training, while daily data points from 24 July 2017 to 30 December 2007 are used for testing.

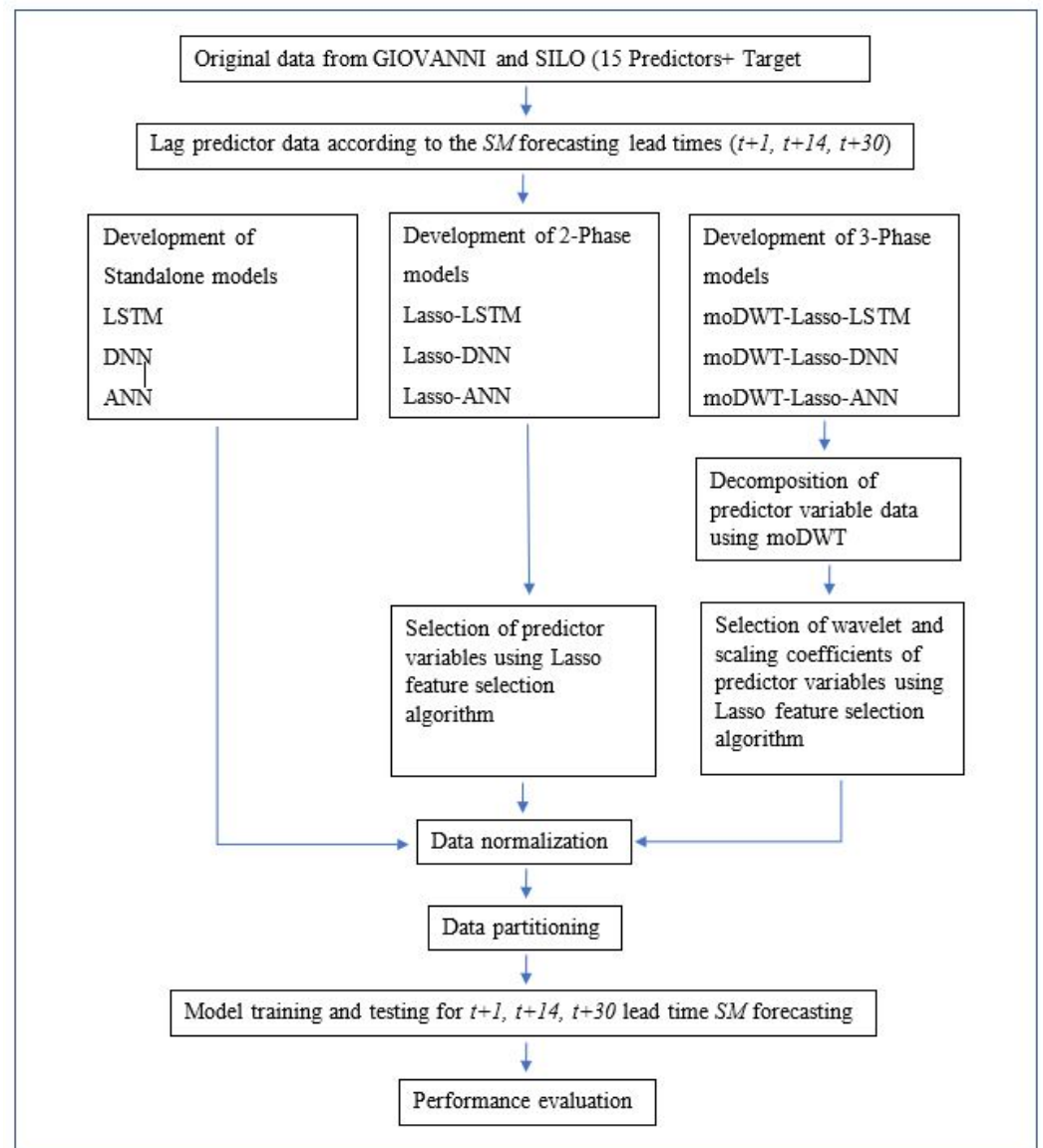
The standalone models, namely ANN, DNN, and LSTM for each lead time, are trained and tested using original un-decomposed lagged data of predictor variables and target variable data. In order to develop 2 phase hybrid models, namely Lasso-ANN, Lasso-DNN and Lasso-LSTM, un-decomposed lagged data of predictor variables selected by Lasso feature selection algorithm are used with target variable data.

In order to develop 3-phase hybrid models, namely moDWT-Lasso-ANN, moDWT-Lasso-DNN, and moDWT-Lasso-LSTM, lagged decomposed data of predictor variables are combined with un-decomposed target variable data. In the training phase of all model development cases, the model can see both input and output variable data. During the testing phase, however, the model can see only the input variable data and has no access the target variable data in the forecasting process. As the testing phase time point range is also historical with respect to the current time, realistically, future data of target variable with respect to all testing phase time points are available.

For setting up a situation exactly similar to the real-world application of the model, target variable data are not made available for the forecasting process and instead let the model to forecast values for the target variable for each lead time with respect to each testing phase time point using the respective historical data of input variables using the skills developed in the training phase. The forecasted values of target variable are then

compared with real future values of target variable available for all testing phase time points and evaluated the accuracy using statistical and graphical tools.

Figure 2 illustrates the schematic view of the all model development process including the 3-phase hybrid moDWT-Lasso-LSTM model for multi-step SM forecasting at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead times.



**Figure 2.** Schematic view of the development of benchmark models and proposed 3-phase hybrid moDWT-Lasso-LSTM model for multi-step SM forecasting at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead times.

### 3.4. Model Performance Evaluation

For ML models development, evaluating the models performance is crucial component. It determines whether a model is suitable for certain applications, compares it with rival models, and identifies areas for improvement [47]. As a result, for SM forecasting at selected sites for the same datasets, the proposed moDWT-Lasso-LSTM model and other benchmark models are evaluated considering forecasting accuracy and errors.

#### Pearson's Correlation Coefficient ( $r$ )

Equation (14) is used to derive the value of  $r$ , which expresses how closely forecasted ( $SM^{FOR}$ ) and observed ( $SM^{OBS}$ ) values are coincided. The values given for this metric are always floating in between  $-1$  to  $+1$  and it equals  $+1$  when perfectly strong and positive correlation exist between two variables (such as the forecasted and observed SM). In

contrast, perfectly strong and negative correlations exist between two variables gives value of  $-1$ . The value  $r$  will be equal to zero if there is no relation between any two variables. However, in this instance, there should be a high and positive correlation between the estimated values by the forecasting model and observed values to consider the forecasting model to be competent enough in prediction works, thus  $r$  value should close or equal to  $+1$ .

$$r = \frac{\sum_{i=1}^n (SM^{OBS} - \langle SM^{OBS} \rangle)(SM^{FOR} - \langle SM^{FOR} \rangle)}{\sqrt{\sum_{i=1}^n (SM^{OBS} - \langle SM^{OBS} \rangle)^2} \sqrt{\sum_{i=1}^n (SM^{FOR} - \langle SM^{FOR} \rangle)^2}} \quad (14)$$

#### Determination of Coefficient ( $R^2$ )

The determination of coefficient  $R^2$  can be explained as the proportion of the variance in the dependent variable that is predicted by the independent variables [48] and it ranges between  $-\infty$  and  $+1$ .  $+1$  is considered as the best value.

$$R^2 = 1 - \frac{\sum_{i=1}^N (SM^{FOR,i} - SM^{OBS,i})^2}{\sum_{i=1}^N (\langle SM^{OBS} \rangle - SM^{OBS,i})^2} \quad (15)$$

#### Root Mean Square Error (RMSE; $\text{kgm}^{-2}$ )

Regression model performances are typically evaluated using the RMSE (Equation (16)). This metric computes the average of prediction error generated by forecasting models, that is the average difference among the forecasted value  $SM^{FOR}$  and the observed value  $SM^{OBS}$ . The value of RMSE can be anywhere between 0 and  $\infty$ , but as model performance increases, the value of RMSE is shifting towards zero.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SM^{FOR} - SM^{OBS})^2} \quad (16)$$

#### Mean Absolute Error (MAE; $\text{kgm}^{-2}$ )

The MAE (Equation (17)) measures the actual forecast errors in relation to the total number of observations; MAE value is expected to fluctuate between 0 and  $\infty$ , however for ideal predictive models, it becomes zero. As the value given for MAE is unaffected by extreme outliers it provides a more reliable estimation of the model's average error relative to the RMSE.

$$MAE = \frac{1}{n} \sum_{i=1}^n |SM^{FOR} - SM^{OBS}| \quad (17)$$

#### Mean Absolute Scaled Error (MASE)

The MASE value (Equation (18)) proposed by Hyndman and Koehler [49] is also used to assess the forecast model's accuracy. The major advantage of this statistical tool is that, the result is independent of the scale of the data. This measures the accuracy of a forecasting model in terms of the in-sample MAE value generated by one period a head naïve forecast method. When the forecasting model performance is better than the average one-step, naïve forecast computed in-sample, the value for MASE will be less than 1 and contrarily, it is greater than 1 if the forecast is inferior to the in-sample average one-step, naïve forecast [49].

$$MASE = \frac{1}{N} \left( \frac{\sum_{i=1}^N |SM^{FOR} - SM^{OBS}|}{\frac{1}{N-m} \sum_{i=m+1}^N |SM^{OBS} - SM^{OBS,i-m}|} \right) \quad (18)$$

#### Symmetric Mean Absolute Percentage Error (SMAPE)

The SMAPE (Equation (19)) is a modification of Mean Absolute Percentage Error (MAPE) to avoid the issue of being infinite or undefined due to zeros in the denominator. Like MASE, SMAPE is also a scale-independent metrics and thus ideal for comparing performances of forecasting algorithms [49]. Smaller percentage values indicate high levels of accuracy in the forecasting models.

$$SMAPE = \frac{200}{N} \sum_{i=1}^N \frac{|SM^{FOR,i} - SM^{OBS,i}|}{SM^{FOR,i} + |SM^{OBS,i}|}, \% \quad (19)$$

#### Willmott's Index (WI)

The WI (Equation (20)) is widely used as a flexible and logical metric with its value ranging from 0 (poor model) to 1 (an ideal model).

$$WI = 1 - \frac{\sum_{i=1}^n (SM^{OBS} - SM^{FOR})^2}{\sum_{i=1}^n (|SM^{FOR} - \langle SM^{OBS} \rangle| + |SM^{OBS} - \langle SM^{OBS} \rangle|)^2} \quad (20)$$

#### Nash-Sutcliffe Index (NS)

The NS (Equation (21)) [50] shows how closely a depicted line between the predicted and observed values fit a 1:1 ratio. If the predicted and observed data match exactly, NS equals 1 so here, the metric lies between  $-\infty < NS \leq 0$  [50].

$$NS = 1 - \frac{\sum_{i=1}^n (SM^{OBS} - SM^{FOR})^2}{\sum_{i=1}^n (SM^{OBS} - \langle SM^{OBS} \rangle)^2} \quad (21)$$

#### Legate and McCabe Index (LM)

The LM (Equation (22)) is considered as a more advanced metric compared to WI and NS. When assessing the quality of model's fit to observed data, this index offers advantages over the correlation based metrics such as WI,  $R^2$  and NS whereby an optimal predictive model is expected to generate a value of 1 over the range  $-\infty$  and 1.

$$LM = 1 - \frac{\sum_{i=1}^n |SM^{OBS} - SM^{FOR}|}{\sum_{i=1}^n |SM^{OBS} - \langle SM^{OBS} \rangle|} \quad (22)$$

It is noteworthy that in Equations (14)–(22),  $SM^{OBS}$  refers to the daily observed soil moisture (0–10 cm depth) and  $SM^{FOR}$  is the daily forecasted soil moisture (0–10 cm depth) whereas  $\langle SM^{OBS} \rangle$  and  $\langle SM^{FOR} \rangle$  are the mean values of  $SM^{OBS}$  and  $SM^{FOR}$  respectively,  $i$  is instance of data point and  $N$  is the number of data points in the testing phase.

## 4. Results and Discussion

The summary of the descriptive statistics for all predictor and target variables is presented in Table 4. Refer to Goos and Meintrup [51] for the details of the calculations and interpretations of these statistics. These descriptive statistics provide information on the central tendency (mean, median) and variability (standard deviation) of the data set, as well as the shape and frequency of data distribution. The mean and median values of variables such as radiation, rh-tmax, ET, mslp, and GWS are close together, indicating that their distributions are nearly symmetric. Other variables have greater differences in the mean and median values, reflecting skewness in their distributions. Specifically, SM, SM10-40, SM100-200, SM40-100 and rain exhibit slightly to moderately right-skewed distributions, indicated by the skewness values just below and above 0.5, respectively. By contrast, the distributions of max-temp, min-temp, rh-tmin, ST40-100, ST10-40, and ST0-10 are slightly left-skewed, indicated by the skewness values close to  $-0.5$ .

In terms of the shape of a distribution's tails and peak, the negative kurtosis values indicate that generally most of the variables in this study are platykurtic, meaning they have thinner tails and flatter peaks compared to a normal distribution. The exceptions are rh-tmax and rh-tmin with positive kurtosis values, which are leptokurtic, indicating fatter tails and sharper peaks. Rain is extremely leptokurtic, suggesting a high frequency of extreme values.

Table 5 summarizes the results of trial-and-error to identify the best decomposition level and wavelet filter combination. In most cases, best suited combinations differ from one another, with the exception of moDWT-Lasso-LSTM and moDWT-Lasso-DNN at  $t + 1$ .

The best model forecasts are obtained using decomposition level 4 and wavelet filter “haar” for those two cases.

**Table 4.** Summary of descriptive statistics values of all predictors and target variable data.

Variable	Mean	Median	Standard Deviation	Skewness	Kurtosis
SM	22.6082	21.5707	3.8999	0.3591	−1.2382
max-temp	27.3613	27.7000	3.5232	−0.3439	−0.1983
min-temp	16.6750	17.3000	4.7370	−0.4436	−0.5403
radiation	18.5636	18.7000	5.9020	−0.2629	−0.6063
rh-tmax	50.4473	50.6000	11.8674	−0.0096	1.1172
rh-tmin	91.5993	96.0000	11.5819	−2.1140	5.6520
ET	3.9792	3.9000	1.3571	0.1223	−0.8611
mslp	1017.5165	1017.7000	4.9808	−0.2450	−0.1818
ST40-100	300.7142	301.8777	4.6184	−0.3690	−1.3350
ST10-40	300.8481	302.3685	5.2255	−0.3879	−1.2737
ST0-10	300.8051	302.5193	5.8458	−0.3975	−1.2068
rain	2.6492	0.0000	10.9918	9.3525	132.1285
SM10-40	80.7710	78.4010	10.3539	0.2873	−1.3730
SM100-200	253.8943	248.9959	19.3150	0.5340	−0.9635
SM40-100	150.0634	145.4409	21.7223	0.3255	−1.3311
GWS	939.5838	939.7291	14.6149	0.0610	−0.3491

**Table 5.** Summary of best decomposition levels and wavelet filters resulted from trial-and-error process for 3-phase hybrid models at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead times.

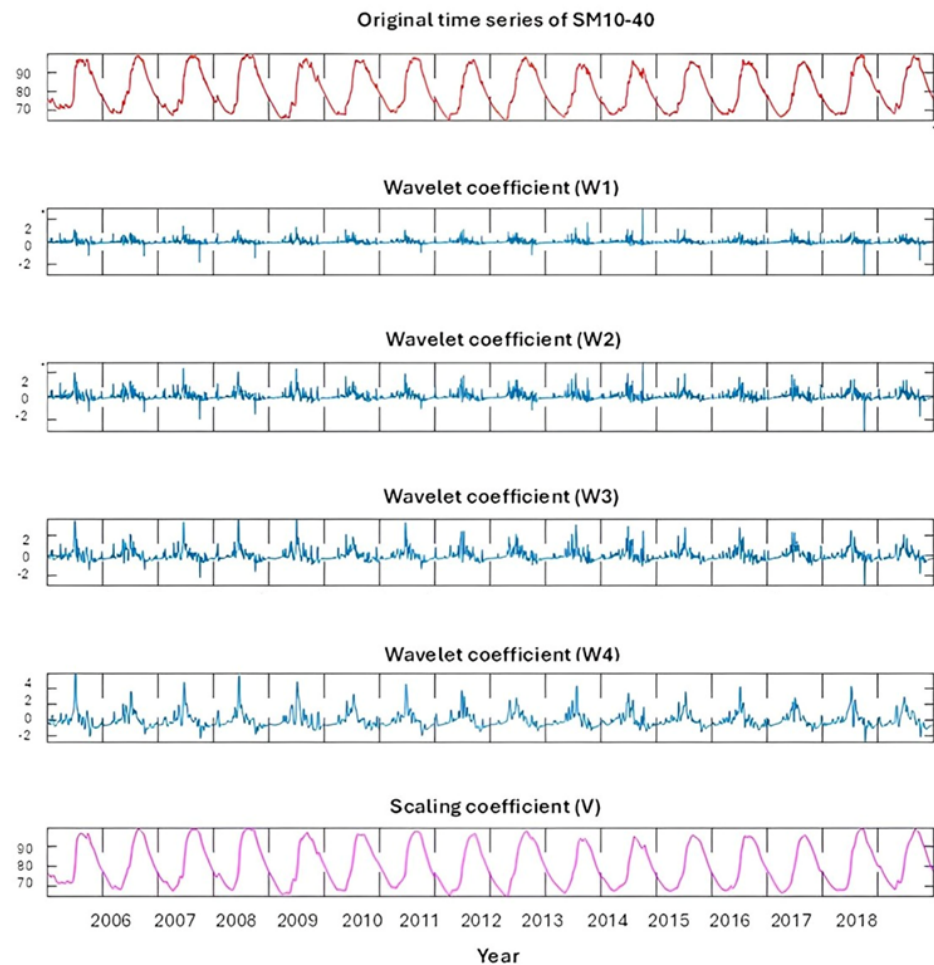
Model	$t + 1$		$t + 14$		$t + 30$	
	Decomposition Level	Filter	Decomposition Level	Filter	Decomposition Level	Filter
moDWT-Lasso-LSTM	4	haar	4	fk4	2	fk4
moDWT-Lasso-DNN	4	haar	4	db4	4	haar
moDWT-Lasso-ANN	2	haar	4	db6	4	db4

According to decomposition level 4, each predictor variable’s time series data is split into four wavelet coefficients and one scaling coefficient, regardless of the wavelet filter used. Figure 3 illustrates the decomposition results for SM10-40 based on decomposition level 4 and wavelet filter “haar”. (Notably, this is the decomposition level and wavelet filter combinations confirm the best model performances in moDWT-Lasso-LSTM and moDWT-Lasso-DNN at  $t + 1$  lead time). When decomposition level 4 is used, the number of predictor variables increased to 75 (60 wavelet coefficients ( $15 \times 4$ ) + 15 scaling coefficients ( $15 \times 1$ )). When decomposition level 2 is used for data decomposition, total number of predictor variables is increased up to 45 ( $=15 \times 2 + 15 \times 1$ ).

The Lasso feature selection algorithm is employed to identify the predictor variables most correlated with the target variable (SM). This algorithm reduces the number of wavelet and scaling coefficient data series used in the forecasting model training and testing. Different wavelet and scaling coefficient data series are selected by the Lasso algorithm for each decomposed data set derived from different combinations of decomposition levels and wavelet filters used in 3-phase model development. Notably, the majority of the coefficient data series selected by the Lasso algorithm are scaling coefficients of predictor variables.

Table 6 summarizes the wavelet and scaling coefficient data series selected by the Lasso feature selection algorithm for all 3-phase hybrid models at  $t + 1$  lead time. For the predictor variable SM10-40, the second, third, and fourth wavelet coefficient data series ( $W2$ ,  $W3$ ,  $W4$ ) and the scaling coefficient data series ( $V$ ) depicted in Figure 3 are selected by the Lasso algorithm for the development of the moDWT-Lasso-LSTM and moDWT-Lasso-DNN models at  $t + 1$  lead time.





**Figure 3.** Wavelet and scaling data series resulted from moDWT decomposition process given for the predictor variable: SM10-40 when decomposition level 4 and wavelet filter “haar” is used at  $t + 1$  lead time.

In developing 2-phase hybrid models (i.e., Lasso-LSTM, Lasso-DNN, and Lasso-ANN), the number of predictor variables selected by the Lasso algorithm for  $t + 1$ ,  $t + 14$ , and  $t + 30$  lead times are as follows:

- For  $t + 1$  lead time: 10 predictor variables (min-temp, radiation, rh-tmax, rh-tmin, ST0-10, rh-tmax, rh-tmin, mslp, ST40-100, ST0-10, rain, SM10-40, SM40-100, SM100-200, and GWS).
- For  $t + 14$  and  $t + 30$  lead times: 12 predictor variables (max-temp, min-temp, radiation, rh-tmin, mslp, ST40-100, ST0-10, rain, SM10-40, SM100-200, SM40-100, and GWS).

Table 7 displays the calculated values of statistical metrics to evaluate the performance of the proposed deep moDWT-Lasso-LSTM model and other benchmark models. These metrics include Pearson’s Correlation Coefficient ( $r$ ), Coefficient of Determination ( $R^2$ ), Root Mean Squared Error ( $RMSE$ ;  $\text{kgm}^{-2}$ ), Mean Absolute Error ( $MAE$ ;  $\text{kgm}^{-2}$ ), Mean Absolute Scaled Error ( $MASE$ ), Symmetric Mean Absolute Percentage Error ( $SMAPE$ ), Legates and McCabe Index ( $LM$ ) and Willmott’s Index ( $WI$ ). In general, the moDWT-Lasso-LSTM model demonstrated superior performance in SM forecasting compared to all the benchmark models in all different lead times. The proposed model has produced the highest values for  $r$ ,  $R^2$ ,  $LM$  and  $WI$ , indicating strong model performance and reliability.

Simultaneously, it has produced the lowest values for *RMSE*, *MAE*, *MASE* and *SMAPE*, reflecting high accuracy and minimal error.

**Table 6.** Summary of selected wavelet and scaling coefficients by Lasso feature selection technique at  $t + 1$  lead time for the 3-phase hybrid moDWT-Lasso-LSTM, moDWT-Lasso-DNN and moDWT-Lasso ANN methods utilising LSTM, DNN and ANN predictive models, respectively.

Model	Predictor	Wavelet Coefficient	Scaling Coefficients	Wavelets/Scaling Coefficient
moDWT-Lasso-LSTM	SM10-40	W2, W3, W4	min-temp	12
	SM100-200	W4	radiation	
	GWS	W4	ST0-10 rain SM10-40 SM100-200 GWS	
moDWT-Lasso-DNN	SM10-40	W2,W3,W4	min-temp	12
	SM100-200	W4	radiation	
	GWS	W4	ST0-10 rain SM10-40 SM100-200 GWS	
moDWT-Lasso ANN	rh-tmin	W2	min-temp	11
	SM10-40	W2	radiation rh-tmax ST0-10 rain SM10-40 SM100-200 SM40-100 GWS	

For the  $t + 1$  lead time, the moDWT-Lasso-LSTM model demonstrates superior performance across almost all metrics. It has the highest  $r \approx 0.97290$ , indicating a very strong correlation between predicted and actual SM values. Its  $R^2 \approx 0.92469$  shows that over 92% of the variance in SM is explained by the model, the highest among all models tested. The lowest values of  $RMSE \approx 0.97808$  and  $MAE \approx 0.76623$  signify more accurate predictions. The  $MASE \approx 4.39700$  and  $SMAPE \approx 3.48910\%$  are both lower than those of other models, further emphasizing the model’s accuracy. Additionally, the values of  $LM \approx 0.78021$  and  $WI \approx 0.98270$  are the highest, indicating superior model performance and agreement with observed data.

Comparatively, the moDWT-Lasso-DNN and moDWT-Lasso-ANN models also perform well, with high  $r$  and  $R^2$  values but slightly higher  $RMSE$ ,  $MAE$ ,  $MASE$ , and  $SMAPE$  values than the moDWT-Lasso-LSTM model. Traditional models such as Lasso-DNN and DNN show significantly lower performance, evidenced by lower  $r$  and  $R^2$  values and higher error metrics. For example, the standalone ANN model ( $r \approx 0.95478$ ) had the lowest correlation with the observed SM values, indicating that it was the least effective in capturing the relationship between the predictor variables and observed SM. This demonstrates the efficacy of the hybrid models, especially the moDWT-Lasso-LSTM, in short-term SM forecasting.

For the  $t + 14$  lead time, the moDWT-Lasso-LSTM model again outperforms other models, though the performance metrics show slight degradation compared to the  $t + 1$  lead time. It achieves an  $r \approx 0.96012$  and an  $R^2 \approx 0.89224$ , indicating strong predictive power. The  $RMSE \approx 1.18054$  and  $MAE \approx 0.96482$ , both are the lowest among all models, further highlighting accurate predictions. The  $MASE \approx 0.79649$  and  $SMAPE \approx 4.01170\%$ ,

which, although slightly higher than the  $t + 1$  lead time, remain the best among the models. The  $LM \approx 0.72280$  and  $WI \approx 0.97491$  also exhibit good model reliability and consistency.

**Table 7.** Values scored in the testing phase for statistical metrics used to evaluate the proposed hybrid moDWT-Lasso-LSTM and benchmark models for lead times  $t + 1$ ,  $t + 14$  and  $t + 30$ . The best values scored for relevant statistical metrics are boldfaced.

Model	$t + 1$							
	$r$	$R^2$	RMSE	MAE	MASE	SMAPE (%)	LM	WI
moDWT-Lasso-LSTM	<b>0.97290</b>	<b>0.92469</b>	<b>0.97808</b>	<b>0.76623</b>	<b>4.39700</b>	<b>3.48910</b>	<b>0.78021</b>	<b>0.98270</b>
moDWT-Lasso-DNN	0.97243	0.90801	1.05142	0.83664	4.80102	4.28050	0.76069	0.97023
moDWT-Lasso ANN	0.96755	0.87927	1.25829	0.99296	5.69808	4.32120	0.71597	0.97211
Lasso-LSTM	0.96916	0.86992	1.24185	0.99203	5.69274	4.29820	0.71543	0.97145
Lasso-DNN	0.96398	0.78780	1.49764	1.22490	7.02904	5.26880	0.64963	0.95672
Lasso-ANN	0.96310	0.86976	1.30536	1.02215	5.86556	4.45690	0.70762	0.96990
LSTM	0.96728	0.89932	1.08161	0.85262	4.89270	3.76450	0.75543	0.97789
DNN	0.96628	0.66048	1.70606	1.38637	7.95562	5.81720	0.60344	0.93937
ANN	0.95478	0.81781	1.58090	1.25067	7.17693	5.51210	0.62531	0.95712
$t + 14$								
	$r$	$R^2$	RMSE	MAE	MASE	SMAPE (%)	LM	WI
moDWT-Lasso-LSTM	0.96012	<b>0.89224</b>	1.18054	0.96482	0.79649	<b>4.01170</b>	0.72280	<b>0.97491</b>
moDWT-Lasso-DNN	<b>0.96149</b>	0.87398	<b>1.19683</b>	<b>0.94721</b>	<b>0.78195</b>	4.13590	<b>0.72846</b>	0.97264
moDWT-Lasso ANN	0.95139	0.85932	1.29359	1.06966	0.88304	4.67810	0.69336	0.96854
Lasso-LSTM	0.93999	0.87467	1.34597	1.05395	0.87006	4.30280	0.69719	0.96878
Lasso-DNN	0.95380	0.88453	1.20330	0.96490	0.79655	4.73540	0.72340	0.97344
Lasso-ANN	0.95167	0.85824	1.30455	1.06954	0.88293	4.65450	0.69340	0.96818
LSTM	0.94245	0.86700	1.36678	1.05309	0.86935	4.71900	0.69744	0.96750
DNN	0.95413	0.77293	1.48918	1.18204	0.97581	5.06000	0.66115	0.95493
ANN	0.93540	0.77400	1.59018	1.28029	1.05692	5.54800	0.63298	0.95122
$t + 30$								
	$r$	$R^2$	RMSE	MAE	MASE	SMAPE (%)	LM	WI
moDWT-Lasso-LSTM	<b>0.96497</b>	<b>0.91564</b>	<b>1.13674</b>	<b>0.91126</b>	<b>0.45417</b>	<b>3.98600</b>	<b>0.73774</b>	<b>0.97849</b>
moDWT-Lasso-DNN	0.95820	0.88818	1.15259	0.95784	0.47738	4.31100	0.72481	0.97516
moDWT-Lasso ANN	0.95528	0.88467	1.22855	1.00449	0.50063	4.44910	0.71140	0.97286
Lasso-LSTM	0.95051	0.88685	1.22393	0.96703	0.48196	4.22980	0.72169	0.97307
Lasso-DNN	0.95665	0.85161	1.26631	0.99443	0.49562	4.30100	0.71429	0.96852
Lasso-ANN	0.93237	0.81717	1.46481	1.22684	0.61145	5.34670	0.64752	0.95895
LSTM	0.95436	0.87888	1.20148	0.97581	0.48634	4.32890	0.71917	0.97277
DNN	0.95139	0.77771	1.47242	1.15331	0.57480	4.91300	0.66865	0.95562
ANN	0.93926	0.83699	1.40469	1.16230	0.57928	5.09100	0.66607	0.96294

The moDWT-Lasso-DNN model shows comparable performance but has slightly lower  $r$  and  $R^2$  values and higher error metrics. Other models, such as Lasso-LSTM and Lasso-DNN, demonstrate relatively good performance but still fall short of the moDWT-based models, highlighting the advantage of the wavelet transform in enhancing forecasting accuracy over a 14-day horizon. Specifically, the Lasso-LSTM model, with an  $r$  score of approximately 0.93999, performs somewhat lower compared to its moDWT-enhanced counterpart. Meanwhile, the ANN model shows the lowest performance for a 14-day lead time, with  $r \approx 0.93540$  and  $R^2 \approx 0.77400$ .

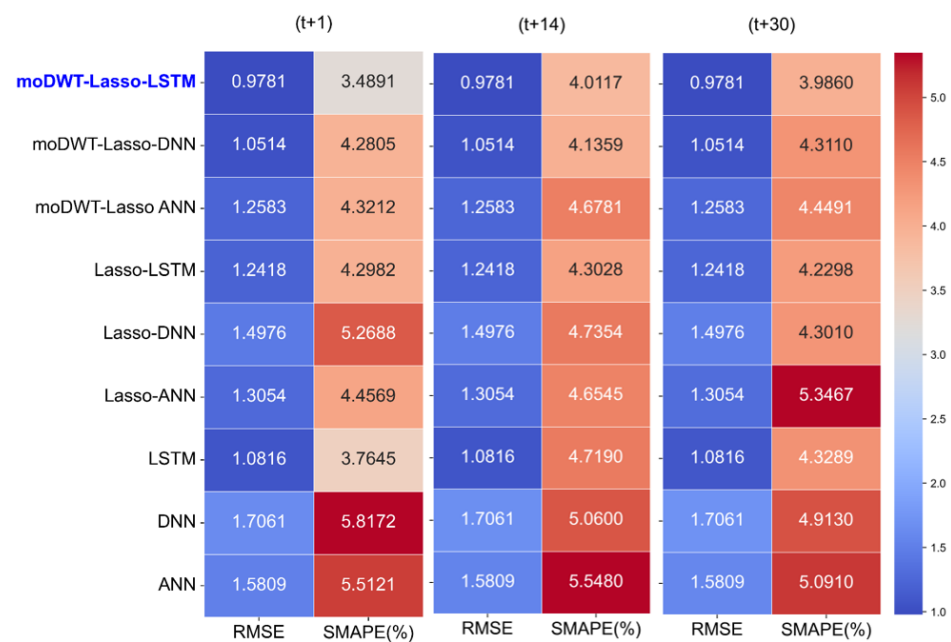
For the  $t + 30$  lead time, the moDWT-Lasso-LSTM model continues to show the best performance, with an  $r \approx 0.96497$  and an  $R^2 \approx 0.91564$ . Its values of  $RMSE \approx 1.13674$  and  $MAE \approx 0.91126$  are the lowest among all models, indicating accurate long-term predictions. The  $MASE \approx 0.45417$  and  $SMAPE \approx 3.98600\%$  are also the lowest. Additionally, the values of  $LM \approx 0.73774$  and  $WI \approx 0.97849$  reflect high model agreement and reliability even for extended forecasts.

Other models, including the moDWT-Lasso-DNN and moDWT-Lasso-ANN, perform well but with slightly higher error metrics and lower correlation coefficients compared to the moDWT-Lasso-LSTM. Traditional models such as DNN and ANN show significantly lower performance, with higher *RMSE*, *MAE*, *MASE*, and *SMAPE* values, indicating less accurate predictions. This further underscores the superiority of the proposed model in long-term SM forecasting.

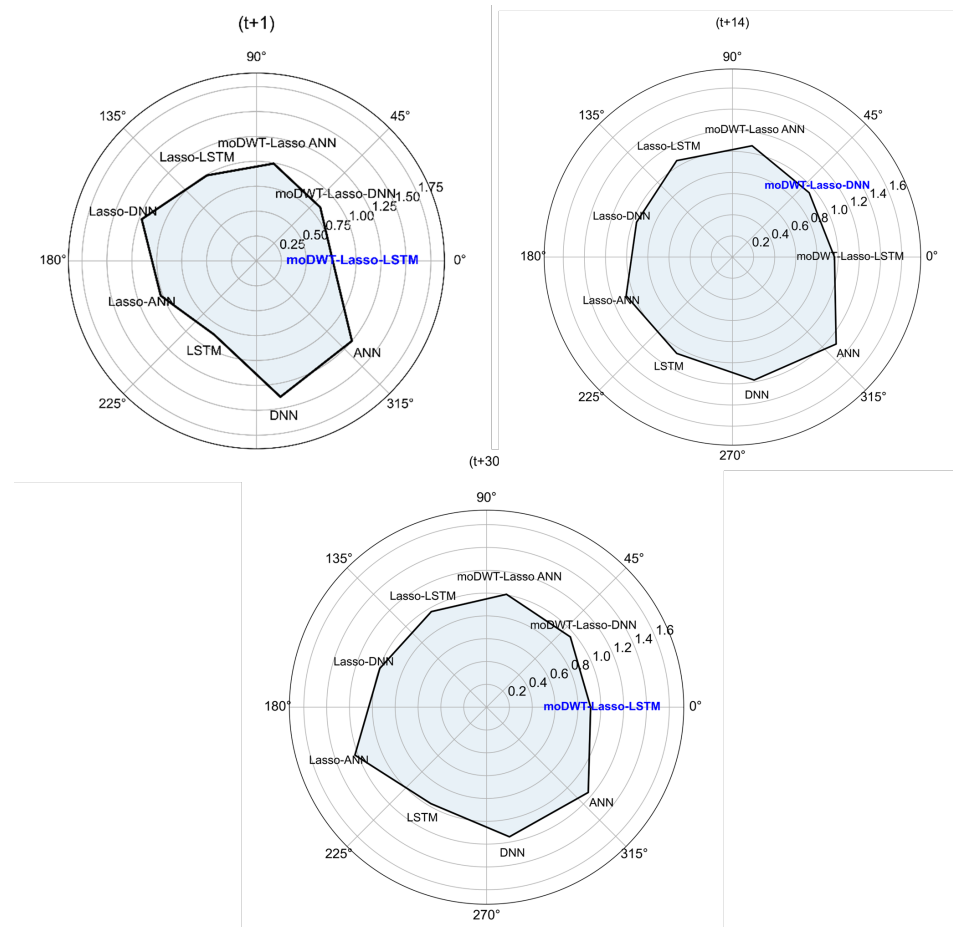
For instance, the heatmap (Figure 4) shows that the moDWT-Lasso-LSTM model consistently performs better in both RMSE and SMAPE across all leading times, suggesting its robustness in reducing relative error. On the other hand, the DNN and ANN models exhibit the highest SMAPE values, particularly at the  $t + 14$  and  $t + 30$  horizons, with values exceeding 5.5%. This indicates that these models may not handle longer-term predictions compared to others.

Moreover, Figure 5 plays three radar charts representing the MAE for all models across three forecasting horizons. The charts illustrate that the moDWT-Lasso-LSTM model is consistently closest to the center, indicating the lowest MAE and the best short-term ( $t + 1$ ) and long-term ( $t + 30$ ) prediction performance. At the mid-term ( $t + 14$ ), the moDWT-Lasso-LSTM continues to outperform the other models, with moDWT-Lasso-DNN showing slight improvement. These radar charts effectively highlight the comparative strength of each model across different forecasting horizons, emphasizing the superior performance of the hybrid moDWT-Lasso-LSTM model.

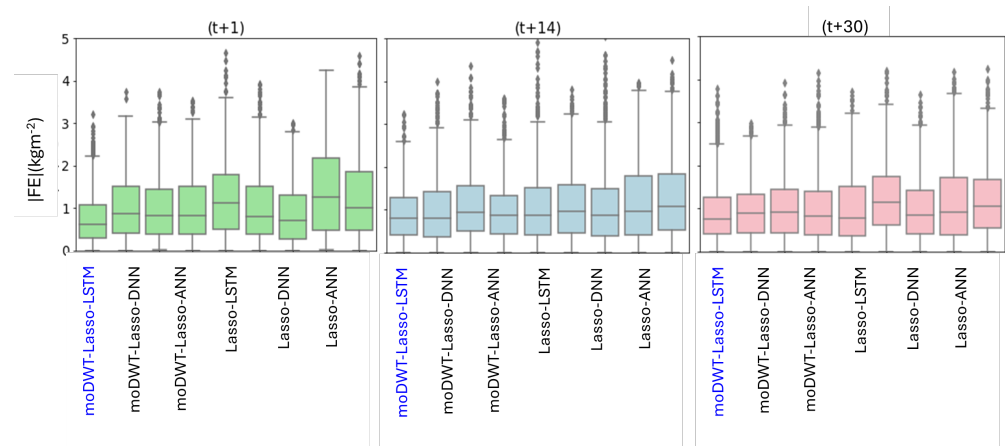
To further affirm the superiority of the proposed moDWT-Lasso-LSTM model in terms of prediction competency over the other benchmark models, the absolute forecasting errors (i.e.,  $|FE| = |\text{observed SM} - \text{forecasted SM}|$ ;  $\text{kgm}^{-2}$ ) of the proposed model and all benchmark models are compared (Figure 6). The distribution of  $|FE|$  during the testing phase, including the upper, median, and lower quartiles for each model for  $t + 1$ ,  $t + 14$ , and  $t + 30$  lead time SM forecasting, is illustrated in the box plots in Figure 6. According to these box plots, the multi-step moDWT-Lasso-LSTM model exhibits the fewest quartiles for  $|FE|$  across all lead times. These results indicate a narrow error distribution for the moDWT-Lasso-LSTM model compared to the benchmark models, further demonstrating its suitability for SM forecasting.



**Figure 4.** Heatmap of the Root Mean Square Error (RMSE) and Symmetric Mean Absolute Percentage Error (SMAPE) for the moDWT-Lasso-LSTM model and other benchmark models at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead time SM forecasting.



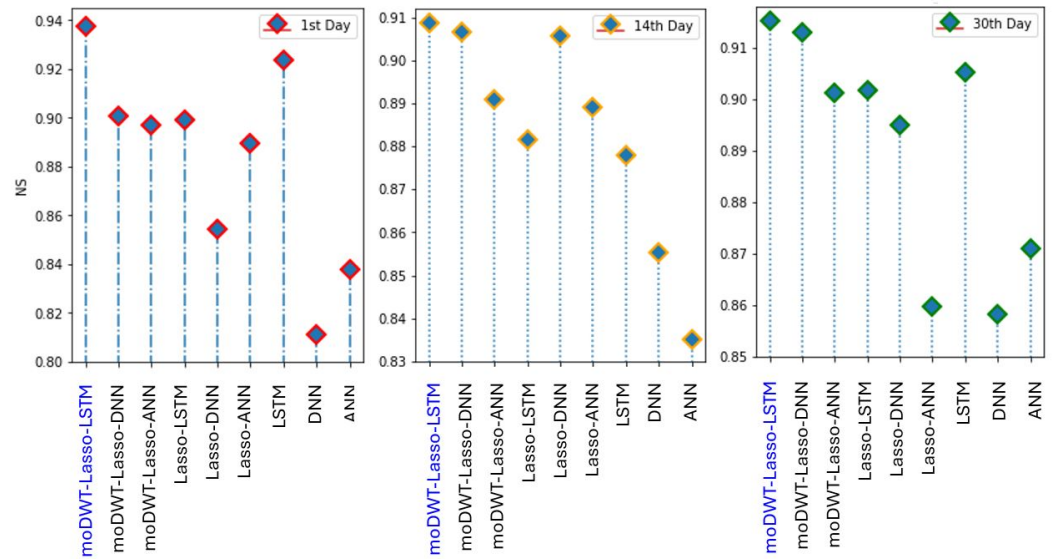
**Figure 5.** Radar plots for the Mean Absolute Error (MAE) for the moDWT-Lasso-LSTM model and other benchmark models at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead time SM forecasting.



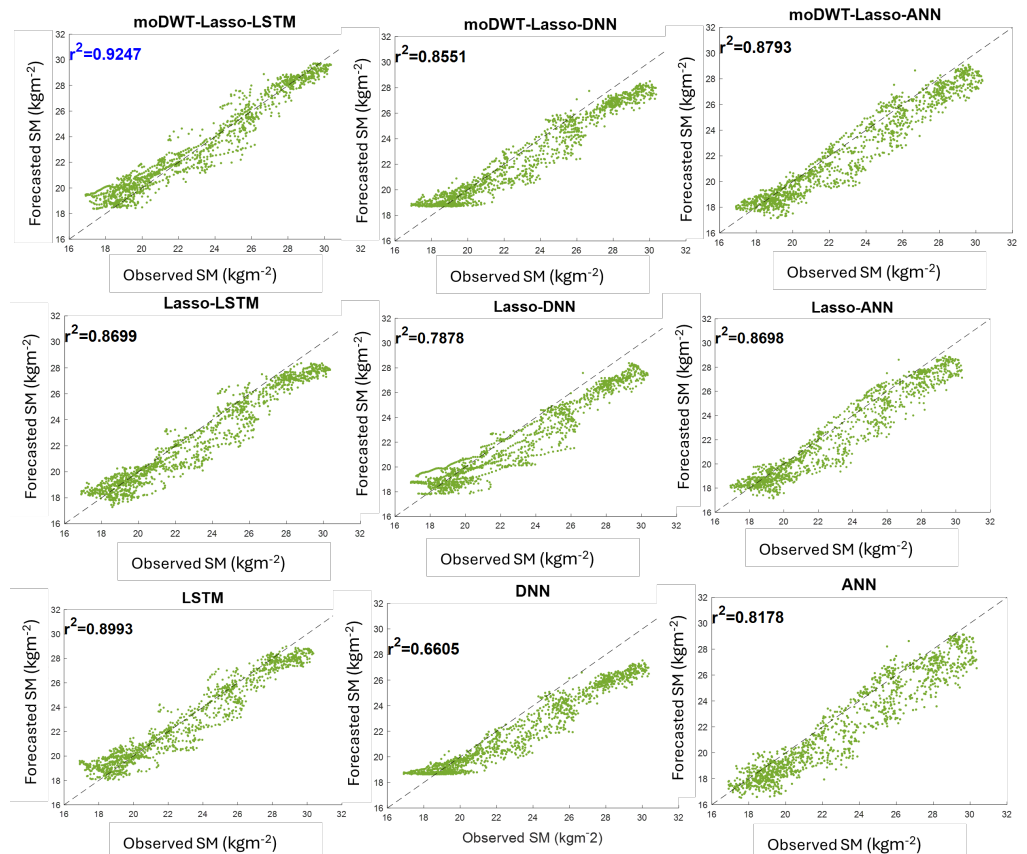
**Figure 6.** Box plot of forecast errors in the testing phase generated by the moDWT-Lasso-LSTM hybrid model and other benchmark models at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead time SM forecasting.

Figure 7 shows the stem plots for the Nash-Sutcliffe Coefficient ( $NS$ ) calculated for the target moDWT-Lasso-LSTM model and benchmark models during the testing phase for  $t + 1$ ,  $t + 14$ , and  $t + 30$  lead times SM forecasting. These graphs present that the moDWT-Lasso-LSTM model exhibits the highest values of  $NS$  for all lead times. Additionally, scatter plots for the  $t + 30$  lead time are provided for all the models tested (Figure 8). In comparison to the scatter plots of other forecasting models, data points for the moDWT-Lasso-LSTM model are more uniformly distributed along the 45-degree line, with fewer outliers and

deviations. This indicates a strongly positive correlation between observed and forecasted SM values for the moDWT-Lasso-LSTM model.



**Figure 7.** Stem plots of the Nash-Sutcliffe Coefficient (NS) for the hybrid moDWT-Lasso-LSTM model and the benchmark models in testing phase at  $t + 1$ ,  $t + 14$  and  $t + 30$  lead time SM forecasting.



**Figure 8.** Scatter plots of moDWT-Lasso-LSTM model and other benchmark models in testing phase at  $t+30$  lead time SM forecasting.

The results consistently show that the moDWT-Lasso-LSTM model outperforms all other benchmark models across different lead times, demonstrating its robustness and

reliability in SM forecasting. The use of the wavelet transform (moDWT) combined with Lasso feature selection and the LSTM model significantly enhances prediction accuracy. This hybrid approach effectively captures the temporal dependencies and intricate patterns in the data, leading to superior performance metrics. Furthermore, the decrease in model performance with increasing lead time is expected due to the complexity and variability of SM dynamics over longer periods. However, the moDWT-Lasso-LSTM model's ability to maintain relatively high accuracy and low error metrics even at  $t + 30$  lead time indicates its potential for practical applications in agricultural and environmental monitoring.

## 5. Conclusions and Future Work

Agricultural decision-making increasingly relies on reliable information regarding climatic and hydrological variables. For instance, farmers commonly use rainfall forecasts to make decisions about crop establishment, crop harvesting, fertilizer application, and land preparation activities. The proposed model is designed to forecast soil moisture (SM), which is a critical hydrological variable. Accurate SM information is essential for determining the need and timing for irrigation, as well as the precise quantity of irrigation water required on a particular day. If sufficient moisture is available in the soil, irrigation can be skipped; if moisture is inadequate, only the deficit should be compensated through irrigation. Reliable SM information greatly aids in such decision-making.

Furthermore, SM forecasts are crucial for fertilizer application. Adequate soil moisture is essential for dissolving nutrients in fertilizers and making them available to plants. This process enhances the plant's ability to absorb essential nutrients, maximizing fertilizer use efficiency while reducing waste. In areas without access to irrigation water, where farming relies entirely on rainfall, knowing the moisture levels in soils in advance can significantly impact activities like land preparation, planting, and fertilizer application. This is particularly relevant as many farmers are transitioning towards precision agriculture to reduce production costs, minimize waste, and conserve resources and inputs.

In this context, this study has developed a multi-step wavelet 3-phase hybrid deep learning SM forecasting (moDWT-Lasso-LSTM) model. This model employs Lasso regression optimization and moDWT decomposition algorithms to forecast SM in Bundaberg, Queensland, Australia. Daily input data from 1 January 2005 to 31 December 2020, were obtained from NASA's Global Land Data Assimilation System (GLDAS), the Land Data Assimilation System (FLDAS), and the ground database SILO. To achieve an accurate model, the extracted data were decomposed using moDWT, and features were selected using the Lasso algorithm for 1 ( $t + 1$ ), 14 ( $t + 14$ ), and 30 ( $t + 30$ ) days ahead forecasts. Incorporating LSTM, moDWT, and Lasso, the proposed deep learning multi-step moDWT-Lasso-LSTM hybrid model was created. Its performance was evaluated using statistical score measures and compared with eight other models: moDWT-Lasso-DNN, moDWT-Lasso-ANN, Lasso-LSTM, Lasso-DNN, Lasso-ANN, LSTM, DNN, and ANN.

In comparison to other benchmark models, The moDWT-Lasso-LSTM hybrid model demonstrates superior performance in forecasting SM across various lead times, with particularly notable improvements for  $t + 1$  and  $t + 30$ . According to the statistical metrics discussed in Table 7, the moDWT-Lasso-DNN model shows performance very similar to that of the moDWT-Lasso-LSTM model for the  $t + 14$  lead time. Visualizing the results using box plots of  $|FE|$  and stem plots of  $NS$ , the moDWT-Lasso-LSTM model consistently outperforms the moDWT-Lasso-DNN and all other benchmark models. Consequently, the moDWT-Lasso-LSTM model proves to be more effective in predicting SM than the other benchmark models.

When considering model complexity, the moDWT-Lasso-LSTM, while offering high accuracy, is relatively complex due to the use of deep learning techniques combined with wavelet based feature selection. These types of models require substantial computational resources and longer training times compared to simpler models. In contrast, models like Lasso-LSTM and Lasso-DNN, although still complex, have fewer parameters and generally shorter training times. On the other hand, standalone model such as LSTM and DNN

offer a balance between performance and complexity but still require considerable computational power compared to the basic models. Overall, while more complex models like the hybrid moDWT-Lasso variants provide improved accuracy, they come with increased computational costs and training times.

Despite the efficacy of the proposed model, the present research considers only 1-day, 14-day, and 30-day ahead SM forecasting. Therefore, the number of lead times used in this study may not impose a limitation on longer-term applications of the proposed model. However, it is important to note that increasing the lead time can potentially cause significant changes in model performance. Future researchers could conduct new studies to assess the forecasting capability of the model with extended lead times. Additionally, future research could develop an SM prediction tool to generate long-term forecasts, such as several months ahead, which could be significantly more important for irrigation, water resource management, and strategic planning than shorter-duration forecasts.

As this study uses satellite data, the inputs to the model are continuously recorded, enabling the methodology to be further implemented for operational use in agriculture and other industries with real-time access to historical input data. Unlike discrete wavelet methods used in earlier studies, the wavelet transform data decomposition procedure adopted here does not require future data to calculate the wavelet and scaling coefficients [23]. This advantage allows the proposed model to be practically implemented in real-time, utilizing accessible historical input data.

Due to time and resource constraints, this methodology has not been tested across the entire Queensland or Australian region. Therefore, it should be tested in other regions to examine the geographical consistency of the proposed model. Additionally, while this model was developed to forecast SM in the topsoil layer (0–10 cm depth), future researchers could examine the methodology's effectiveness in forecasting SM in deeper soil layers.

For methodological improvements over the present moDWT technique coupled with Lasso feature selection or LSTM model, future studies might adopt alternative decomposition methods such as the atrous (AT) algorithm [23] that can address issues related to using future data in model design. Moreover, the moDWT-Lasso-LSTM model could be applied to predict important drought indices such as the Palmer Drought Severity Index (PDSI), Standardized Precipitation Index (SPI), and Standardized Precipitation and Evaporation Index (SPEI), which are time-series methods where data splitting through multi-resolution analysis can be utilized.

**Author Contributions:** Conceptualization, W.J.M.L.P.J., and R.C.D.; methodology, W.J.M.L.P.J. and R.C.D.; software, W.J.M.L.P.J.; validation, Z.M.Y., W.J.M.L.P.J. and R.C.D.; formal analysis, Z.M.Y., W.J.M.L.P.J. and T.N.-H.; investigation, W.J.M.L.P.J.; resources, R.C.D.; data curation, W.J.M.L.P.J.; writing—original draft preparation, Z.M.Y., W.J.M.L.P.J. and T.N.-H.; writing—review and editing, Z.M.Y., R.C.D., N.R., S.G., T.N.-H. and A.G.; visualisation, W.J.M.L.P.J. and T.N.-H.; supervision, R.C.D., N.R., S.G. and A.G.; project administration, R.C.D.; funding acquisition, W.J.M.L.P.J. and R.C.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the University of Southern Queensland International PhD Scholarship and study leave grant from Wayamba University of Sri Lanka.

**Data Availability Statement:** The data presented in this study are available on request from the first author, excluding some data protected by copyright.

**Acknowledgments:** The authors are grateful to NASA and SILO for providing free access to GIOVANNI satellite and ground-based meteorological data. The authors would like to express their gratitude to the University of Southern Queensland (UniSQ), Australia, and the Wayamba University of Sri Lanka for generously funding this research. In addition, the authors would like to thank Barbara Harmes for proofreading this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.



## Appendix A

Table A1 shows the list of symbols used in this study.

**Table A1.** List of Symbols.

Symbol	Description
ReLU	Rectified Linear Unit
SM <sup>OBS</sup>	Observed Soil Moisture
SM <sup>FOR</sup>	Forecasted Soil Moisture
$r$	Pearson's Correlation Coefficient
$R^2$	Determination of Coefficient
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
SMPE	Symmetric Mean Absolute Percentage Error
WI	Willmott's Index
NS	Nash-Sutcliffe Index
LM	Legate and McCabe Index

Table A2 shows the list of prefixes used in this study.

**Table A2.** List of Prefixes.

Symbol	Description
ML	Machine Learning
SSM	Surface Soil Moisture
EWT	Empirical Wavelet Transform
SM	Soil Moisture
DL	Deep Learning
LSTM	Long Short-Term Memory
DWT	Discrete Wavelet Transformation
moDWT	Maximum Overlap Discrete Wavelet Transform
Lasso	Least Absolute Shrinkage and Selection Operator
FLDAS	Famine Early Warning Systems Network Land Data Assimilation System
GIOVANNI	Goddard Online Interactive Visualization and Analysis Infrastructure
GLDAS	Global Land Data Assimilation System
SILO	Scientific Information for Landowners
ANN	Artificial Neural Network
DNN	Deep Neural Network

## References

- Liao, R.; Yang, P.; Wang, Z.; Wu, W.; Ren, S. Development of a soil water movement model for the superabsorbent polymer application. *Soil Sci. Soc. Am. J.* **2018**, *82*, 436–446. [\[CrossRef\]](#)
- Chang, X.; Zhao, W.; Zeng, F. Crop evapotranspiration-based irrigation management during the growing season in the arid region of northwestern China. *Environ. Monit. Assess.* **2015**, *187*, 699. [\[CrossRef\]](#)
- Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos. Res.* **2017**, *197*, 42–63. [\[CrossRef\]](#)
- Silverman, D.; Dracup, J.A. Artificial neural networks and long-range precipitation prediction in California. *J. Appl. Meteorol.* **2000**, *39*, 57–66. [\[CrossRef\]](#)
- Khan, N.; Sachindra, D.; Shahid, S.; Ahmed, K.; Shiru, M.S.; Nawaz, N. Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.* **2020**, *139*, 103562. [\[CrossRef\]](#)
- Belayneh, A.; Adamowski, J.; Khalil, B.; Quilty, J. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmos. Res.* **2016**, *172*, 37–47. [\[CrossRef\]](#)
- Zhu, B.; Feng, Y.; Gong, D.; Jiang, S.; Zhao, L.; Cui, N. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* **2020**, *173*, 105430. [\[CrossRef\]](#)
- Huang, S.; Chang, J.; Huang, Q.; Chen, Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J. Hydrol.* **2014**, *511*, 764–775. [\[CrossRef\]](#)

9. Jamei, M.; Ali, M.; Karbasi, M.; Sharma, E.; Jamei, M.; Chu, X.; Yaseen, Z.M. A high dimensional features-based cascaded forward neural network coupled with MVMD and Boruta-GBDT for multi-step ahead forecasting of surface soil moisture. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105895. [[CrossRef](#)]
10. Basak, A.; Schmidt, K.M.; Mengshoel, O.J. From data to interpretable models: Machine learning for soil moisture forecasting. *Int. J. Data Sci. Anal.* **2023**, *15*, 9–32. [[CrossRef](#)]
11. Emmert-Streib, F.; Yang, Z.; Feng, H.; Tripathi, S.; Dehmer, M. An introductory review of deep learning for prediction models with big data. *Front. Artif. Intell.* **2020**, *3*, 4. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2018**, *561*, 918–929. [[CrossRef](#)]
13. Gauch, M.; Kratzert, F.; Klotz, D.; Nearing, G.; Lin, J.; Hochreiter, S. Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 2045–2062. [[CrossRef](#)]
14. Elsaadani, M.; Habib, E.; Abdelhameed, A.M.; Bayoumi, M. Assessment of a Spatiotemporal Deep Learning Approach for Soil Moisture Prediction and Filling the Gaps in Between Soil Moisture Observations. *Front. Artif. Intell.* **2021**, *4*, 636234. [[CrossRef](#)] [[PubMed](#)]
15. Li, Q.; Li, Z.; Shangguan, W.; Wang, X.; Li, L.; Yu, F. Improving soil moisture prediction using a novel encoder-decoder model with residual learning. *Comput. Electron. Agric.* **2022**, *195*, 106816. [[CrossRef](#)]
16. Suebsombut, P.; Sekhari, A.; Sureephong, P.; Belhi, A.; Bouras, A. Field data forecasting using LSTM and bi-LSTM approaches. *Appl. Sci.* **2021**, *11*, 11820. [[CrossRef](#)]
17. Zeynoddin, M.; Bonakdari, H. Structural-optimized sequential deep learning methods for surface soil moisture forecasting, case study Quebec, Canada. *Neural Comput. Appl.* **2022**, *34*, 19895–19921. [[CrossRef](#)]
18. Liu, Y.; Yue, Q.; Wang, Q.; Yu, J.; Zheng, Y.; Yao, X.; Xu, S. A framework for actual evapotranspiration assessment and projection based on meteorological, vegetation and hydrological remote sensing products. *Remote Sens.* **2021**, *13*, 3643. [[CrossRef](#)]
19. Wu, M. Wind speed forecasting by spatial-temporal data-driven models using atmospheric input variables. *Ocean Eng.* **2024**, *308*, 118191. [[CrossRef](#)]
20. Nikdad, P.; Mohammadi Ghaleni, M.; Moghaddasi, M.; Pradhan, B. Enhancing a machine learning model for predicting agricultural drought through feature selection techniques. *Appl. Water Sci.* **2024**, *14*, 125. [[CrossRef](#)]
21. Alizadeh, Z.; Shourian, M.; Yaseen, Z.M. Simulating monthly streamflow using a hybrid feature selection approach integrated with an intelligence model. *Hydrol. Sci. J.* **2020**, *65*, 1374–1384. [[CrossRef](#)]
22. Chu, H.; Wei, J.; Wu, W. Streamflow prediction using LASSO-FCM-DBN approach based on hydro-meteorological condition classification. *J. Hydrol.* **2020**, *580*, 124253. [[CrossRef](#)]
23. Quilty, J.; Adamowski, J. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J. Hydrol.* **2018**, *563*, 336–353. [[CrossRef](#)]
24. Adib, A.; Zaerpour, A.; Lotfirad, M. On the reliability of a novel MODWT-based hybrid ARIMA-artificial intelligence approach to forecast daily snow depth (Case study: The western part of the Rocky Mountains in the USA). *Cold Reg. Sci. Technol.* **2021**, *189*, 103342. [[CrossRef](#)]
25. Yong, B.; Hong, Y.; Ren, L.L.; Gourley, J.J.; Huffman, G.J.; Chen, X.; Wang, W.; Khan, S.I. Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin. *J. Geophys. Res. Atmos.* **2012**, *117*, D09108. [[CrossRef](#)]
26. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Remote Sens. Environ.* **2018**, *212*, 176–198. [[CrossRef](#)]
27. Armstrong, J.S. *Long-Range Forecasting. From Crystal Ball to Computer*; John Wiley and Sons: New York, NY, USA, 1985; Volume 348, pp. 1–34.
28. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [[CrossRef](#)]
29. Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. *Soil Tillage Res.* **2018**, *181*, 63–81. [[CrossRef](#)]
30. Shirsath, P.B.; Singh, A.K. A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. *Water Resour. Manag.* **2010**, *24*, 1571–1581. [[CrossRef](#)]
31. Le, X.H.; Nguyen, D.H.; Jung, S.; Lee, G. Deep neural network-based discharge prediction for upstream hydrological stations: A comparative study. *Earth Sci. Inform.* **2023**, *16*, 3113–3124. [[CrossRef](#)]
32. El Bilali, A.; Abdeslam, T.; Ayoub, N.; Lamane, H.; Ezzaoui, M.A.; Elbeltagi, A. An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. *J. Environ. Manag.* **2023**, *327*, 116890. [[CrossRef](#)] [[PubMed](#)]
33. Al-Musaylh, M.S.; Deo, R.C.; Li, Y. Electrical energy demand forecasting model development and evaluation with maximum overlap discrete wavelet transform-online sequential extreme learning machines algorithms. *Energies* **2020**, *13*, 2307. [[CrossRef](#)]
34. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [[CrossRef](#)]

35. Karevan, Z.; Suykens, J. Spatio-temporal feature selection for black-box weather forecasting. In Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2016, Bruges, Belgium, 27–29 April 2016; pp. 611–616.
36. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
37. Chen, J.; Zeng, G.Q.; Zhou, W.; Du, W.; Lu, K.D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Convers. Manag.* **2018**, *165*, 681–695. [[CrossRef](#)]
38. Zhang, X.; Zhang, Q.; Zhang, G.; Nie, Z.; Gui, Z.; Que, H. A Novel Hybrid Data-Driven Model for Daily Land Surface Temperature Forecasting Using Long Short-Term Memory Neural Network Based on Ensemble Empirical Mode Decomposition. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1032. [[CrossRef](#)]
39. Spark, W. Weather Spark. Available online: <https://weatherspark.com> (accessed on 11 August 2024).
40. Queensland Government. *Climate Change in the Wide Bay-Burnett Region*; Department of Energy and Climate: London, UK, 2023.
41. PINTEREST. Land Use Map of Queensland. Available online: <https://www.pinterest.com/pin/land-use-map-of-queensland--510877151477439946> (accessed on 11 August 2024).
42. Teng, W.; Rui, H.; Vollmer, B.; de Jeu, R.; Fang, F.; Lei, G.D.; Parinussa, R. NASA Giovanni: A Tool for Visualizing, Analyzing, and Intercomparing Soil Moisture Data. In *Remote Sensing of the Terrestrial Water Cycle*; Wiley Online Library: Hoboken, NJ, USA, 2014; pp. 331–346.
43. Morshed, A.; Aryal, J.; Dutta, R. Environmental spatio-temporal ontology for the Linked open data cloud. In Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, Melbourne, VIC, Australia, 16–18 July 2013; pp. 1907–1912.
44. Ketkar, N. *Introduction to Keras*; Deep learning with Python; Springer: Berlin/Heidelberg, Germany, 2017.
45. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
46. Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D.D. Hyperopt: A Python library for model selection and hyperparameter optimization. *Comput. Sci. Discov.* **2015**, *8*, 014008. [[CrossRef](#)]
47. Pearce, J.; Ferrier, S. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **2000**, *133*, 225–245. [[CrossRef](#)]
48. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [[CrossRef](#)]
49. Hyndman, R.J. Another look at forecast-accuracy metrics for intermittent demand. *Foresight Int. J. Appl. Forecast.* **2006**, *4*, 43–46.
50. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
51. Goos, P.; Meintrup, D. *Statistics with JMP: Graphs, Descriptive Statistics and Probability*; Wiley: West Sussex, UK, 2015.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.