

# Advancing remote sensing and artificial intelligence-driven frameworks for groundwater recharge forecasting in Morocco linking Groundwater depletion to Groundwater availability and reduction of trees vitality

Azeddine Elhassouny, *Member, IEEE*,

**Abstract**—While groundwater is a vital component of maintaining the world’s food supply, the primary source of fresh water, and an essential component of preserving the ecological balance of the planet, it is hardly ever fully utilized. Despite its significance and limited availability, groundwater management is rarely done in most countries of the world [1]. Continuous monitoring and precise projections of spatiotemporal groundwater recharge change can aid in sustainable development and effective groundwater resource management. Open public Remote sensing datasets were utilized to develop Artificial intelligence-based models such as Random Forest, XGBoost, long short-term memory (LSTM), and convolutional neural network (CNN) models to anticipate groundwater sheet recharging. The publicly available datasets are organized into three sections for training, testing, and validating: 2002–2009, 2010–2015, and 2015–2020. A comparison of the predictive models’ estimates of groundwater recharge in Morocco revealed that the LSTM model was more accurate with root mean square error (RMSE) equal to 20.05 mm/month. The study can be extrapolated to any other site (region) using our created pipeline. Our model’s performance on validation datasets demonstrates the utility and scalability of our combined remote sensing and artificial intelligence-based technology, opening up a new pathway for large-scale groundwater management.

**Index Terms**—Groundwater, Artificial Intelligence, Remote sensing, Forecasting, Time series, Recharge, Water withdrawal, Land subsidence, Water-Intensive Crops.

## I. INTRODUCTION

**I**N recent years, remote sensing-based groundwater management has incorporated artificial intelligence (AI) more frequently, providing new perspectives and new instruments for predicting and forecasting groundwater behavior. An overview of the most recent AI strategies for groundwater management via remote sensing is provided in this paper, along with used pertinent references [1]–[6] in this study.

Remote sensing techniques provide an efficient way to get insights into the status of groundwater supplies. Groundwater levels and depth, recharge and withdrawal rates, and other hydrological variables can be monitored over wide areas using remote sensing data from satellite or airborne platforms,

providing critical information for groundwater management. With the increased availability of remote sensing data sets, gridded hydrometeorological data, and digitized hydrography data, large-scale regions can now be monitored for a wide range of hydrologic applications. Some of the most utilized publically available data sets include official data portals of the NASS [7], CHC-UCSB, OpenlandMap [8], CSIRO.

Data from remote sensing has been extensively analyzed using AI approaches for groundwater management. Machine learning (ML) including deep learning (DL) are AI algorithms that have the capacity to learn from vast amounts of data and produce predictions, forecasting, classifications, and so on. Support vector machines (SVM), random forests (RF), and artificial neural networks (ANN), among others, are examples of machine learning (ML) approaches that have been used to predict groundwater depth levels [9], [10], calculate recharge rates [11], and map groundwater potential zones [12]. For instance, using data from the MODIS (Moderate Resolution Imaging Spectroradiometer). Support vector machine and data assimilation framework for Groundwater Level Forecasting using GRACE satellite data are used to predict groundwater levels in the northeast United States using the Gravity Recovery and Climate Experiment (GRACE) mission-informed groundwater anomalies data (GWA) and in-situ climate variables [13]. Random Forests (RF) was used to forecast groundwater withdrawals in Arizona using publicly available datasets and in situ groundwater withdrawal data available for Arizona Active Management Areas (AMA) and Irrigation Non-Expansion Areas (INA) [1]. With the aid of remote sensing data, ANN was used to map the possible groundwater zones [14].

Regarding DL methods, groundwater management applications have used convolutional neural networks (CNN) and recurrent neural networks (RNN) to analyze remote sensing data. To anticipate groundwater storage spatiotemporal change [15], two deep learning prediction models, long short-term memory (LSTM) and convolutional neural network-LSTM (CNN-LSTM), were developed using data from remote sensing. A comparison of the two deep learning predictive models with in situ measurements from South Korea’s National Groundwater Monitoring Network (NGMN) revealed that the CNN-LSTM

A. Elhassouny is with the Department of Software engineering, ENSIAS, Mohammed V University in Rabat, Rabat, Morocco e-mail: (see <http://ensias.um5.ac.ma/professor/m-azeddine-elhassouny>).

Manuscript received May 25, 2023; revised July 17, 2023.

model performed better [3].

Even though AI techniques have proven significant potential for groundwater management based on remote sensing datasets, in addition, these techniques have been experiments within a few regions and countries, there are still a number of difficulties that must be addressed. These problems include AI model interpretability, data availability, and data quality. Furthermore, further research is needed to combine multiple data sources and produce more reliable AI models.

In this study, we employ remote sensing-based datasets and deep learning models to estimate spatiotemporal changes in groundwater charges based on several influencing factors (meteorological, hydrological, and Eva-transpiration). Using deep learning for groundwater prediction models based on remote sensing data in Morocco is, to our knowledge, a new undertaking. We create an artificial intelligence framework for groundwater recharge forecasting using publicly available datasets and deep learning models, and then collect data on groundwater withdrawals in two locations in Morocco, both in the Taounate region (a cannabis cultivation area and a non-cannabis cultivation area). We created an end-to-end and integrated workflow that begins with obtaining datasets from multiple sources and progresses through preprocessing, fusion, scaling, normalizing, and so on, before windowing datasets into an LSTM and CNN-driven deep learning framework and providing a thorough deep analysis related to target scale and Gaussian filtering of land-use products. Furthermore, we conduct a more comprehensive study by devising numerous data windowing strategies (temporal, spatial, and spatiotemporal) and discover that the temporal data-splitting technique is the most effective. The LSTM window and its relationship to spatial scales are also examined. We also created a new method for estimating the loss of constrained aquifer storage and improving the characterization of aquifer characteristics and conditions by combining multivariate and groundwater storage utilization data.

The purpose of this study was to (i) in addition to using ensemble learning algorithms such as Random Forest(RF), and XGBoost to perform prediction of GWR, investigate several CNN and RNN-based time series forecasting methodologies, such as CNN, LSTM, residual LSTM, and LSTM autoregressive, for predicting groundwater recharge changes from multi-satellite data and images, (ii) generate long-term spatial groundwater recharge maps for Morocco, and (iii) identify groundwater recharges changes based on factor changes. This was done in order to evaluate the potential usefulness of combining deep-learning models with remote-sensing picture data.

## II. PROPOSED METHOD

The deep learning and remote sensing-based GWR forecasting framework (GWR-FF) workflow is divided into many stages. The workflow, as depicted in the figure (Fig. 1), may be divided into three major sections: Remote sensing Data Preparation-Thornthwaite-Mather (RSDP-TM) that consists of two parts (1) Data: Getting, preprocessing, and fusion(DGPF) and (2) TM procedure, and then Artificial intelligence-based models(AIM).

### A. Remote sensing data preparation and Thornthwaite-Mather(TM) procedure

1) *Data: Getting, preprocessing, and fusion(DGPF)*: The workflow is as follows: we use the Google Earth Engine platform to manage datasets from various sources, the Thornthwaite-Mather(TM) [16], [17] approach to provide spatiotemporal Groundwater recharge (GWR), and Tensorflow with Keras and Python 3-based libraries to build AI-based models.

The data collecting and pre-processing workflows are comparable to cutting-edge works [18]. We obtain Data from 2002 to 2020 using the Google Earth Engine platform [19] with the official data portals of NASA (MOD16A2.006: Terra Net Evapotranspiration(PET) 8-Day Global 500m ), CHC-UCSB(The Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) for Precipitation (Pr)) [20], OpenLandMap Soil Texture Class (USDA System), CSIRO(global plant root depth) datasets.

In the case of groundwater recharge, OpenLandMap records are utilized to define the clay, sand, and organic carbon content of the soil. Hengl et al. [21] make available a global dataset of soil water content at field capacity with a resolution of 250 m. Now that the soil parameters have been specified, the water content at the field capacity and at the wilting point was determined using the Thornthwaite-Mather(TM) method.

The entire pipeline is automated and relies on open-source or freely available programming languages, tools, and libraries. Tensorflow [22] with Keras [23] is used in this study to provide an approachable, highly-productive interface for solving machine learning problems, Python 3 [24] is used as the main programming backend for data acquisition, pre-processing, and implementing the machine learning model. The primary Python libraries used in our workflow include NumPy [25], SciPy [26], scikit-learn [27], GeoPandas [28], Pandas [29], Folium [30], seaborn [31], matplotlib [32], xgboost [33]. QGIS [34] is used for statistical analysis and visualization.

2) *Thornthwaite-Mathe (TM) Algorithm* [35]: Soil attributes datasets from OpenLandMap will be studied. The wilting point and field capacity of the soil will be determined as hydraulic properties of soil by applying some mathematical formulae to a variety of image collections. After iterating through the meteorological image collection and hydraulic parameters of soil, the Thornthwaite-Mather(TM) technique will be used to determine groundwater recharge in the area of interest.

- To relate water content at wilting point  $\theta_{WP}$  to the soil texture, Saxton and Rawls' equations [8] are utilized.

$$\theta_{WP} = \theta_{1500t} + (0.14\theta_{1500t} - 0.002) \quad (1)$$

with:

$$\theta_{1500t} = -0.024S + 0.487C + 0.006OM + 0.005(S \times OM) - 0.013(C \times OM) + 0.068(S \times C) + 0.031 \quad (2)$$

where:  $S$ : represents the sand content of the soil (mass percentage),  $C$ : represents the clay content of the soil

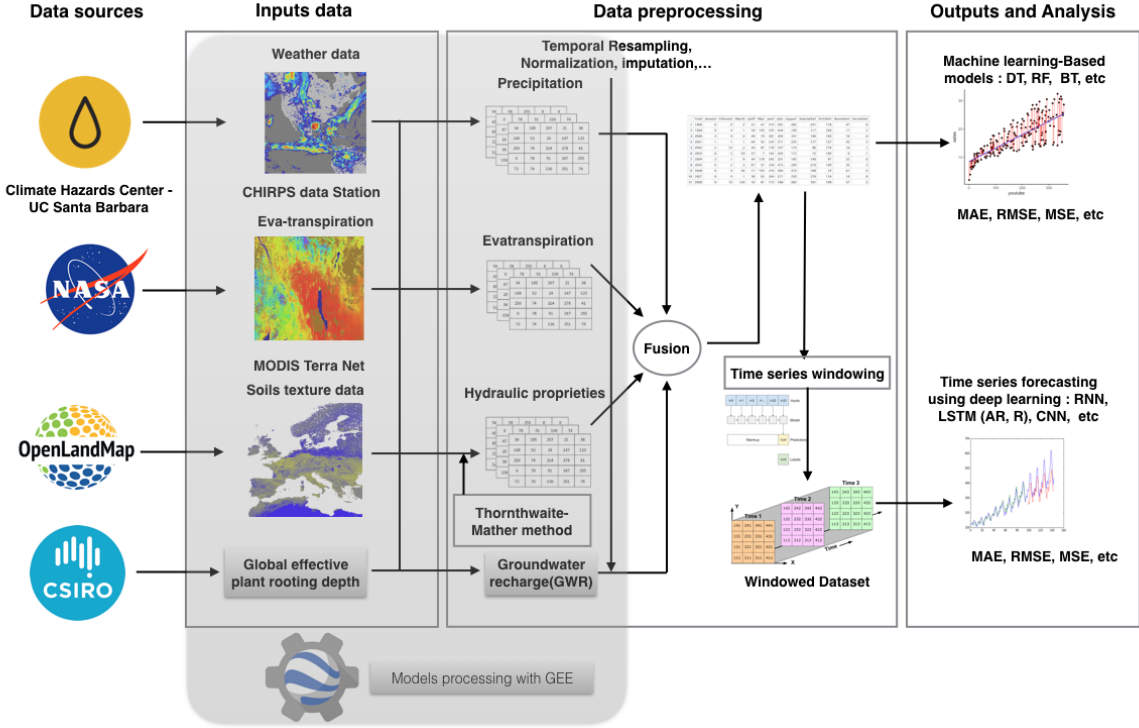


Fig. 1: The blocks in this diagram indicate the key steps in our workflow. The downloading of the data is demonstrated in Block 1, following which we create the predictor variables using the preprocessing block. The groundwater recharge dependent variable is created based on soil characteristics in this step using Thornthwaite-Mather(TM) procedure, along with additional procedures like reprojection, resampling, and statistical operations like temporal sum. The AI-based algorithmic steps, including model fitting and evaluation, are represented by the final block.

(mass percentage), and  $OM$ : represents the organic matter content of the soil (mass percentage).

$$OM = 1.724 \times OC \quad (3)$$

- Similar to the above, the following formula can be used to get the water content at field capacity:

$$\theta_{FC} = \theta_{33t} + (1.283\theta_{33t}^2 - 0.374\theta_{33t} - 0.15)w \quad (4)$$

with:

$$\begin{aligned} \theta_{33t} = & -0.251S + 0.195C + 0.011OM + 0.006(S \\ & \times OM) - 0.027(C \times OM) + 0.452(S \times C) + 0.299 \end{aligned} \quad (5)$$

For the Thornthwaite-Mather(TM) process to be formalized, certain definitions are required. According to Allen [36], the following definitions are provided:

$$T_{AW} = 1000 \times (\theta_{FC} \times \theta_{WP}) \times Z_r \quad (6)$$

where:  $T_{AW}$ : total soil water accessible in the root zone, expressed in millimeters,  $\theta_{FC}$ : the amount of water in the field when it is filled to capacity ( $m3m - 3$ ), and  $\theta_{WP}$  refers to the water content at the wilting point ( $m3m - 3$ ). ,  $Z_r$ : rooting depth (in millimeters),

In Table 19 of Allen et al. [36], typical values of  $\theta_{FC}$  and  $\theta_{AW}$  for various soil types are provided .

The equation readily available water ( $\mathcal{R}_{AW}$ ) :

$$\mathcal{R}_{AW} = p \times \mathcal{T}_{AW} \quad (7)$$

where is  $\mathcal{T}_{AW}$  total available water and  $p$  is the average percentage of total available water that can be exhausted from the root zone before moisture stress (between 0 and 1).

This global dataset leads to the reasonable assumption that the effective rooting depth in the vicinity of our region of interest Taounate (Olive) is  $Z_r = 1.45$ . The parameter  $p$  is likewise assumed to be constant and equal to  $p = 0.65$ , which is consistent with the typical values listed in Table 22 in [36], [37].

## B. Artificial Intelligence-based Models(AIM)

Given the varieties of AI algorithms output prediction or forecasting, machine learning or deep learning, and different dimensions temporal, spatial, or spatiotemporal, we experiment with ensemble learning for prediction and deep learning for forecasting.

1) *Machine learning-based groundwater recharge prediction (MLGWRP)*: The objective of a multi-variate regression problem is to accurately anticipate annual groundwater recharge (GWR) given a number of predictors. Ensemble Learning Algorithms such as random forest (RF), and XG-Boost are among the machine learning techniques utilized to tackle this problem. The simplest model to construct using Ensemble Learning Algorithms or any other machine learning

**Algorithm 1** DGPf Algorithm**Input:**

$i_{date}, f_{date}$ ; { %Initial and final dates of interest.}  
 $\mathcal{P}_{oi}(L_{POI}, l_{POI})$ ; { % Longitude and latitude that defines the location of interest with a point.}  
 $\mathcal{F}_{VB}$ ; { % Van Bemmelen factor.}  
 $Scale$ ; //A nominal scale in meters of the projection to work in [in meters].  
 $Source$ ; //Source of dataset.

**Output:**

$\mathcal{DF}_m^{POI}$  //meteorological dataset as a DataFrame at the location of interest

0: **procedure** DGPf

//Initialization

- 1: Compute the wilting point  $\theta_{WP}$ ; //based on  $\theta_{1500t}$  using Eq. (1)
- 2: Compute the field capacity  $\theta_{FC}$ ; //based on  $\theta_{33t}$  using Eq. (4)
- 3:  $\mathcal{S} \leftarrow \text{GSPF}(\text{OpenLandMap datasets}, Scale, SP = \text{'Sand'})$
- 4:  $\mathcal{S} \leftarrow \text{GSPF}(\text{OpenLandMap datasets}, Scale, SP = \text{'Sand'})$
- 5:  $\mathcal{C} \leftarrow \text{GSPF}(\text{OpenLandMap datasets}, Scale, SP = \text{'Clay'})$
- 6:  $\mathcal{OC} \leftarrow \text{GSPF}(\text{OpenLandMap datasets}, Scale, SP = \text{'Orgc'})$
- 7: Compute Organic Matter  $\mathcal{OM}$ ; //based on Organic Carbon content  $\mathcal{OC}$  using Eq. (3);
- 8:  $\mathcal{Pr} \leftarrow \text{EE.IMAGECOLLECTION}(\text{UCSB-CHG/CHIRPS/DAILY}, i_{date}, f_{date}, \mathcal{P}_{oi}, SP = \text{'precipitation'})$
- 9:  $\mathcal{Pet} \leftarrow \text{EE.IMAGECOLLECTION}(\text{UCSB-CHG/CHIRPS/DAILY}, i_{date}, f_{date}, \mathcal{P}_{oi}, SP = \text{'evapotranspiration'})$
- 10:  $\mathcal{Pr}(month) \leftarrow \text{SUMRESAMPLER}(\mathcal{Pr}, \text{freq} = 1, \text{scale\_factor} = \text{'month'}, \text{band\_name} = \text{'Precipitation'})$
- 11:  $\mathcal{Pet}(month) \leftarrow \text{SUMRESAMPLER}(\mathcal{Pet}, \text{freq} = 1, \text{scale\_factor} = \text{'month'}, \text{band\_name} = \text{'evapotranspiration'})$
- 12:  $\mathcal{DF}_{weather} \leftarrow \text{FUSION}(\mathcal{Pr}(month), \mathcal{Pet}(month))$
- 12: **end procedure**=0

algorithms on a dataset with a standard form is one that predicts the value of a target variable, such as GWR, only based on the values of the current predictor variables. XGBoost, [38]–[40], Random forest (RF) [1], [5] are some of the recently developed algorithms, that are used to tackle a multi-variate regression problem where the goal is to properly predict yearly groundwater withdrawals (GWD) given several predictors.

2) *Deep learning-based groundwater recharge time series forecasting (DLGWRf)*: In this study, we experimented with various AI-based time series forecasting algorithms such LSTM, CNN, with different strategies, including :

- Univariate and Multivariate: using one dependent variable which is recharge, or with other impacting factors such as precipitation, evaporation, etc.
- Forecast for a single time step: a single feature, or all

**Algorithm 2** TM Algorithm**Input:**

$i_{date}, f_{date}$  : { %Initial and final dates of interest.}  
 $\mathcal{P}_{oi}(L_{POI}, l_{POI})$ : { % Longitude and latitude that defines the location of interest with a point.}  
 $\mathcal{F}_{VB}$ : { % Van Bemmelen factor.}  
 $\mathcal{Z}_r(\mathcal{P}_{oi}(L_{POI}, l_{POI}))$  //rooting depth around our region of interest.  
 $\sqrt{\quad}$  /parameter constant according to Table 22 of Allen et al. (1998).

**Output:**

$\mathcal{R}_m$  //Monthly Recharge

$\mathcal{ST}_m, \mathcal{APWL}_m$  //Available/Amount of water stored in the root zone for the month  $m$

0: **procedure** TM

//Initialization

- 1: Compute averaged value between reference depths of the water content at field capacity ( $\theta_{WP}(b_i)$ ; //based on  $\theta_{WP}(b_i)$ )
- 2: Compute averaged value between reference depths of the water content at wilting point ( $\theta_{FC}(b_i)$ ); //based on  $\theta_{FC}$
- 3: Calculate the theoretical available water  $\mathcal{T}_{AW}$ ; //based on ( $\theta_{FC}$ ) and ( $\theta_{WP}$ ) using Eq. (6)
- 4: **if**  $\mathcal{P}_{et} \notin \mathcal{P}\nabla$  **and**  $\mathcal{APWL}_m = \mathcal{APWL}_{m-1} + (\mathcal{PET} - \mathcal{P}\nabla)$  **then**
- 5:  $\mathcal{ST}_m \leftarrow \mathcal{ST}_{m-1} + (P_m - \mathcal{PET}_m)$ ; //amount of water stored in the root zone for the month  $m$
- 6: **else if**  $\mathcal{ST}_m \notin \mathcal{ST}_{FC}$  **then**
- 7:  $\mathcal{R}_m \leftarrow \mathcal{ST}_m - \mathcal{ST}_{FC} + P_m - \mathcal{PET}_m$ ; //recharge
- 8:  $\mathcal{ST}_m = \mathcal{ST}_{FC}$ ; // the water stored at the end of the month becomes equal to water stored at field capacity
- 9: **else if**  $\mathcal{ST}_m \in \mathcal{ST}_{FC}$  **then**
- 10:  $\mathcal{APWL}_m \leftarrow \mathcal{ST}_{FC} \times \ln(\mathcal{ST}_m / \mathcal{ST}_{FC})$ ; // the accumulated potential water loss for the month  $m$
- 11:  $\mathcal{ST}_m = \mathcal{ST}_{FC}$ ; // and no percolation occurs
- 12: **end if**
- 12: **end procedure**=0

features.

- Forecast multiple steps: Single-shot that consists on make the predictions all at once, or Autoregressive which makes one prediction at a time and feed the output back to the model.

3) *Deep learning-based groundwater recharge Spatio-temporal forecasting (DLGWRSTF)*: To the best of our knowledge, multivariate or univariate forecasting for groundwater recharge, storage, or withdrawals has only been done at this time using either simple predictions [1], temporal updates [4], or distinct spatial and temporal updates which still a model without any application on groundwater management.

The goal of multivariate or univariate groundwater recharge prediction is to forecast actual values in light of real-world circumstances [1]. Groundwater recharge time series forecasting, on the other hand, concentrates on predicting future values based on the previous context. Modern sequence-to-sequence models rely on neural attention between timesteps, which sup-



**Algorithm 3** Main Algorithm**Input:**

$\mathcal{DF}_m^{\mathcal{POI}}$  //meteorological dataset as a DataFrame at the location of interest  
 $\mathcal{R}_m$  //Monthly Recharge  
 $\mathcal{ST}_m, \mathcal{APWL}_m$  //Available/Amount of water stored in the root zone for the month  $m$   
 $i_{date}, f_{date}$ ; { %Initial and final dates of interest.}  
 $\mathcal{P}_{oi}(L_{POI}, l_{POI})$ ; { % Longitude and latitude that defines the location of interest with a point.}  
 $\mathcal{F}_{VB}$ ; { % Van Bemmelen factor.}  
 $Scale$  ; //A nominal scale in meters of the projection to work in [in meters].

**Output:**

$C_i$  //Updated category of patient's readmission after performing the algorithm  
 $\mathcal{C}_m, \mathcal{C}_m^{(n)}$  //Prior/New Confusion matrix, i.e., before/after performing the proposed algorithm  
 $\mathcal{O}_a, \mathcal{O}_a^{(n)}$  //Prior/New Pveral accuracy

0: **procedure** MAIN

//Initialization;

- 1:  $\mathcal{DF}_m^{\mathcal{POI}} \leftarrow \text{DGPF}(i_{date}, f_{date}, \mathcal{P}_{oi}(L_{POI}, l_{POI}), \mathcal{F}_{VB}, Scale, Source)$
- 2:  $\mathcal{R}_m, \mathcal{ST}_m, \mathcal{APWL}_m \leftarrow \text{TM}(i_{date}, f_{date}, \mathcal{P}_{oi}(L_{POI}, l_{POI}), \mathcal{F}_{VB}, Scale, \sqrt{\mathcal{Z}_r(\mathcal{P}_{oi}(L_{POI}, l_{POI}))}, Source)$
- 3:  $\mathcal{GDF}_m^{\mathcal{POI}} \leftarrow \text{FUSION}(\mathcal{DF}_m^{\mathcal{POI}}, \mathcal{R}_m, \mathcal{ST}_m, \mathcal{APWL}_m)$
- 4: Reshape  $\mathcal{C}_m$ ; //based on  $T_i$  and  $C_i$
- 5: MSE, MAE, RMSE  $\leftarrow \text{AIM}(\mathcal{DF}_m^{\mathcal{POI}}, \mathcal{R}_m, \mathcal{ST}_m, \mathcal{APWL}_m)$
- 5: **end procedure**=0

ports temporal learning but ignores clear spatial correlations between variables. On the other hand, graph neural network-based techniques explicitly model variable relationships. These techniques, however, frequently carry out independent spatial and temporal updates and rely on established graphs that are immutable through time. We solve these issues in this paper by developing a novel transformer model and translating multi-variate forecasting into a "spatiotemporal sequence" formulation [41] in which each Transformer input token represents the value of a single variable at a particular time. Along this lengthy sequence, Long-Range Transformers can learn interactions between space, time, and value information. While learning spatiotemporal associations simply from data, our system, which we call Spacetimeformer, produces competitive results on benchmarks ranging from traffic forecasting to electricity demand and weather prediction.

## III. STUDY AREA AND DATA

## A. Study area

Groundwater is heavily used in agriculture in several Moroccan locations to grow year-round crops and trees such as Watermelon, Tomato, Avocado, Sours, Oranges, and a variety of other specialized crops, and agriculture consumes 87% of water resources [42]. Like with other regions in Morocco,

Taounate has endured two scourges in the past twenty years: Cannabis farming and a droughts climate with very variable precipitation. Groundwater resources are severely utilized due to rising freshwater demands and the ongoing drought. This is especially true in Taounate (cannabis region), where there is little access to surface water and pumping typically occurs from thick, unconsolidated aquifers, which has a significant impact on groundwater availability and trees vitality. In the past, around 20 years ago, in Taounate regions such as Sidi Yahia Bni Zeroual, the water streams would pour down from mountains throughout the year, providing the community with a distinctive source of drinking water. Since the year 2000, when the region began growing cannabis plants of various sorts that require irrigation water, several wells of varying depths have been drilled, and people have begun to notice that as many wells have been drilled, so many water sources are drying up. On the other hand, there are still several drinking mountain water sources in the other regions like Ourtzagh (a Taounate non-cannabis territory), which has access to surface water thanks to the El-Wehda dam. Due to the large disparities in crop production between the two zones, descriptions of the in situ groundwater depletion, and the availability of data, we chose Taounate with both zones (Sidi Yahia Bni Zeroual and Ourtzagh) as the test sites for testing our suggested approach and hypotheses. Another reason for choosing this area was a knowledge of the connections between groundwater depletion for Cannabis irrigation and tree viability.

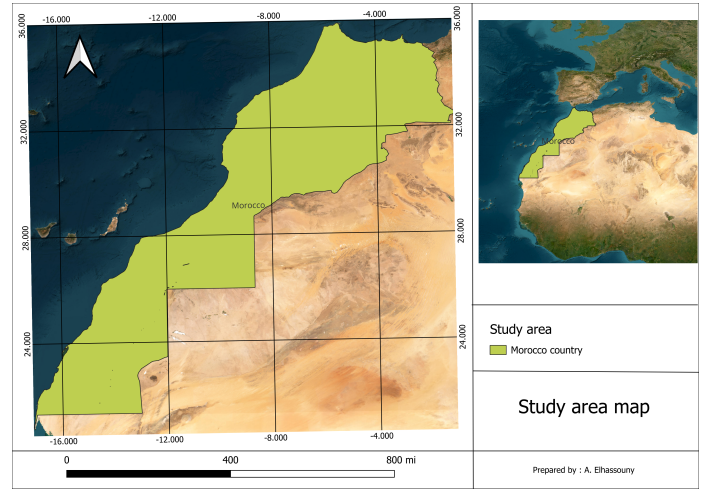


Fig. 2: Map of Morocco as the study area, as well as the Earth's largest hot desert, covering nearly all of northern Africa.

## B. Data

Remote sensing-based data of characteristics that directly or indirectly affect groundwater recharge were used as input for deep learning models. For this investigation, we combined data such as meteorological, hydrological, etc from numerous sources. The meteorological datasets consist of the following: MODIS Terra Net provides evapotranspiration on an 8-day basis, and Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS) provides precipitation on a daily basis, both of them with a resolution of  $500m^2$ . Datasets from

OpenLandMap are used to describe the amounts of clay, sand, and organic carbon in the soil. Hengl et al. have made a global dataset of soil water content at the field capacity with a resolution of  $250m^2$  available (2019). As a result, both parameters will be determined in the following utilizing the global datasets that show the soil's sand, clay, and organic matter contents together with the prior equations. As stated in the summary. In the TM method, the wilting point and field capacity of the soil are two hydraulic parameters that are frequently utilized. The wilting point denotes the depth beyond which plant roots cannot extract water, while the field capacity denotes the depth at which soil can no longer store water. When gravitational pull exceeds a certain threshold, water begins to seep into the lower levels.

#### IV. EXPERIMENTATION AND DISCUSSION

In this study, we successfully advance recent state-of-the-art works such as [1], [5], and extend the prediction of current groundwater withdrawals to forecast the future recharge/level withdrawals of a new region by offering new insights and new methods, especially on using univariate and multivariate, forecasting future based on the past with different windowing. We also establish a pipeline to expand the application of our approach to any new region and relate the groundwater flow from mountains, trees vitality, and groundwater availability to the anticipated groundwater recharge and withdrawals. Finally, we suitably illustrate the extensibility of our approach to a framework, taking into account the fact that the used data are available for any region in the world.

*a) Training Paradigm:* We employed two ways for data splitting. First, we divided the dataset into three sections: training, testing, and validation, with percentages of 50%, 25%, and 25%, respectively. Following that, we set up and looped the pipeline for numerous iterations to conduct all necessary tasks such as data preparation, feature engineering, model building, optimization, hyperparameter tuning, and validation. Second, we replicate the data for the second time, dividing it into two parts training and testing, with 75% and 25% respectively. We train the obtained models using ideal parameters gotten from the first time, and then we test them on the testing part as depicted in Figure 3.

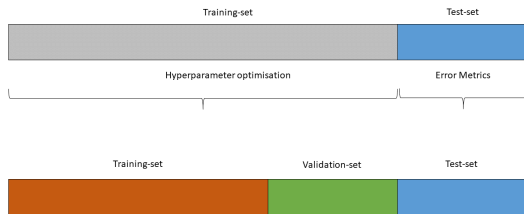


Fig. 3: Time series dataset is divided into three sections for training, validation, and testing

*b) Performance Metrics:* Many metrics, including the Nash-Sutcliffe efficiency (NSE), squared Pearson's correlation coefficient (R2), absolute and relative root mean squared errors (RMSE and rRMSE), absolute and relative biases (Bias and

rBias), and the persistency index (PI), are used to assess forecast accuracy.

##### A. Ensemble learning models

The most straightforward model one can build on a dataset with a standard shape using XGBOOST and/or Random Forest, or any other machine learning algorithms, is one that predicts the value of a target variable, like GWD, based only on the values of the current predictor variables.

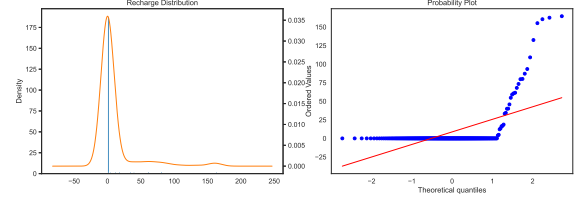


Fig. 4: Standardized residuals restricted within the  $[-2, 2]$  interval and Q-Q plots.

Figure 4, the left subfigure shows the standardized residuals restricted within the  $[-2, 2]$  interval (the orange line represents the Gaussian probability density function), and the Right subfigure shows Q-Q plots. As seen in the GWR distribution plot and supported by the Q-Q(quantile-quantile) plot, the GWR distribution is not close to being normal.

Figure 5 illustrates the fact that the expected recharge for both the training period (2002–2013) and the validation period (2013–2020) is rather accurate, and our models RF (red line) and XGBoost (green line) are able to capture the temporal variations precisely. However, we see that the RF model does not adequately capture the temporal patterns for the peak values low or high. This may be due to the hyperparameters max depth and minsamplesleaf, which control model overfitting. That conclusion is supported by the performance metrics mean absolute error (MAE) values of both models RF and XGBoost, which are 7.56 and 6.65, respectively. As a result, the XGBoost model outperforms the RF model. The precise groundwater recharge estimations for the test or validation data continue to demonstrate our model's significant generalizability.

The Random forest feature importance for Pr, pet, apwl, and st is 0.60, 0.063, 0.15, and 0.18, respectively, whereas for XGBoost it is 0.62, 0.047, 0.34, and  $6.98e-08$ , in that order. The features' importance demonstrates that recharge is primarily dependent on the amount of precipitation; the more precipitation, the more recharge. This is logical because, in the absence of precipitation, there is no water available for evapotranspiration, plant consumption, or recharge, regardless of whether there is or is not evapotranspiration.

##### B. Deep learning-based models for time series forecasting

In this study, we extend machine learning-based models to deep learning-based models for time series forecasting. We investigate two advanced deep learning algorithms: Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)-based architecture, Long short-term memory (LSTM) in residual and autoregressive forms, with varied

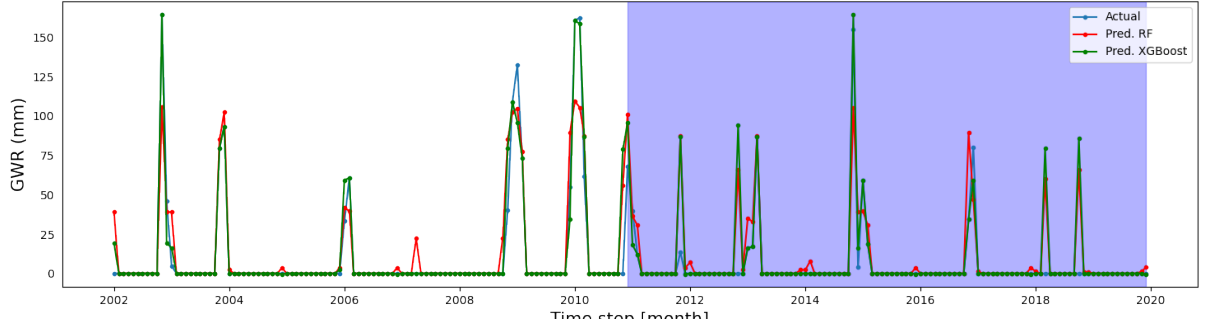


Fig. 5: Mean actual, XGBoost, and RF-based predicted groundwater recharge (GWR) over the region for each month, with 2010–2020 being the validation years.

dataset windowing shapes, one last step to forecast one future step, many past steps to forecast one future step, many past steps to forecast many future steps. The main goals of this work are to forecast GWR in a given location. In this section's exploration part, we investigate the GWR variable and other aspects of the dataset in order to state hypotheses.

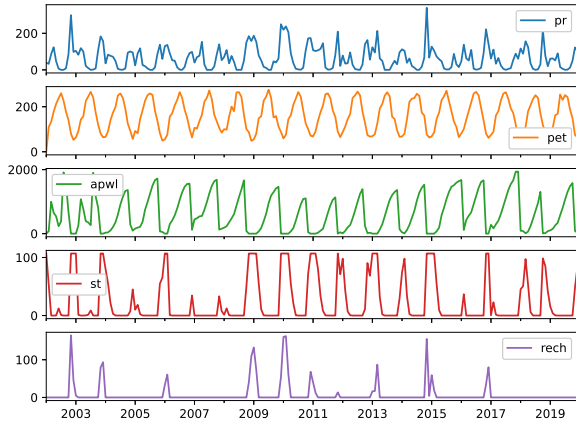


Fig. 6: Evolution of features over time

Figure 6 depicts the evolution of all features such as precipitation (pr), evapotranspiration (pet), water lost in the field (apwl), water storage (st), and GWR (rech) over time from 2002 to 2020. We illustrate the evolution of characteristics over time to create periodicity hypotheses for time series-based algorithms. Also, Figure 6 shows that there are not many disruptions, which is normal for natural occurrences. Figure 7 illustrates the average GWR resampled over Month, Quarter, and Year for exploration purposes in order to identify the time unit of periodicity.

There are numerous techniques to deal with periodicity. For example, sine and cosine transform can be used to figure out the time of season and year signals, making it possible to obtain acceptable signals. However, the time of day is not a suitable model input. Having data on groundwater recharge, it has a distinct quarterly and annual cycle, as shown in figure 8.

The most crucial frequency features are accessible to the model through figure 8. In this instance, it is possible to understand the importance of time frequencies like year and season. For instance, by extracting features from the Fast

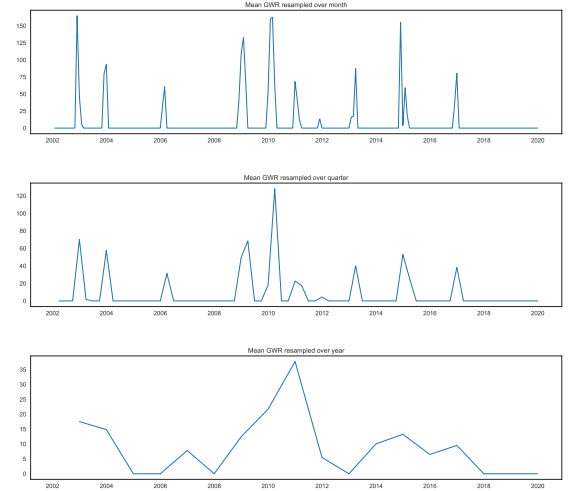
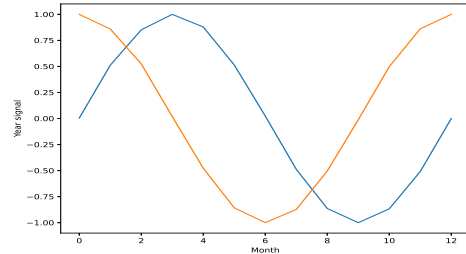
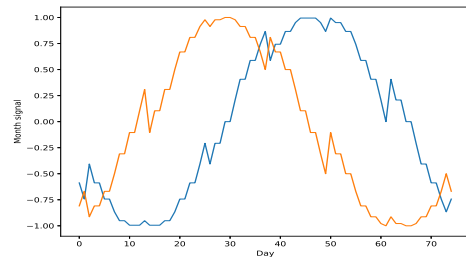


Fig. 7: Average GWR resampled over Month, Quarter, and Year.



(a) Time of year signal



(b) Time of month signal

Fig. 8: Sine and cosine transform of GWR signals  $A_i^{(c)}$ .

Fourier Transform, one can figure out which frequencies are crucial. Figure 9 shows that the noticeable peaks occur at a frequency close to  $\frac{1}{Year}$ .

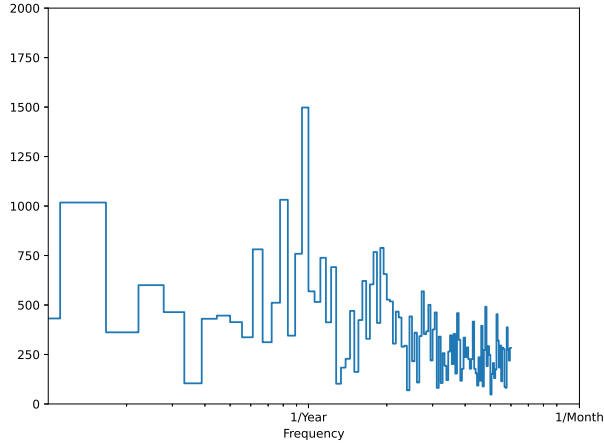
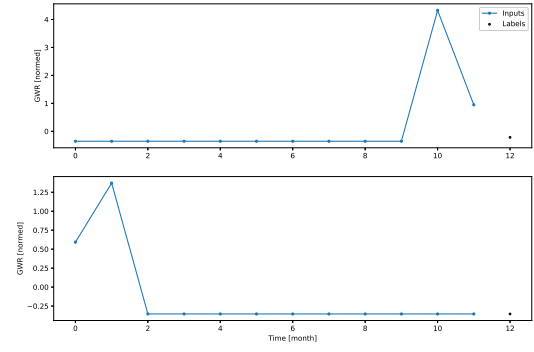


Fig. 9: Evolution of features over time

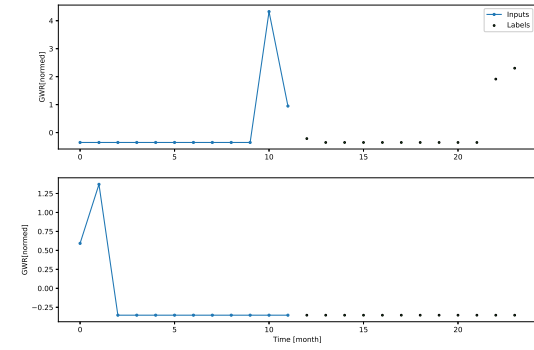
a) *Data windowing*: There are four data windowing paradigms in the literature, One2One, One2Seq, Seq2One, and Seq2Seq [43], [44], and the choice of one of them is based on a hypothesis made upon many factors such as problem formulation, dataset shape, etc. However, a model that solely relies on the present circumstances to forecast the value of a single feature one-time step into the future. The current input values of a single-time-step model are unrelated to any previous values. It is unable to observe how the input features evolve over time. In order to handle the problem of forecasting Groundwater Recharge (GWR) based on predictors, the model requires access to numerous time steps for that reason we used Seq2One and Seq2Seq.

In order to develop a model that can forecast future outputs based on the previous inputs, one needs to take a time series dataset which is a list of consecutive entries and convert it into a window-shaped dataset of entries and label pairs (inputs, labels). Data windowing is a crucial stage in time series-based algorithms. One needs to specify the window(s) that determine the time step(s) in the past, time step(s) in the future, and offset time. Forecast one step at a time using all features or just one. To predict numerous phases at once, use single-shot forecasting. Make one prediction at a time with autoregressive, then feed the results back into the model. Data windowing is one of the primary pretreatments used to reshape data in a suitable manner for time series-based models. One may want to construct a variety of data windows depending on the task and type of model. To make a prediction one month into the future based on 12 months of history, for instance, one might define a window with (*input* = 12, *labels* = 1, and *offset* = 1), alternatively, one could base a prediction 12 months into the future on one or several years of history, as indicated in the figures below ( Fig. 10).

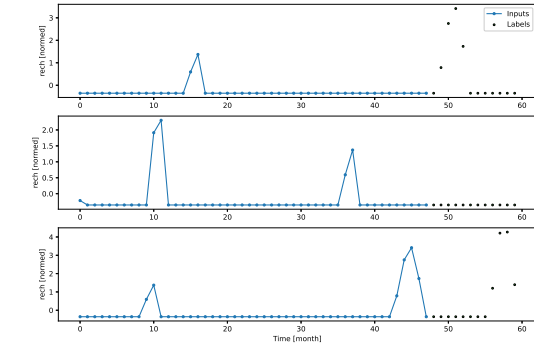
b) *Univariate vs. Multivariate Forecasting*: In terms of the features involved, there are two types: univariate and multivariate forecasting. Univariate forecasting uses only one



(a) Two examples of data windows of GWR variable with  $Window(input = 12, labels = 1, shift = 1)$ .



(b) Two examples of data windows of GWR variable with  $Window(input = 12, labels = 12, shift = 12)$ .



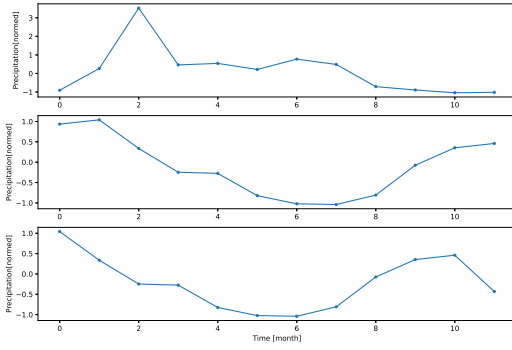
(c) Three examples of data windows of GWR variable with  $Window(input = 36, labels = 12, shift = 1)$ .

Fig. 10: Examples of different data windows of GWR as feature and target variable in GWR univariate forecasting.

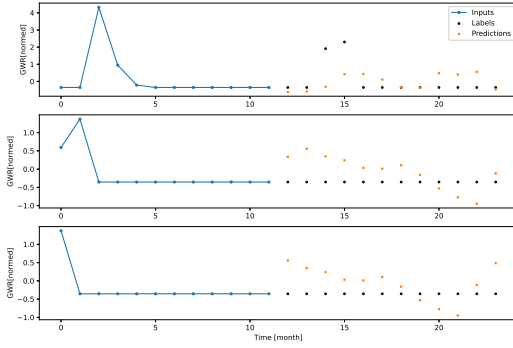
feature, in this case, the groundwater recharge (GWR) variable, whereas multivariate forecasting includes many features (more than two variables), including precipitation (pr), evapotranspiration (pet), water lost in the field (apwl), and water storage (st) features in addition to the GWR.

In a window with 12 timesteps of input and 12 timesteps of output, for the precipitation variable as well as other variables, one may only provide a series of timesteps of past values, for instance, 12 timesteps as shown in Figure 11a, but for the GWR variable, we provide 12 timesteps of past values and 12 timesteps as labels, resulting in 24 timesteps, as shown in Figure 11b. The 12 timesteps as labels are used to validate





(a) Three examples of windows of precipitation feature with  $Window(input = 12, labels = None, shift = None)$



(b) Three examples of Data windows of GWR as feature and target variable with  $Window(input = 12, labels = 12, shift = 12)$

Fig. 11: Examples of Data window of used features in GWR multivariate forecasting

model predictions and calculate performance indicators.

#### 1) CNN and LSTM-based Univariate GWR Forecasting:

In this study, LSTM and CNN models powered by remote sensing data were built and evaluated. The models were built using TensorFlow, Keras, and Python 3. The LSTM model had an input layer, an output layer, a dense layer completely connected to hidden nodes, and an LSTM layer composed of each cell memory. Data are communicated through the max-pooling layer and flattened layer in the CNN model before being delivered from the input layer to the 2D convolution layer. To avoid interrupting network convergence, the remote sensing-based input variables were adjusted using the z-score scaling method. This approach is defined by Eq. 8, which can be altered to limit the data range to a small range.

$$z = \frac{x - \mu}{\sigma} \quad (8)$$

We built an LSTM model with 100 *neurons* in the first hidden layer and 1 *neurons* in the output layer to forecast groundwater recharge (GWR). The suggested model is assessed using the MSE loss function, and the model is fitted with a batch size of 36 *instances* using the effective *Adam* version of stochastic gradient descent. We adjusted a number of hyperparameters, including the number of training epochs, the dropout rate (20%), and the input shape, which

will be many timesteps in the past to forecast one in the feature, depending on the forecasting window shape. Due to its ability to converge more quickly than the sigmoid or tangent hyperbolic functions, the ReLU function was chosen for use as an activation function.

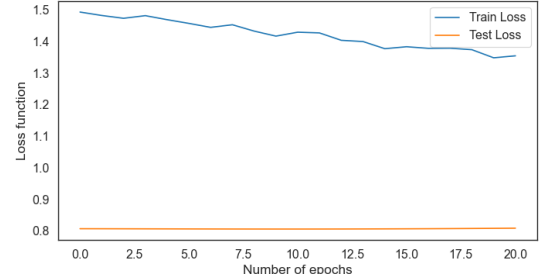


Fig. 12: LSTM Loss function changes according to the number of epochs for training and testing

Deep learning models' performance over Region of interest (ROI) is shown by the time series in Figure 13. The magnitude and variability of GWRs obtained from the LSTM (red dotted line in Figure 13) model were consistent with those of in situ measurements (blue line). The LSTM univariate model captured both the monthly and seasonal dynamics of in situ GWR, apart from showing stable performance for the test period (March 2013 to December 2019). However, a comparison of the LSTM model's predictions with in situ measurements indicated that the model tended to underestimate large values while overestimating small or zero values of GWRs.

2) *LSTM and CNN-based Multivariate Forecasting:* In the previous section, we attempted to forecast GWR based just on fluctuations in GWR, however, in this section, LSTM, and CNN models were utilized to accurately predict GWR over Morocco utilizing various feature combinations from multi-satellite data. The deep learning models included a variety of input variables (precipitation, evapotranspiration, changes in soil moisture storage, and so on). It is critical to understand the sensitivity of their predictive performance in order to avoid overfitting issues. The forecast accuracies of several parameter combinations were compared during hyperparameter optimization based on Bayesian optimization.

#### a) Seq2One Multivariate LSTM, and CNN forecasting:

A comparison of the two deep learning models' predictions of groundwater storage change with in situ measurements of the GWR in Morocco revealed that the LSTM model was more accurate with Normalize RMSE (is determined using the normalized values of the true values and the anticipated values)  $RMSE = 0.2952mm/month$ , than the CNN model with  $RMSE = 0.3501mm/month$ . The outcomes are depicted in Figure 14.

The deep learning models' performance over ROI is shown by the time series in Figure 15. The magnitude and variability of GWRs obtained from the LSTM (red dotted line; Figure 15) and CNN (green line; Figure 15) models were consistent with those of in situ measurements (blue line). The LSTM and CNN Multivariate models captured both the monthly and seasonal dynamics of in situ GWR associated with all features such as precipitation, evapotranspiration, APWL, ST, and Recharge itself, apart from showing stable performance

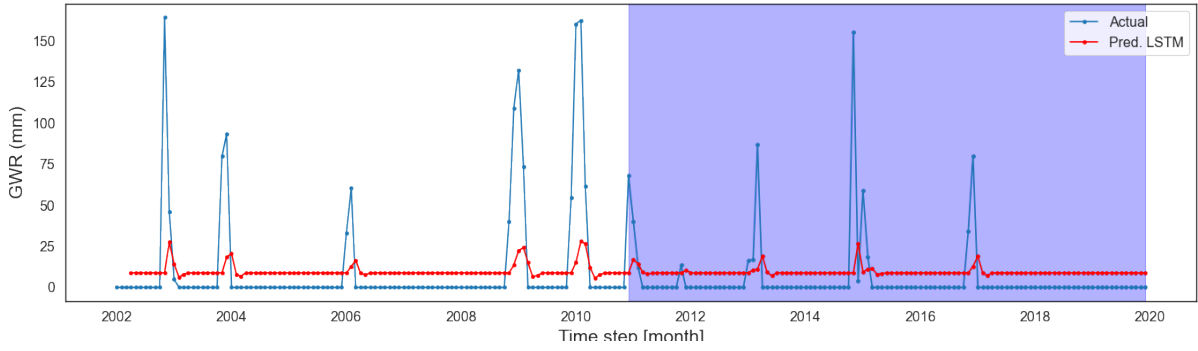


Fig. 13: GWR time-series comparisons predicted by the LSTM (red dotted line) Univariate model with in situ measurements (blue line) CNN green Line corresponding to TMPA precipitation.

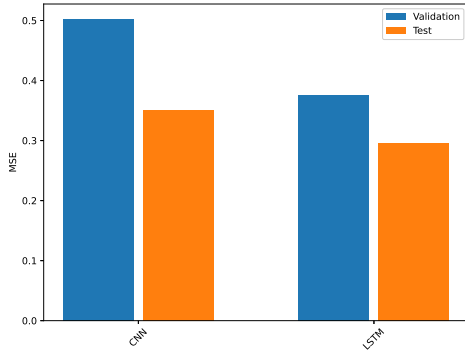


Fig. 14: Normalize RMSE error values for LSTM and CNN models for Test and Validation

for the test period (March 2013 to December 2019). However, a comparison of the LSTM model's predictions with in situ measurements indicated that the model tended to overestimate GWRs.

*b) LSTM residual model:* Building models that forecast how the value will change in the following time step rather than the upcoming value is a typical practice in time series analysis. Similar to this, in deep learning, residual networks, or ResNets, refer to topologies where each layer increases the model's cumulative output. That is how one benefits from the understanding that the change should be minimal. In essence, a model with a residual link facilitates faster convergence and somewhat improved performance. Any model can be utilized in conjunction with this strategy. The LSTM model, to which it is being applied in this instance, has been parametrized to make sure that the initially anticipated changes are modest and don't overwhelm the residual connection.

Figure 16 illustrates how the residual technique outperformed LSTM in terms of forecasting, with RMSE falling from 0.4539 for the LSTM model to 0.4094 for Residual LSTM. With an RMSE of 0.3501, CNN remains the best model out of all those with and without residuals.

*c) Seq2Seq Multivariate LSTM, CNN, and LSTM Autoregressive forecasting:* Both the single-output and multiple-output models in the previous sections made single-time step predictions, one month into the future. This section looks at how to expand these models to make multiple time-step predictions. In a multi-step prediction, the model needs to learn to predict a range of future values. Thus, unlike a single-step

model, where only a single future point is predicted, a multi-step model predicts a sequence of future values. There are two rough approaches to this: (i) Single shot predictions where the entire time series is predicted at once. (ii) Autoregressive predictions where the model only makes single-step predictions and its output is fed back as its input. In this section, all the models will predict all the features across all output time steps. For the multi-step model, the training data again consists of monthly samples. However, here, first, the models will learn to predict 12 months (one year) into the future, given 12 months of the past, and then to predict one year into the future, given 3 years months of the past.

Autoregressive refers to that the output of each model can be fed back into itself at each stage, and predictions can be made based on the preceding one. We used both of the above models (CNN and LSTM) in an autoregressive feedback loop, but we built an LSTM-Autoregressive (LSTM-AR) model that was specifically designed for autoregressive feedback. Figure 19 depicts the obtained result, which shows that LSTM-AR (sky blue line) captures well the change of GWR, after which CNN and LSTM executed an autoregressive feedback loop. The obtained RMSE for the three approaches are 0.4185, 0.4419, and 0.4590, in that order: LSTM-AR, CNN, and LSTM as shown in Figure 18.

*d) Long-term Autoregressive LSTM forecast:* Given the importance of long-term forecasting systems in general, and groundwater in particular, we anticipated a lengthy duration of investigation to pave the way for building lengthy-term memory by investigating autoregressive principles. For predicting the future value of GWR, the RMSE function is used in accordance with the number of time steps to be taken into account. One can learn about the best window shape for accurate prediction. Figure 20 shows the plot showing the correlation between the RMSE and the number of history timesteps used to forecast the GWR value during the whole period. As seen in Figure 20, the value of a window from which we obtain the minimum MSE is the three timesteps in a month. When we try to anticipate the far future, the MSE value rises; the further we go into the future, the more imprecision we receive. We discovered that a good period in the past to predict one year in the features is three years after hyperparameters tweaking with regard to the past period and future period. Figure A displays the predictions made using the LSTM autoregressive model and the ground truth of the GWR.

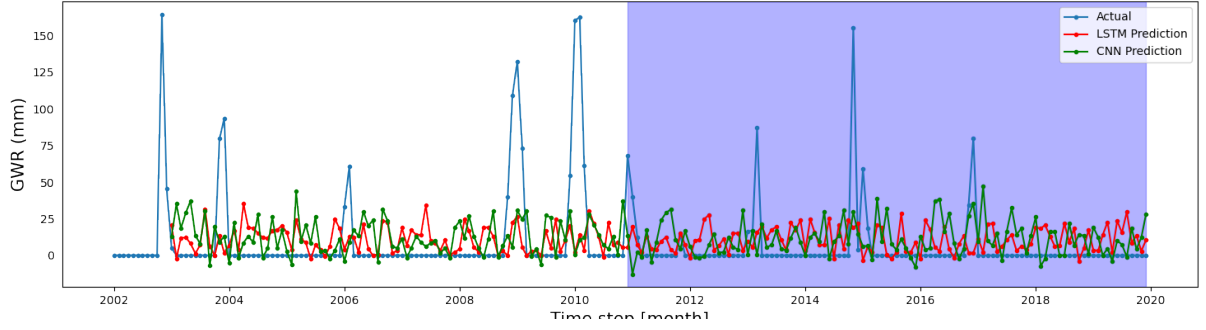


Fig. 15: GWR time-series comparisons Seq2One Multivariate forecasting by the LSTM (red dotted line) Multivariate model with in situ measurements (blue line) CNN green Line. corresponding to TMPA precipitation.

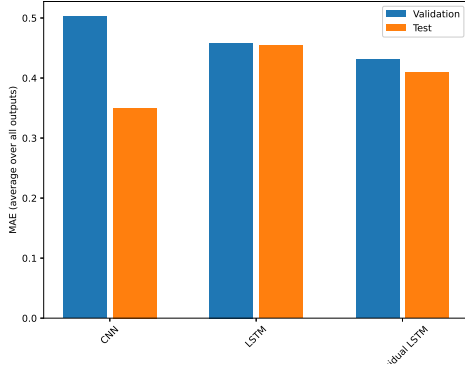


Fig. 16: Normalized RMSE error values for Residual LSTM, LSTM, and CNN models for test and validation

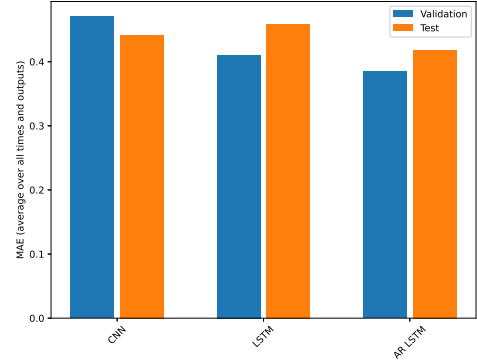


Fig. 18: Normalize RMSE error values for LSTM-Autoregressive (LSTM-AR), LSTM, and CNN models for Test and Validation.

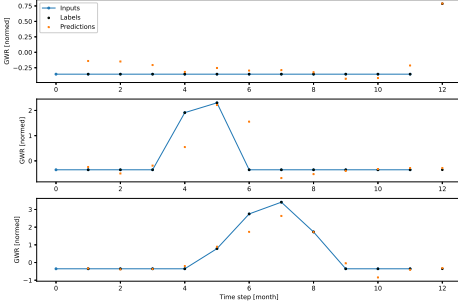


Fig. 17: Example of data windows of GWR variable with the blue dotted line as inputs, black dots as actual labels, and orange dots as forecasted values:  $Window(input = 36, labels = 12, shift = 1)$ .

More than any other model before it, the model accurately represents the high and low values in GWR. This conclusion is supported by Normalized MAE  $RMSE = 0.4815$  as depicted in Figure 20. As shown in figure 21, we investigate 3 years in the past to predict 1 years in the future.

It is well understood that groundwater recharge cannot be readily monitored and is influenced by spatially and temporally complicated processes. In addition, Groundwater is critical to the preservation of ecosystems and the adaptation of humanity to climate change which is with long-term impacts. However, given the importance of long-term forecasting systems in general, groundwater recharge, storage, and withdrawal in particular, and as shown in Figure 24, we investigate the prospect

of long-term future forecasting, which, while not precise, is a significant tool for capturing groundwater dynamics and providing scientific guidance for decision-makers.

e) *Hyperparameters tuning*: In this study, the hyperparameters tuning for the deep learning algorithms such as CNN, and LSTM were found using cross-validation using Bayesian optimization techniques based on Python's Keras tuning library. The optimization range of values and hyperparameter sets were set up as depicted in Table I.

TABLE I: Hyperparameters of the deep learning models and the optimization range of values.

Model	Hyperparameter	Optimization range of values
LSTM	Number of hidden layers	$\{1, 2, \dots, 10\}$
	Number of nodes per layer	$\{5, 10, \dots, 150\}$
	Time lag (month)	$\{1, 2, \dots, 12\}$
	Dropout rate	$\{0.1, 0.2, \dots, 0.5\}$
	Batch size	$\{16, 32, \dots, 128\}$
	Epoques	$\{1, 2, \dots, 100\}$
	Seq length	$\{1, 2, \dots, 52\}$
CNN	Number of filters	$\{16, 32, \dots, 128\}$
	Number of hidden layers	$\{1, 2, \dots, 10\}$
	Number of nodes per layer	$\{1, 10, \dots, 100\}$
	Time lag (month)	$\{1, 6, \dots, 12\}$
	Dropout rate	$\{0.1, 0.2, \dots, 0.5\}$
	Batch size	$\{5, 10, \dots, 128\}$
	Epoques	$\{1, 10, \dots, 100\}$
	Dense Size	$\{1, 2, \dots, 256\}$
	Seq length	$\{1, 2, \dots, 52\}$
	Number of filters	$\{1, 2, \dots, 256\}$

Table II shows the most performance measures for both



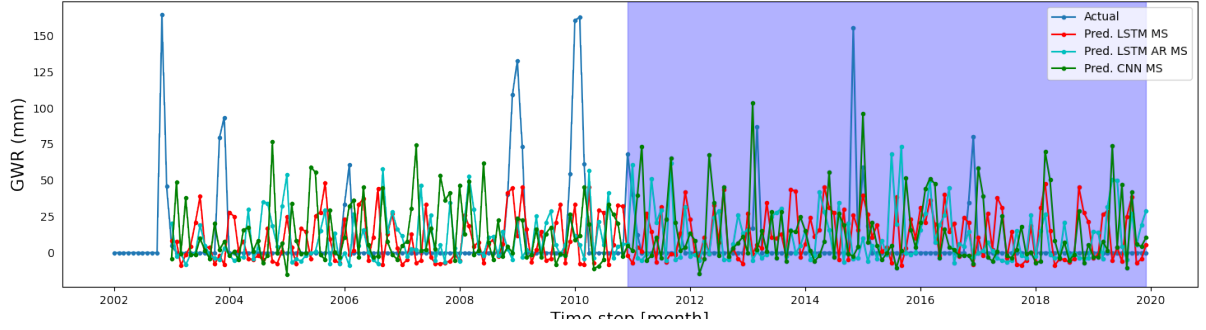


Fig. 19: GWR time-series comparisons Seq2One Multivariate forecasting by the LSTM AutoRegressive (red dotted line) Multivariate model with in situ measurements (blue line) CNN green Line, corresponding to TMPA precipitation.

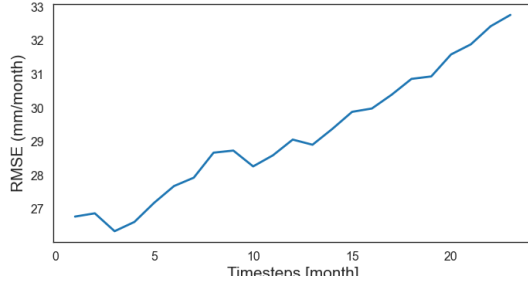


Fig. 20: RMSE according to the number of timesteps of the past to take into consideration to predict the future value of GWR

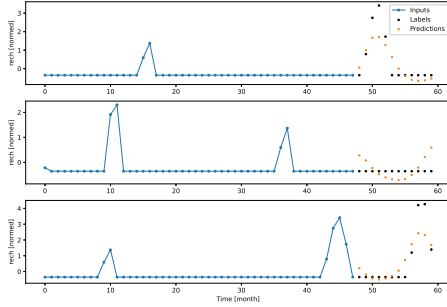


Fig. 21: Example of data windows of GWR variable, blue dotted line as inputs, black dots as actual labels, and orange dots as forecasted values:  $Window(input = 36, labels = 12, shift = 1)$ .

techniques with tuned hyperparameters, which favor the LSTM technique over CNN.

### C. Hypothesis validation with the in situ information

Figure 23a displays the annual precipitation average in millimeters per year (mm/year), while Figure 23b displays the annual recharge average in millimeters per year (mm/year). While recharge varies between 0 mm/y and 500 mm/y, precipitation rates range from 500 mm/y to 1500 mm/y. The blue dots in Figure 23a indicate a rate of precipitation of about 900 mm/year; we observed consistency in Figure 23b, and the green points indicate a rate of recharge of about 300 mm/year, and so on for all points pairwise comparisons between Figures 23a and 23b. We can observe that the rates

TABLE II: Test error metrics obtained using different AI-based algorithms with tuned hyperparameters

Model	Error metric	Value
LSTM	NSE	-0.87
	$R^2$	0.10
	RMSE	20.05
	rRMSE	12.19
	Bias	15.13
	rBias	9.19
	PI	-0.43
CNN	NSE	0.05
	$R^2$	0.09
	RMSE	22.31
	rRMSE	13.56
	Bias	2.80
	rBias	1.70
	PI	-0.07

of recharge (Fig. 23b) and precipitation (Fig. 23a) are completely correlated. Recharge increases in direct proportion to precipitation, and vice versa. This conclusion backs up the previous one that was reached using temporal data. We can infer from this that precipitation and recharge are spatially correlated. Consequently, there is a spatiotemporal regression between precipitation and recharging.

According to the graph, Sidi Yahia Bni Zeroual's (place 1) annual recharge is approximately twice as high as Ourtzagh's (place 2). The outcome also demonstrates a high degree of yearly recharge variability, with values in places 1 and 2 ranging from 0mm/year to 430mm/year and 0mm/year to 220mm/year, respectively. The result also demonstrates that there has been no recharging in the two locations over the last three years. We observed during our visit to the two locations that, despite the first location receiving more recharge due to its higher precipitation, the vitality of the trees has decreased and the habitats are suffering from thirst and well depletion, whereas the second location, despite receiving less recharge and precipitation, has healthy trees and wells that are still filled with water. Because of the changing climate, Morocco and the rest of the world are experiencing lower precipitation. Given the established spatiotemporal relationship between groundwater recharge and precipitation, we can conclude that groundwater is a limited resource, and uncontrolled depletion of this priceless substance poses a serious risk to the sustainability of life. The scenario was as follows, according to the data we have gathered from the locals of

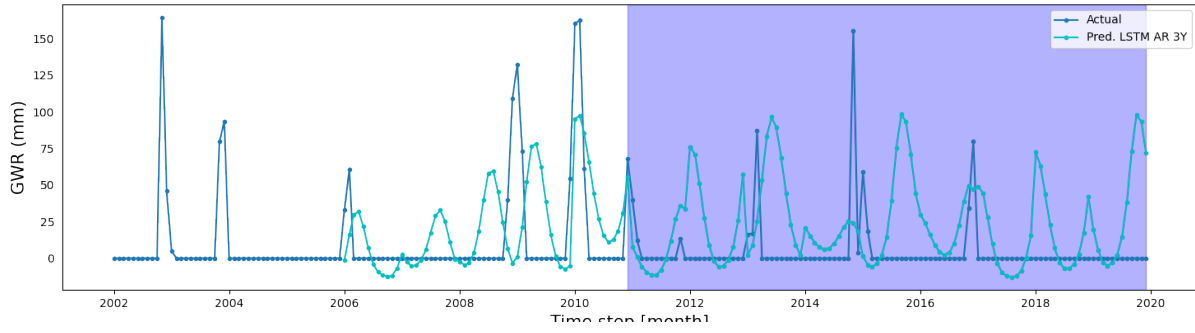


Fig. 22: Multivariate forecasting of one year(12 months) in the future based on 3 years in the past.

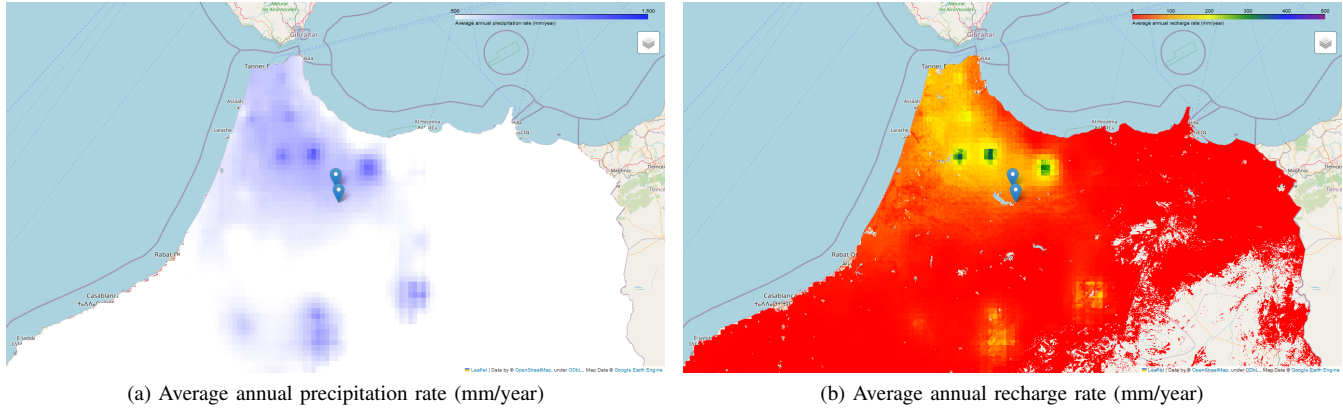


Fig. 23: Average annual of precipitation and recharge rates (mm/year), including two areas of study, in north of Morocco

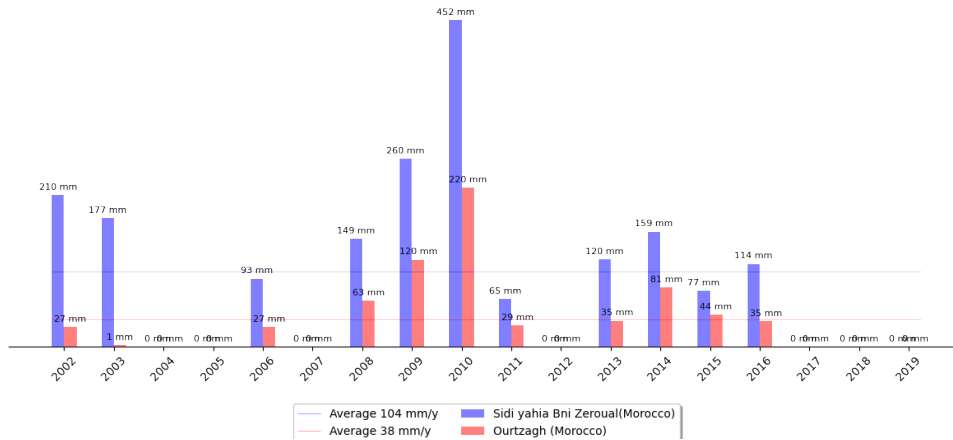


Fig. 24: Groundwater recharge comparison between the two places

the area: Around 20 years ago, people used surface water for agriculture and the water came from mountains. Agriculture consisted of subsistence crops such as wheat, vegetables, and melons. Since the year 2000, residents of location 1 have switched from subsistence farming to cannabis farming, which is a spring-summer culture with little precipitation. Farmers began digging wells illegally and without authorization, and as people realized that each well meant a loss of a source of water flow from mountains, more wells were drilled until there was no longer a source of water flow from mountains, and the locals begin drinking water from wells. When the water in the wells first started to run out, they were around 60 meters

deep. Farmers then dug new, deeper wells, and so on. More than 50 wells of various depths were now counted in the area of  $6 \text{ km}^2$  in place 1. People in that area suffer from thirst, so the local authority brings drinking water pipes from another location because there are only two wells with a high deep depth and little water available in Place 1. We were informed by locals during our tour that many families had left the region and moved to large cities like Tanger, Fes, and Casablanca.

## V. CONCLUSION, PERSPECTIVE, AND RECOMMANDATION

Physical modeling of the GWSC requires extensive data for numerous variables, and highly skilled researchers, and is

a nonlinear and demanding undertaking. A viable alternative to traditional models for capturing complicated interactions between GWR-related factors is artificial intelligence (AI). Additionally, it is simple to include satellite data for multiple factors to forecast spatiotemporal GWR with AI models. In this study, several feature combinations from multi-satellite data were used to forecast GWR across Morocco accurately using CNN and LSTM models, including different variations of residual and autoregressive models. Different input variables (precipitation, average evapotranspiration, changes in soil moisture storage, AWPL, and ST) were used in the deep learning models. To prevent overfitting issues, it is important to comprehend the sensitivity of their predictive performance. Different parameter combinations were used for hyperparameter optimization based on Bayesian optimization, and the combinations' prediction accuracies were compared. Overall, the findings demonstrated that, in terms of the RMSE, the LSTM autoregressive model predicted GWR maps somewhat better than the other models described in this study.

As was previously noted, managing groundwater involves keeping track of a variety of processes, including recharge, storage, withdrawal, depth level, etc. The first natural process is a recharge, in which freshwater seeps into the earth to restock subsurface aquifers. The other processes mentioned above follow and are the focus of our following studies.

Active monitoring of groundwater recharge, storage, and withdrawals is still restricted to just a few places worldwide, despite the growing global drive toward sustainable groundwater management techniques and water security in general. In this work, we effectively illustrated the applicability and adaptability of our AI and remote sensing methods-based approach, which might help water managers achieve their goals for water management. In Morocco, agriculture is becoming increasingly reliant on irrigation, and while we recognize the importance of agriculture to the national economy and its value in providing food for ecosystems, rising withdrawal rates in unmanaged basins are dangerous and pose a threat to the national reserve.

#### ACKNOWLEDGMENT

The authors would like to thank the Remote Sensing Frameworks for making datasets available to the public for free.

#### REFERENCES

- [1] S. Majumdar, R. Smith, B. D. Conway, and V. Lakshmi, "Advancing remote sensing and machine learning-driven frameworks for groundwater withdrawal estimation in arizona: Linking land subsidence to groundwater withdrawals," *Hydrological Processes*, vol. 36, no. 11, p. e14757, 2022.
- [2] A. Wunsch, T. Liesch, and S. Broda, "Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (lstm), convolutional neural networks (cnns), and non-linear autoregressive networks with exogenous input (narx)," *Hydrology and Earth System Sciences*, vol. 25, no. 3, pp. 1671–1687, 2021. [Online]. Available: <https://hess.copernicus.org/articles/25/1671/2021/>
- [3] J. Y. Seo and S.-I. Lee, "Predicting changes in spatiotemporal groundwater storage through the integration of multi-satellite data and deep learning models," *IEEE Access*, vol. 9, pp. 157 571–157 583, 2021.
- [4] A. Wunsch, T. Liesch, and S. Broda, "Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (lstm), convolutional neural networks (cnns), and non-linear autoregressive networks with exogenous input (narx)," *Hydrology and Earth System Sciences*, vol. 25, no. 3, pp. 1671–1687, 2021.
- [5] S. Majumdar, R. Smith, J. Butler Jr, and V. Lakshmi, "Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning," *Water Resources Research*, vol. 56, no. 11, p. e2020WR028059, 2020.
- [6] J. Grigsby, Z. Wang, N. Nguyen, and Y. Qi, "Long-range transformers for dynamic spatiotemporal forecasting," *arXiv preprint arXiv:2109.12218*, 2021.
- [7] G. B. Senay, S. Bohms, R. K. Singh, P. H. Gowda, N. M. Velpuri, H. Alemu, and J. P. Verdin, "Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the sseb approach," *JAWRA Journal of the American Water Resources Association*, vol. 49, no. 3, pp. 577–591, 2013.
- [8] K. E. Saxton and W. J. Rawls, "Soil water characteristic estimates by texture and organic matter for hydrologic solutions," *Soil science society of America Journal*, vol. 70, no. 5, pp. 1569–1578, 2006.
- [9] H. Yoon, S.-C. Jun, Y. Hyun, G.-O. Bae, and K.-K. Lee, "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer," *Journal of hydrology*, vol. 396, no. 1–2, pp. 128–138, 2011.
- [10] T. Zhou, F. Wang, and Z. Yang, "Comparative analysis of ann and svm models combined with wavelet preprocess for groundwater depth prediction," *Water*, vol. 9, no. 10, p. 781, 2017.
- [11] X. Huang, L. Gao, R. S. Crosbie, N. Zhang, G. Fu, and R. Doble, "Groundwater recharge prediction using linear regression, multi-layer perceptron network, and deep learning," *Water*, vol. 11, no. 9, p. 1879, 2019.
- [12] S. A. Naghibi, K. Ahmadi, and A. Daneshi, "Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping," *Water Resources Management*, vol. 31, pp. 2761–2775, 2017.
- [13] D. Liu, A. K. Mishra, Z. Yu, H. Lü, and Y. Li, "Support vector machine and data assimilation framework for groundwater level forecasting using grace satellite data," *Journal of Hydrology*, vol. 603, p. 126929, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169421009793>
- [14] M. Panahi, N. Sadhasivam, H. R. Pourghasemi, F. Rezaie, and S. Lee, "Spatial prediction of groundwater potential mapping based on convolutional neural network (cnn) and support vector regression (svr)," *Journal of Hydrology*, vol. 588, p. 125033, 2020.
- [15] J. Y. Seo and S.-I. Lee, "Predicting changes in spatiotemporal groundwater storage through the integration of multi-satellite data and deep learning models," *IEEE Access*, vol. 9, pp. 157 571–157 583, 2021.
- [16] T. Steenhuis and W. Van der Molen, "The thornthwaite-mather procedure as a simple engineering method to predict recharge," *Journal of hydrology*, vol. 84, no. 3–4, pp. 221–229, 1986.
- [17] C. W. Thornthwaite, "Instructions and tables for computing potential evapotranspiration and the water balance," *Publications on Climatology*, vol. 10, pp. 185–310, 1957.
- [18] S. Majumdar, R. Smith, J. J. Butler, and V. Lakshmi, "Groundwater Withdrawal Prediction Using Integrated Multitemporal Remote Sensing Data Sets and Machine Learning," *Water Resources Research*, vol. 56, no. 11, p. e2020WR028059, Nov. 2020.
- [19] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.
- [20] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell *et al.*, "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," *Scientific data*, vol. 2, no. 1, pp. 1–21, 2015.
- [21] T. Hengl and S. Gupta, "Soil water content (volumetric%) for 33kpa and 1500kpa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (version v01)," *Zenodo*. Available at: <https://zenodo.org/Accessed 06 April 2022>, 2019.
- [22] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [23] F. Chollet *et al.* (2015) Keras. [Online]. Available: <https://github.com/fchollet/keras>
- [24] R. Vallat, "Pingouin: statistics in python." *J. Open Source Softw.*, vol. 3, no. 31, p. 1026, 2018.

- [25] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: <https://doi.org/10.1038/s41586-020-2649-2>
- [26] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [28] K. Jordahl, J. V. den Bossche, M. Fleischmann, J. Wasserman, J. McBride, J. Gerard, J. Tratner, M. Perry, A. G. Badaracco, C. Farmer, G. A. Hjelle, A. D. Snow, M. Cochran, S. Gillies, L. Culbertson, M. Bartos, N. Eubank, maxalbert, A. Bilogur, S. Rey, C. Ren, D. Arribas-Bel, L. Wasser, L. J. Wolf, M. Journois, J. Wilson, A. Greenhall, C. Holdgraf, Filipe, and F. Leblanc, "geopandas/geopandas: v0.8.1," Jul. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3946761>
- [29] T. pandas development team, "pandas-dev/pandas: Pandas," Feb. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3509134>
- [30] python visualization, "Folium." [Online]. Available: <https://python-visualization.github.io/folium/>
- [31] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: <https://doi.org/10.21105/joss.03021>
- [32] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [33] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 785–794. [Online]. Available: <http://doi.acm.org/10.1145/2939672.2939785>
- [34] QGIS Development Team, *QGIS Geographic Information System*, Open Source Geospatial Foundation, 2009. [Online]. Available: <http://qgis.osgeo.org>
- [35] G. Attard, "Groundwater recharge estimation using earth engine."
- [36] R. G. Allen, L. S. Pereira, D. Raes, M. Smith *et al.*, "Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56," *Fao, Rome*, vol. 300, no. 9, p. D05109, 1998.
- [37] Y. Yang, R. J. Donohue, and T. R. McVicar, "Global estimation of effective plant rooting depth: Implications for hydrological modeling," *Water Resources Research*, vol. 52, no. 10, pp. 8260–8276, 2016.
- [38] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [39] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.
- [40] T. Chen, T. He, M. Benesty, and V. Khotilovich, "Package 'xgboost,'" *R version*, vol. 90, pp. 1–66, 2019.
- [41] J. Grigsby, Z. Wang, and Y. Qi, "Long-range transformers for dynamic spatiotemporal forecasting," 2021.
- [42] A. Laouina, "Prospective "maroc 2030": Gestion durable des ressources naturelles de la biodiversité au maroc," *Rapport pour le compte du HAUTCOMMISSARIAT AU PLAN ROYAUME DU MAROC*, 2006.
- [43] C.-S. J. Chu, "Time series segmentation: A sliding window approach," *Information Sciences*, vol. 85, no. 1, pp. 147–173, 1995. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S002002559500021G>
- [44] T. chung Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952197610001727>



**Azeddine Elhassouny** (IEEE Member) He received the M.S. and Ph.D. degrees in Computer Science from Ibn Zohr University, Agadir, Morocco, in 2008 and 2013, respectively. He was awarded a Fulbright Scholarship to work as a postdoctoral researcher in AI and computer vision at the University of New Mexico in the United States. He defended his HDR in Artificial intelligence at ENSIAS College of Engineering, Mohammed V University, Rabat, in 2019. In 2014, he joined the National High School, IT/ENSIAS College of Engineering, Mohammed V University, Rabat, where currently he is an associate Professor. He supervised many Ph.D. and master's students at Mohammed V University. So far, his research efforts have culminated in, three books, numerous papers in a wide variety of international conferences and journals. His current research interests include Artificial intelligence, Computer vision and Multi-criteria decision making. He has been involved as a session chair, TPC member in various conferences and journals.