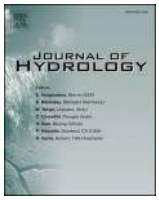
Journal of Hydrology 627 (2023) 130359

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Artificial intelligence models to evaluate the impact of climate change on groundwater resources

Daniele Secci [a,b,*], Maria Giovanna Tanda [a], Marco D'Oria [a], Valeria Todaro [a]

[a] Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy
[b] Instituto Universitario de Investigación de Ingeniería del Agua y Medio Ambiente, Universitat Politècnica de València, 46022 Valencia, Spain

ARTICLE INFO

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Gokmen Tayfur, Associate Editor

Keywords:
Artificial neural network
NARX
LSTM
CNN
Groundwater resources
Climate change

ABSTRACT

This study develops three different artificial intelligence (AI) models in order to investigate the effects of climate change on groundwater resources using historical records of precipitation, temperature and groundwater levels together with regional climate projections. In particular, the Non-linear Autoregressive Neural Network (NARX), the Long-Short Term Memory Neural Network (LSTM) and the Convolutional Neural Network (CNN) were compared. Considering an aquifer located in northern Italy as a case study, the neural networks were trained to replicate observed groundwater levels by taking as input precipitation and temperature records, and in the case of the NARX also antecedent groundwater levels, on a monthly scale. The trained networks were used to infer groundwater levels until the end of the century based on precipitation and temperature projections provided by an ensemble of 13 Regional Climate Models (RCMs) from the EURO-CORDEX initiative. Two emission pathways were considered: the RCP4.5 and RCP8.5. All the AI models show good performance metrics during the training phase, but NARXs perform poorly compared to the other models during validation and testing. For the future, the NARX and LSTM models predict a decline in groundwater levels, especially for the RCP8.5 scenario, while slight changes are expected using the CNN. As NARXs are not deep learning techniques and CNNs may not be able to extrapolate values outside the training range, LSTMs appear to be better suited for climate change impact evaluations.

1. Introduction

Climate change is affecting all regions of the world underlining the urgent need to assess its impacts on various aspects of the environment, society and economy, including the future availability of good quality freshwater. Surface water resources are already facing qualitative and quantitative issues, which means that groundwater is frequently the primary source of water supply. This can lead to unwise overexploitation of aquifers that, compounded by the direct impact of climate change on aquifer recharge, jeopardizes their quantity and quality.

The aforementioned issues have prompted several studies aimed at quantifying the climate change effects on groundwater resources. This task is extremely challenging, especially relying on a conventional approach that requires the use of complex hydrological models driven by future climate model projections. Developing an accurate numerical model of the surface/subsurface system necessitates a multitude of data that may not always be available, including the geological structure, model parameters, boundary and initial conditions. In addition, the computational demand of this approach can be significant and may result in limiting the analysis to short periods and only few scenarios. In fact, to assess the impacts of climate change on groundwater, future projections of meteorological variables provided by several climate models under different scenarios are usually employed. Malcolm and Soulsby (2000) evaluated the possible influence of climate change on a shallow coastal aquifer in northern Scotland using a numerical MODFLOW model. Different climate change scenarios were analyzed to determine future variations in climate variables (e.g., precipitation and temperature) and their effects on groundwater levels (GWLs). A similar approach was used by Croley and Luukkonen (2003) for the Lansing area in Michigan. Brouyère et al. (2004) made use of an integrated hydrological model (MOHISE) driven by climate future projections of three General Circulation Models (GCMs) to assess the impacts of climate change on the groundwater availability in the Geer basin, Belgium. Citrini et al. (2020) assessed the impacts of climate change on the discharges of a karst spring using the GR4J model and the data of three climate models under different scenarios (namely Representative

* Corresponding author at: Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy.
E-mail address: daniele.secci@unipr.it (D. Secci).

https://doi.org/10.1016/j.jhydrol.2023.130359
Received 12 January 2023; Received in revised form 16 May 2023; Accepted 9 October 2023
Available online 29 October 2023

Concentration Pathways, RCPs: RCP2.6, RCP4.5 and RCP8.5). Azizi et al. (2021) developed a MODFLOW model for the Varamin plain in Iran to evaluate the change in GWLs until 2050 using as climate drivers the data of ten Atmosphere-Ocean GCMs, under the RCP2.6, RCP4.5 and RCP8.5 scenarios.

Surrogate models, by replacing complex and time consuming numerical models with computationally less expensive numerical tools, can provide simpler and faster responses (Razavi et al., 2012; Asher et al., 2015; Rajaee et al., 2019) and can help in various groundwater management problems. For example, surrogate models can be used to forecast groundwater levels based on a set of drivers or proxy variables, such as precipitation, temperature and withdrawals. To accomplish this task, they just require a preliminary training process that makes use of available historical data of the drivers (e.g., precipitation and temperature) and corresponding responses (e.g., groundwater levels). Most surrogate models are categorized as either statistical approaches or machine learning algorithms, which have become increasingly popular in recent decades, thanks to the enhancement of computing capabilities.

Several authors (e.g. Khan et al, 2008; Bloomfield and Marchant, 2013; Kumar et al, 2016; Leelaruban et al, 2017; Soleimani Motlagh et al, 2017; Van Loon et al, 2017; Uddameri et al, 2019; Guo et al, 2021) used statistical approaches to investigate the relationships between groundwater levels and different proxy variables. In addition, space–time modeling and spatiotemporal geostatistics are also used to analyze these relationships (e.g. Bierkens et al., 2001; Varouchakis and Hristopulos, 2013). In many applications, it is common to use standardized indices to represent the variables of interest, such as the standardized precipitation index (SPI), the standardized precipitation-evapotranspiration index (SPEI) and the standardized groundwater index (SGI). As an example, Secci et al. (2021) analyzed the correlation between SGIs and both SPIs and SPEIs for several wells in northern Tuscany (Italy), revealing that the groundwater index and the meteorological indices were often correlated. Linear regression models were used to establish relationships between SGIs and SPIs or SGIs and SPEIs, which were then used to forecast future SGIs based on projections of the meteorological indices derived from precipitation and temperature data obtained from an ensemble of Regional Climate Models (RCMs) under the RCP4.5 and RCP8.5 scenarios. The authors pointed out that changes in SGIs are mainly driven by temperature, which is expected to undergo a significant increase in the future, rather than precipitation that shows limited variation for the investigated area.

Also machine-learning algorithms can be used to capture the relationships between groundwater levels and meteorological data, see Rajaee et al. (2019) and Tao et al. (2022) for extensive reviews. Several studies employed different machine learning techniques to evaluate piezometric levels over a historical period using a subset of the available observations. These techniques include artificial neural networks (Coppola et al., 2003; Lallahem et al., 2005; Mohanty et al., 2010; Trichakis et al., 2011; Karthikeyan et al., 2013; Sahoo and Jha, 2013; Shiri et al., 2013; Taormina et al., 2012; Emamgholizadeh et al., 2014; Chang et al., 2015; Mohanty et al., 2015; Yan and Ma, 2016; Ghose et al., 2018; Lee et al., 2019), neuro-fuzzy systems (ANFIS, Fallah-Mehdipour et al., 2013; Shiri et al., 2013; Emamgholizadeh et al., 2014) and support vector machines (Yoon et al., 2011; Shiri et al., 2013; Suryanarayana et al., 2014). In some cases, precipitation is the sole input feature, while other applications consider additional variables such as temperature, humidity, runoff and evapotranspiration. Tidal levels were used as input data for coastal aquifers, while antecedent groundwater levels, pumping rates and water demand were included in other studies.

In recent years, some studies investigated the impact of climate change on groundwater levels making use of machine learning techniques in combination with future climate projections (Chen et al., 2010; Shakiba and Cheshmi, 2013; Chang et al., 2015; Jeihouni et al., 2019; Idrizovic et al., 2020; Javadinejad et al., 2020; Ghazi et al., 2021; Gonzalez and Arsanjani, 2021). Afrifa et al. (2022) provide a review of mathematical and machine learning models implemented to examine the effects of climate change on groundwater level fluctuations. Shakiba and Cheshmi (2013) used the output of one GCM under the AR2 scenario to simulate variations in groundwater levels through a non-linear autoregressive network (NARX). However, the use of a single climate model is highly discouraged (IPCC, 2018) due to the significant uncertainty of the results. Chen et al. (2010) investigated the effect of climate change and human activities on shallow groundwater levels in Wuqiao in North China Plain using projections from 20 GCMs. The authors trained a Back-Propagation Artificial Neural Network (BP-ANN) with observed meteorological and pumping rates data to replicate groundwater levels. The ensemble mean of the climate models was considered to project future climate variables and to simulate deviations in groundwater levels with the trained network. Chang et al. (2015) developed two ANNs to simulate and predict suprapermafrost groundwater levels. The models were trained using historical information of antecedent groundwater levels, temperature and precipitation as input or only precipitation and temperature. The authors then assessed the impact of climate change on groundwater levels by analyzing different scenarios of precipitation and temperature increases. Idrizovic et al. (2020) investigated the possible influence of climate change on the Toplica River catchment in Serbia. Historical precipitation, temperature and potential evapotranspiration data were used to simulate runoff via a calibrated hydrological model (HBV-light). Then, an ANN was trained to reproduce groundwater levels using runoff values as input. An ensemble of seven RCMs, from the EURO-CORDEX initiative (Jacob et al., 2014), was considered to compute precipitation and temperature projections and to simulate future values of runoff. The trained ANN was finally used to estimate future groundwater levels under the RCP4.5 and RCP8.5 scenarios. Ghazi et al. (2021) studied groundwater level fluctuations for the Tasuj plain, Iran, under climate change scenarios comparing different machine learning techniques: ANN, SVM and NARX. Four GCMs were selected to compute future projections of precipitation and temperature for the period 2022–2050 under three RCP scenarios (RCP2.6, RCP4.5 and RCP8.5). According to the results, NARX exhibited the best accuracy in predicting observed groundwater levels.

The above-mentioned machine learning techniques, despite their usefulness for estimating groundwater levels, are supervised and therefore are limited in their ability to extrapolate information beyond the range of the training data. As global warming and extreme weather events are expected to increase in the future (Jiménez Cisneros et al., 2015), the input data for machine learning models (such as, precipitation and temperature) can fall outside the range of the learning dataset, leading to inaccurate results. Innovative deep learning techniques were proposed to address the extrapolation problem. In particular, Long-Short Term Memory (LSTM) neural networks, a type of recurrent neural networks, were developed to handle sequential time series prediction problems, but very few papers report the use of deep learning models to predict groundwater levels. Nourani et al. (2022) compared the LSTM neural network with a classic feedforward neural network and an auto regressive integrated moving average model with exogenous data to estimate groundwater level time series in a historical period. The results highlighted that LSTMs outperforms the other two methodologies. Gharehbaghi et al. (2022) tested three different layer structures of meteorologically sensitive Gated Recurrent Unit (GRU) neural networks, which are a variant of the LSTMs, to estimate groundwater levels in a historical period based on observed meteorological and hydrological data. The results showed a great potential of these deep learning methodologies for the evaluation of future groundwater levels. Wunsch et al. (2022) investigated the impacts of climate change on groundwater resources implementing a deep learning method based on Convolutional Neural Networks for 118 well-distributed sites in Germany. The authors used historical meteorological input to train the networks for reproducing piezometric levels. Then, they used the trained CNN models to investigate future groundwater levels, using precipitation and temperature projections derived from different climate models under different RCP scenarios (RCP2.6, RCP4.5 and RCP8.5).

This paper attempts to fill the gap regarding applications of deep learning models in assessing the impacts of climate change on groundwater levels. Three methods based on artificial intelligence (AI) techniques are compared: a machine learning method (Non-Linear Autoregressive Neural Network, NARX) and two deep learning methods (Long Short-Term Memory neural network, LSTM, and Convolutional Neural Network, CNN). The MATLAB computing environment was utilized for the development of the three machine learning models.

The LSTM and the CNN networks take precipitation and temperature as inputs, while the NARXs consider also antecedent groundwater levels; in all cases the outputs are groundwater levels. The training and validation of the networks are done using historical meteorological and groundwater level time series. Once trained, the networks are driven by climate model projections to estimate future groundwater levels. The proposed AI techniques are applied to a regional area located in northern Italy. To account for inter-model variability, an ensemble approach is adopted (Teutschbein and Seibert, 2012; D'Oria et al., 2017; D'Oria et al., 2018; Todaro et al., 2022) by considering the output of 13 RCMs of the EURO-CORDEX initiative (Jacob et al., 2014). Only the two most relevant RCP scenarios, the RCP4.5 and RCP8.5, of the phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012) were used. A downscaling and bias correction process (Teutschbein and Seibert, 2012) of the RCM outputs was applied to preserve local heterogeneities.

The novelties of this paper are: i) the comparison of different AI techniques for assessing the impact of climate change on groundwater resources and, in particular, the innovative implementation of a sequential deep learning method, namely the LSTM neural network, in this field; ii) the use of neural networks in a long-term forecasting process where the network also operates beyond the ranges encountered during the training process; iii) the use of a large ensemble of regional climate projections under different scenarios, providing a comprehensive characterization of the uncertainty of the results.

This paper is organized as follows: Section 2 presents the study area, the available historical data, the climate models adopted and the methodologies implemented. The results of the AI techniques are shown in Section 3 and discussed in Section 4; conclusions are drawn in Section 5.

2. Material and methods

2.1. Study area and available data

The investigated area is the northern part of the Tuscany region in Italy (Fig. 1) and involves four river basins. The main characteristics of the study area have already been presented in previous studies (D'Oria et al., 2017; D'Oria et al., 2019; Secci et al., 2021) to which reference is made for further details. According to the different characteristics of the watersheds, the area was divided into four zones (Fig. 1): Magra, Coastal basins, Serchio, Arno portion. The precipitation, temperature and groundwater level data used in this study are the same analyzed by Secci et al. (2021).

Daily climate data were available from 18 precipitation gauges and 14 temperature stations (Fig. 1 and Table 1) in the period 1934–2020. The groundwater level data were available from 10 wells (Fig. 1 and Table 1) in the period 2005–2020 on a daily scale. The data are published by the Environmental Agency of the Tuscany, Liguria and Emilia Romagna regions (SIR, 2021; ARPAE, 2021; OMIRL, 2021). To fill the gaps in the observed time series, the FAO method (Allen et al., 1998) was considered and interpolation techniques were used to have contemporary records of temperature and precipitation at the same locations (Secci et al., 2021).

The annual precipitation and annual mean temperature over the four basins in the historical period 1934–2020 are summarized in Table 2. The Magra, Coastal basins and Serchio basins present similar climate in the analyzed period; the average annual mean temperature is about 13 °C with variability in the range of about 11 ÷ 15 °C and the annual precipitation is about 1500 mm with a natural variability in the range 800 ÷ 2600 mm. The warmest and driest area is the Arno portion with an average annual mean temperature of about 15 °C and annual

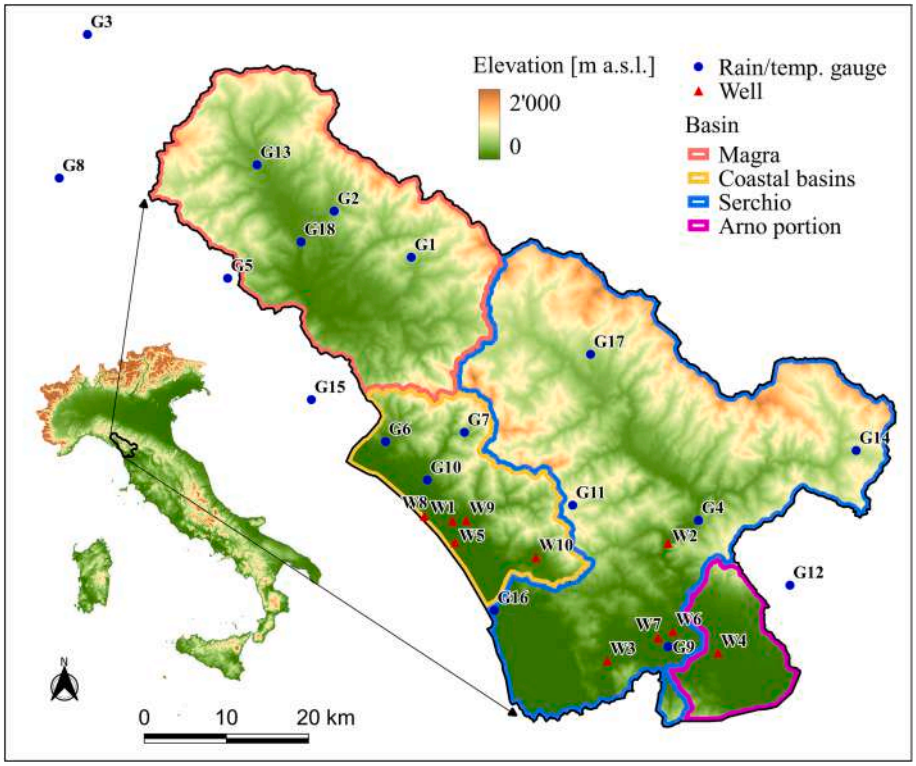


Fig. 1. Location of the study area with indication of the climate stations, monitoring wells and river basins. Elevation is expressed in meters above sea level (m a.s.l.).

Table 1
ID, name, type and elevation of the precipitation and temperature gauges and ID, name and ground elevation of the monitoring wells.

Temperature and precipitation gauges				Monitoring wells		
ID	Name	Type	Elevation (m a.s.l.)	ID	Name	Elevation (m a.s.l.)
G1	Arlia	Rain/Temp.	460	W1	Cugnia	4
G2	Bagnone	Rain/Temp.	195	W2	Diecimo	65
G3	Bedonia	Rain/Temp.	500	W3	Nozzano	16
G4	Borgo a Mozzano	Rain	100	W4	Paganico	13
G5	Calice al Cornoviglio	Rain/Temp.	402	W5	Percorso vita	2
G6	Carrara	Rain/Temp.	55	W6	Salicchi	27
G7	Casania	Rain	845	W7	S.Alessio	19
G8	Cembrano	Rain/Temp.	410	W8	Sat 1	2
G9	Lucca	Rain/Temp.	16	W9	Via Barsanti	20
G10	Massa	Rain/Temp.	150	W10	Via Romboni	38
G11	Palagnana	Rain	861			
G12	Pescia	Rain/Temp.	78			
G13	Pontremoli	Rain/Temp.	340			
G14	S. Marcello Pistoiese	Rain/Temp.	618			
G15	Sarzana	Rain/Temp.	26			
G16	Viareggio	Rain/Temp.	0			
G17	Villacollemandina	Rain	502			
G18	Villafranca Lunigiana	Rain/Temp.	156			

Table 2
Annual mean temperature and annual precipitation over the basins: average, maximum and minimum values in the period 1934–2020.

Annual mean temperature (°C)	**MAGRA**	**COASTAL BASINS**	**SERCHIO**	**ARNO PORTION**
Average	13.2	13.2	12.9	14.8
Max	14.8	14.8	14.3	16.0
Min	11.3	11.8	11.4	13.3
Annual precipitation (mm)	**MAGRA**	**COASTAL BASINS**	**SERCHIO**	**ARNO PORTION**
Average	1539	1578	1536	1205
Max	2608	2579	2650	2039
Min	810	803	825	444

precipitation of about 1200 mm.

The future daily precipitation and daily mean temperature data were obtained from an ensemble of 13 climate models from the EURO-CORDEX initiative (Jacob et al., 2014); they are combinations of different GCMs and RCMs (Table 3). The climate model data consist of a simulated historical period (1976–2005) and a projection period from 2006 to 2100, under two Representative Concentration Pathways (RCPs): the RCP4.5 and RCP8.5. The climate model data are provided on a regular grid (EUR-11 grid); therefore, a downscaling procedure was applied to obtain them at the gauging station locations. The data were bias corrected with reference to the historical period 1976–2005 using the Distribution Mapping method. For more details on the climate model data and the downscaling and bias correction methods adopted for the study area, see D'Oria et al. (2017).

Table 3
Combinations of GCMs and RCMs from the EURO-CORDEX initiative used in this study.

		GCM				
		CNRM-CM5	**EC-EARTH**	**HadGEM2-ES**	**MPI-ESM-LR**	**IPSL-CM5A-MR**
RCM	**CCLM4-8–17**	x	x	x	x	
	HIRHAM5		x			
	WRF331F					x
	RACMO22E		x	x		
	RCA4	x	x	x	x	x

2.2. Description of the AI methods

In the following paragraphs, the three neural networks developed in this study are presented. For all the networks, the exogenous input data are monthly precipitation and mean monthly temperature of the 18 climate stations reported in Table 1 (for a total of 36 features). The output dataset corresponds to monthly groundwater levels of the 10 wells reported in Table 1 (10 responses). The target data, used to build the models, span from March 2005 to December 2020 (190 months), when observed monthly groundwater levels are available. The precipitation and temperature data span from July 2004 to December 2020 (198 months), in order to consider potential delayed responses of groundwater level data to meteorological variables.

Both the input and target datasets were standardized (Z_i) in order to facilitate the process of updating parameters (weights and biases):

$$Z_i = \frac{z_i - \overline{z}}{\sigma_z} \tag{11}$$

where z_i is a single data point of the input/target vector, $\overline{z}$ is the arithmetic mean and σ_z is the standard deviation. The learning (training and validation) and test phases of the neural networks were based on a dataset that covers the period 2005–2018 (training 90 % and validation 10 %) and 2019–2020, respectively.

The network was trained in order to minimize the following Loss function $L(\Theta)$:

$$L(\Theta) = \frac{1}{N \cdot s_3} \sum_{i=1}^{N} \sum_{j=1}^{s_3} \left(h_\Theta \left(\boldsymbol{x}^{(i)} \right)_j - \boldsymbol{y}_j^{(i)} \right)^2 \tag{12}$$

where N is the number of the training input-target pairs associated with each well and corresponds to the number of time steps considered, s_3 is the number of wells, $h_\Theta \left(\boldsymbol{x}^{(i)} \right)_j$ is the output of the network related to the i-th input data $\boldsymbol{x}^{(i)}$ depending on all network parameters Θ and $\boldsymbol{y}_j^{(i)}$ is the vector target.

The AI models, once trained, were used to evaluate future groundwater levels for the period 1976–2095 using precipitation and temperature data provided by the 13 climate models as input.

2.2.1. Non-linear Autoregressive neural network

The Non-linear Autoregressive Neural Networks (NARXs) are an evolution of the feedforward neural networks in which the neurons transfer information not only in the forward direction but also in the backward direction through recurring cycles. In this way, the outputs of

the network become new inputs to the network itself, thus favoring the learning process and making these models suitable for working with time-series forecasting problems.

The NARXs maintain the same conceptual structure as feedforward networks, i.e., they are divided into three main layers: the input layer, the hidden layer and the output layer. These layers are Euclidean spaces and have dimensions s_1, s_2, s_3 (number of neurons of which they are composed), respectively. The dimension s_1 depends on the number of features, s_2 on the number of hidden units and s_3 on the number of outputs to be estimated. Conceptually, the operation of the neural network provides a composition of vector-valued functions such as pointwise operation and linear and non-linear activation functions (sigmoid/tangent transfer function and linear transfer function) capable of transferring information from the dimensional space $\mathbb{R}^{s_1}$ to $\mathbb{R}^{s_3}$. More specifically, the output predicted by NARX, which also uses exogenous variables as input, turns out to be:

$$\boldsymbol{h}(t+1) = S(\boldsymbol{x}(t+1), \boldsymbol{x}(t), \cdots, \boldsymbol{x}(t-n), \boldsymbol{h}(t), \boldsymbol{h}(t-1), \cdots, \boldsymbol{h}(t-m)) \tag{13}$$

where $\boldsymbol{h}(t+1)$ represents the response (e.g., groundwater levels) predicted by the network at time $t+1$, $\boldsymbol{x}(t+1), \boldsymbol{x}(t), \cdots, \boldsymbol{x}(t-n)$ represent the exogenous inputs (e.g., precipitations and temperatures), $\boldsymbol{h}(t), \boldsymbol{h}(t-1), \cdots, \boldsymbol{h}(t-m)$ are the target/output data (e.g., groundwater levels) used as given inputs, n and m represent the input delay and the feedback delay, respectively, while S is the structure of the network. This architecture can be of two types: open-loop (Series-Parallel Architecture) or closed-loop (Parallel Architecture). The open-loop network is implemented for the training phase as it directly uses the available observations, together with the exogenous inputs, as input information. The closed scheme is used in the prediction phase when observations are no longer available and therefore the output produced by the network also becomes an input. In detail, the initial information undergoes an affine transformation and then it is activated by a transfer function. The output $\boldsymbol{a}^{(2)}$ produced by the input and hidden layers is:

$$\boldsymbol{a}^{(2)} = f_1(\boldsymbol{IW}^{(1)}\boldsymbol{x}^{(1)} + \boldsymbol{b}^{(1)} + \boldsymbol{LW}^{(1)}\boldsymbol{h}^{(1)}) \tag{14}$$

where the superscript indicates the reference layer of the network, f_1 represents the tangent sigmoid transfer function, $\boldsymbol{IW}^{(1)} \in \mathbb{R}^{s_2 \times (s_1 \cdot n)}$ represents the weight matrix related to exogenous inputs, $\boldsymbol{x}^{(1)} \in \mathbb{R}^{s_1 \cdot n}$ is the exogenous input vector, $\boldsymbol{b}^{(1)} \in \mathbb{R}^{s_2}$ is the bias term vector, $\boldsymbol{LW}^{(1)} \in \mathbb{R}^{s_2 \times (s_3 \cdot m)}$ is the matrix of the input weights relative to the observations and $\boldsymbol{h}^{(1)} \in \mathbb{R}^{s_3 \cdot m}$ is the observation vector used as given input. The output is then transferred from the hidden layer to the output layer. This process is described by:

$$\boldsymbol{h}_{w,b}(\boldsymbol{x}, \boldsymbol{h})^{(3)} = f_2(\boldsymbol{OW}^{(2)}\boldsymbol{a}^{(2)} + \boldsymbol{b}^{(2)}) \tag{15}$$

where f_2 represents the linear transfer function, $\boldsymbol{OW}^{(2)} \in \mathbb{R}^{s_3 \times s_2}$ is the matrix of the weights of the outputs of the hidden layer $\boldsymbol{a}^{(2)} \in \mathbb{R}^{s_2}$, $\boldsymbol{b}^{(2)} \in \mathbb{R}^{s_3}$ is the bias term vector. The results of the network $\boldsymbol{h}_{w,b}(\boldsymbol{x}, \boldsymbol{h})^{(3)}$, i.e., the predicted value at time t, are sent back to the initial layer as a new input to the network at time $t+1$ for the closed loop structure. Fig. 2 shows the structure of the closed-loop type NARX. The Levenberg-Marquardt numerical optimization technique (Hagan and Menhaj, 1994) was selected as training algorithm to estimate the correction of the weights. It is based on the backpropagation technique, where derivatives are processed from the last layer to the first, and it is widely used as it does not require the calculation of the Hessian matrix, leading to a reduction in terms of computational burden.

In this work, the number of training epochs was set equal to 50, with a mini-batch size of 18 and a learning rate equal to 0.001. The network was trained in open-loop mode using historical precipitation and temperature data and observed groundwater levels. Then, precipitation and temperature projections from the ensemble of climate models were used as input to the NARX to estimate future groundwater levels. In the prediction phase, the network was used according to the closed-loop scheme. The input delay n was set equal to 9, in agreement with the findings of Secci et al. (2021), where the maximum correlation between SGI and SPEI occurs for an accumulation period of 9 months. This means that, in order to predict the groundwater levels at time $t+1$, precipitation and temperature data spanning from time $t-8$ to time t are considered exogenous inputs. The feedback delays (m) were set to 2 due to the rapid response time of the aquifer highlighted by Secci et al. (2021). Therefore, it was assumed that the predicted levels at time $t+1$ are related to the groundwater levels at times t and $t-1$. At each time step, the size of the input matrix is 36×9, where 9 are the monthly values given by the input delay and 36 are the features represented by the monthly values of precipitation and temperature provided by the 18 climate stations. The input matrix is processed by one hidden layer, made up of 10 neurons. The size of the output vector is 10, corresponding to the number of wells. The total number of time steps is 190, which corresponds to the time series length of groundwater levels.

2.2.2. Long-Short term Memory network

Long-Short Term Memory networks, hereafter LSTMs, are an evolution of recurrent neural networks (Hochreiter and Schmidhuber, 1997). They are designed for handling long time series, as their structure is not vulnerable to the so-called vanishing gradient problem (Hochreiter, 1991), and are effective in learning long-term dependencies. LSTMs are schematized within a deep-learning architecture composed of various layers. The first layer is the Sequence Input Layer, which enters sequence data into the network. The second layer is the LSTM layer, which is the main layer of the deep learning chain,

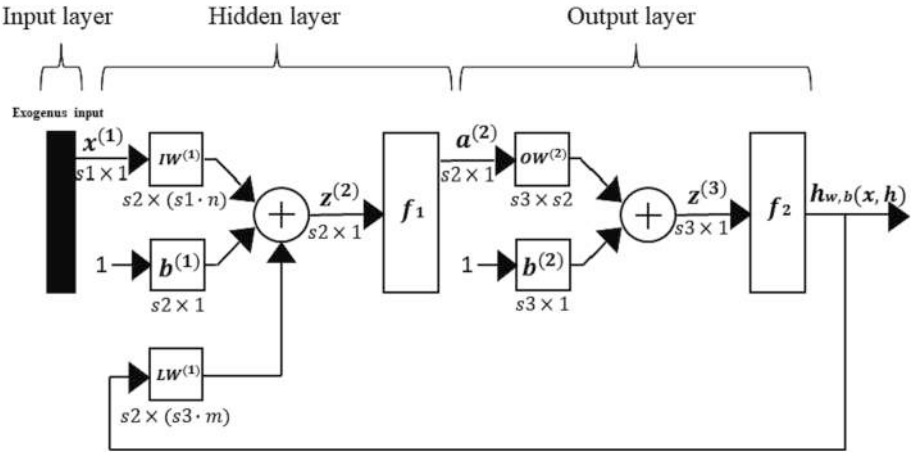


Fig. 2. NARX architecture (closed loop).

represented by the cell unit (Fig. 3). This cell is the core of the mechanism that allows LSTMs to be used for handling sequential prediction problems. The cornerstone of the cell unit is the Cell state $\boldsymbol{C}(t)$, which can be considered as the brain of the network where important information is stored. A classic LSTM layer has three main gates: forget gate layer (f_t), input gate layer (i_t), output gate layer (o_t). The first two gates work to update the Cell state, while the third, together with the updated Cell state, is used to produce the output of the LSTM layer, which is the hidden state at the current time t. Considering the three Euclidian spaces of dimension s_1, s_2 and s_3, the individual input vector of dimension s_1 is given by the concatenation of the exogenous input $\boldsymbol{x}(t)$ at time t and the hidden state vector $\boldsymbol{h}(t-1)$ at time $t-1$. For each gate, the matrix of weights $\boldsymbol{W} \in \mathbb{R}^{s_2 \times s_1}$ and the bias term $\boldsymbol{b} \in \mathbb{R}^{s_2}$ are defined. The forget gate, as a first step, defines which information must be eliminated from the previous chain produced at time $t-1$. This layer gives as output:

$$\boldsymbol{f}_t = sig(\boldsymbol{W}^{(f_t)}[\boldsymbol{x}_t, \boldsymbol{h}_{t-1}] + \boldsymbol{b}^{(f_t)}) \tag{16}$$

The next step is to select the new information to be provided to the chain. This procedure is divided into two parts: first, the input gate layer ($\boldsymbol{i}_t$) is activated via a sigmoid function to define which inputs will be considered for updating the Cell state. Then, another layer, defined as "new candidate gate layer" ($\boldsymbol{nc}_t$), produces a new set of possible candidates through a hyperbolic tangent function. From a mathematical point of view:

$$\boldsymbol{i}_t = sig(\boldsymbol{W}^{(i_t)}[\boldsymbol{x}_t, \boldsymbol{h}_{t-1}] + \boldsymbol{b}^{(i_t)}) \tag{17}$$

$$\boldsymbol{nc}_t = tanh(\boldsymbol{W}^{(nc_t)}[\boldsymbol{x}_t, \boldsymbol{h}_{t-1}] + \boldsymbol{b}^{(nc_t)}) \tag{18}$$

Then, the results of the previous steps are combined to define the Cell state at the current time:

$$\boldsymbol{C}_t = \boldsymbol{f}_t \cdot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \cdot \boldsymbol{nc}_t \tag{19}$$

The output gate layer ($\boldsymbol{o}_t$) defines which part of the Cell state will represent the output $\boldsymbol{h}_t$, activating the received signal through a sigmoid function:

$$\boldsymbol{o}_t = sig(\boldsymbol{W}^{(o_t)}[\boldsymbol{x}_t, \boldsymbol{h}_{t-1}] + \boldsymbol{b}^{(o_t)}) \tag{20}$$

Then, this output is multiplied by the Cell previously activated through the hyperbolic tangent function in such a way as to distribute the values between -1 and 1. In summary, the final output of the unit cell will be:

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{C}_t) \tag{21}$$

The LSTM layer process is iterated for each time step of the considered time series. Then, the procedure follows considering the hidden state obtained at the last time step. To prevent overfitting, the vector is processed by the Dropout Layer, which randomly sets input elements to zero with a given probability (0.5 in the considered application). The output produced by the Dropout Layer is managed by the Fully Connected Layer, which creates the synaptic connections necessary to connect the dropped out hidden state with the Regression Output Layer. The generated weight matrix will have dimension $\mathbb{R}^{s_3 \times s_2}$, which multiplied by the hidden state vector, produces the final output vector of dimension $\mathbb{R}^{s_3}$. Finally, the Regression Output Layer computes the half-mean-squared-error loss between the reproduced output and the target one. The objective function obtained is minimized using a backpropagation algorithm known as "Adam", widely used in the literature (Kingma and Ba, 2015).

In this work, the number of epochs was set equal to 100, the mini-batch size equal to 18, the initial learning rate equal to 0.005, the learning drop period equal to 40 and the learning drop factor equal to 0.1. The dimension of the sequence matrix is 190×1 of data type cell, where each row corresponds to an input block composed by a 36×9 matrix, where rows represent the number of features (precipitation and temperature provided by the 18 climate stations) and columns indicate the input sequence length. The number of hidden units was set equal to 100. At each time step, an output vector of dimension 10, corresponding to the number of wells, is obtained.

2.2.3. Convolutional neural network

A Convolutional Neural Network (CNN) is a deep learning model designed to automatically learn from raw input data, avoiding the procedure of manually extracting features typically required by traditional machine learning models. It is based on a grid architecture, typical of images, where the spatial hierarchies are recognized through specific convolutional layers, enabling the automatic identification of the most significant patterns within the initial grid (input image). Given their nature, CNNs were originally developed to work and solve image segmentation and pattern recognition problems, now they are widely used to classify non-image data such as sound, signal and time series data. Convolutional networks are characterized by three main layers: convolutional, pooling and fully connected. Convolutional and pooling layers, collaborating with specific activation functions, take care of the extraction of the features, while the fully connected layer processes the data extracted from the previous layers according to a classic forward neural network scheme, producing the desired output. In addition to these layers, other layers contribute to the structure of the network. In the present case (Fig. 4), the first layer is the Image Input Layer, which enters the image input data into the network. Then, it is sent to a block composed of the Convolutional Layer, Batch Normalization Layer, ReLu Layer and Average Pooling 2D Layer. Three of these blocks are placed in

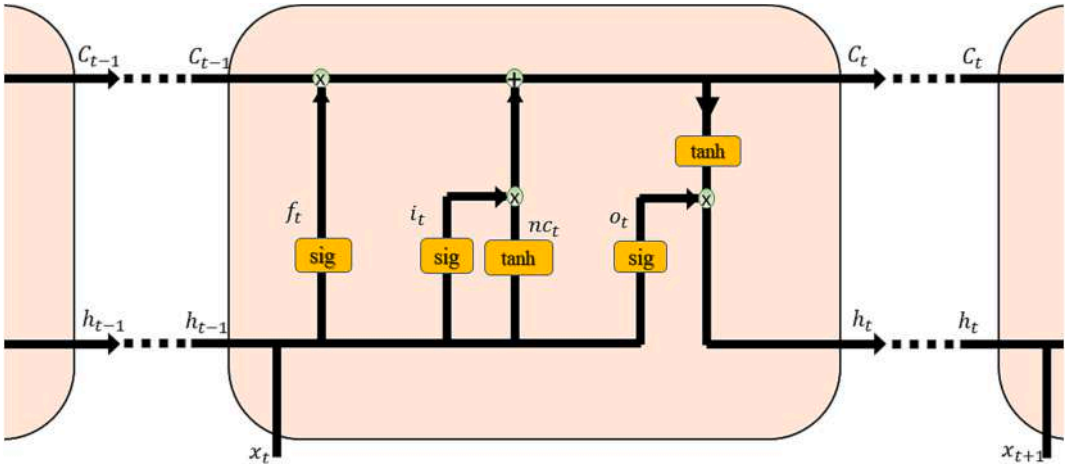


Fig. 3. LSTM cell architecture.

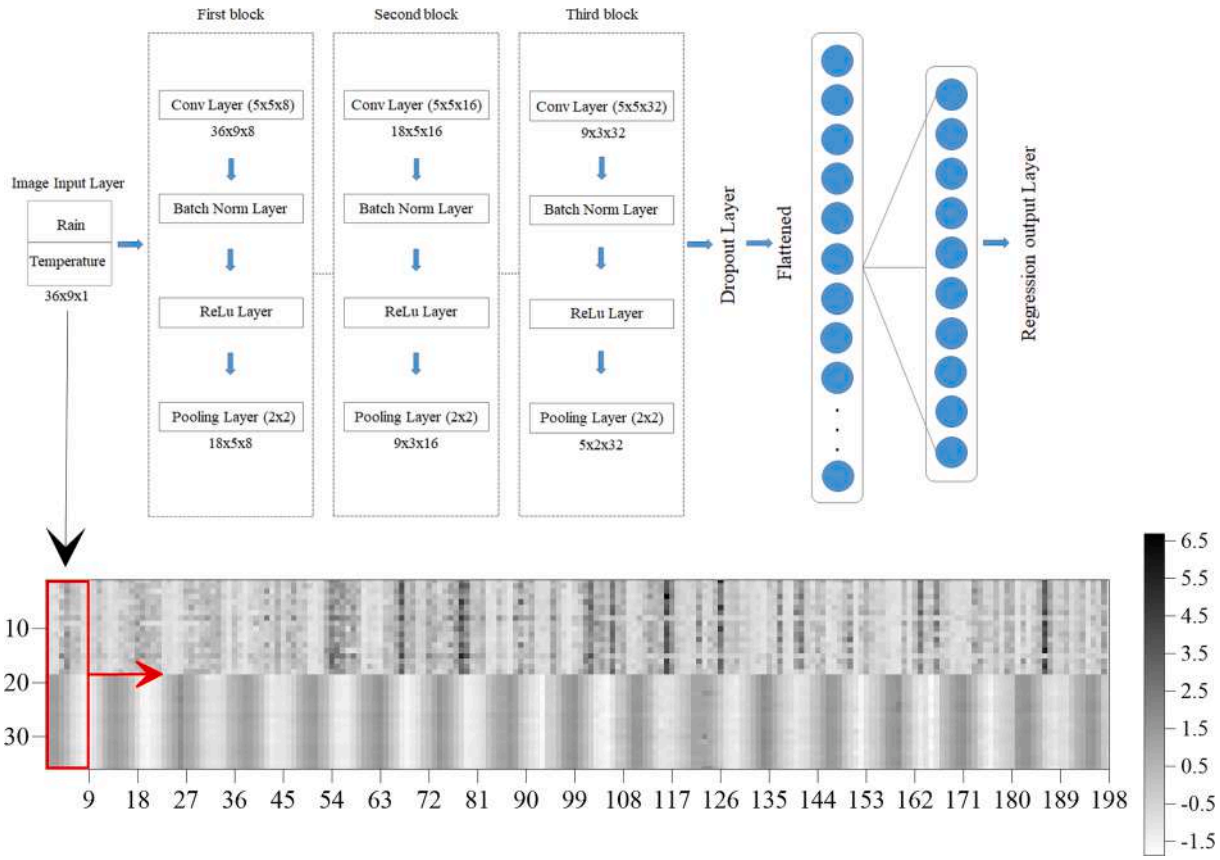


Fig. 4. Schematic view of the CNN (top) and sample of the standardized input data (bottom).

series in the implemented model. The Convolutional Layer represents the main part of CNNs. It consists of sets of filters, known in literature as Kernel or feature detectors, which represent the parameters that the network must learn during the training process. Typically, multiple filters of smaller size than the input image are used. Each filter is scrolled along the input image for both directions (W-E and N-S) and the scalar product between the filter and the input is computed for each spatial position in order to produce different feature maps (or activation maps). Let's consider an input image with dimensions $h \times w \times c$, where h and w are the height and width of the image, respectively, while c is the number of channels. This image will be processed by a Convolutional Layer composed by a number of filters (nf) of dimensions equal to $n \times m \times nf$ where n and m are the height and width of the filter, respectively. The number of filters determines the number of channels in the output. Once the number and size of the filters is defined, two other parameters must be set up, known as stride (s) and padding (p). The stride defines the step size at which the convolutional filter moves across the input image vertically and horizontally, while the padding defines the number of additional border pixels around the input image. Usually, the stride is set equal to 1, while the padding is defined in such a way that the output reproduced by the convolutional layer does not modify the initial size of the input. The size of the output reproduced by the convolutional layer is:

$$[row; column; channel] = \left[\frac{(h-n+2p)}{s}+1; \frac{(w-m+2p)}{s}+1; nf\right] \quad (22)$$

The Batch Normalization Layer normalizes a mini-batch of the extracted features, for each channel, independently. The goal of this procedure is to accelerate the training of the convolutional neural network and reduce the sensitivity to initial values of the network. The ReLu Layer activates the signal with the Rectified Linear Unit Function. Convolutional Layers are able to outline the presence of features in an input image; however, the output reproduced by these layers is very sensitive to the position of features in the input. By changing the feature location in the input image, the learned filters will reproduce a different output, making the model inefficient. The Pooling Layer deals with this problem by down sampling the spatial size of the feature maps, making the model translation-invariant and reducing the number of parameters, with a consequent reduction in terms of computational burden. The normalized and activated feature maps in the Average Pooling 2D Layer are divided into no-overlapping zones with a size equal to a defined filter (pooling filter). As for the LSTM, the Dropout Layer randomly sets the down sampled information to zero with a given probability (0.5). The dropped output is flattened into a single vector and then connected by means of the Fully Connected Layer to the Output Layer. The Regression Layer computes the half-mean-squared-error loss between the reproduced output and the target one. As for the LSTM approach, the objective function was minimized using the Adam algorithm.

In this work, the number of epochs was set equal to 100, the mini-batch size equal to 10, the initial learning rate equal to 0.001, the learning drop period equal to 50 and the learning drop factor equal to 0.1. The total number of input images is equal to 190, defined by the groundwater level time series length. Each input image has dimension $36 \times 9 \times 1$, where 36 are the representative features (monthly precipitation and temperature data provided by the 18 climate stations), 9 is the time window considered to select the input data and 1 is the number of channels of the image. For instance, the first image, which is used to obtain the groundwater levels in March 2005, contains the precipitation and temperature data from June 2004 to February 2005; the second image, which is the input related to the groundwater levels in April 2005, contains the climate variables from July 2004 to March 2005. The filter height and width are both equal to 5 for all three Convolutional Layers, while the number of filters is equal to 8, 16 and 32 for the first, second and third block, respectively. The pooling filter, with a height

and width of 2 and a stride of 2, is used to extract the average value. At each time step, an output vector of dimension 10, corresponding to the number of wells, is obtained. A schematic representation of the CNN for a single image is depicted in Fig. 4. In addition, the figure shows the raw map, made up of the sample of the standardized input data, used to generate all the input images. It is evident that the precipitation values, ranging from 1 to 18 on the Y-axis, exhibit significant variability, reflecting the natural fluctuations in patterns. Conversely, the temperature values, which are plotted from 19 to 36, show a pronounced seasonal pattern. The red rectangle denotes the image dimension; a sequential one-step scrolling is performed to generate a total of 190 images.

Again, once trained, the neural network was used to estimate groundwater levels as a function of precipitation and temperature data provided by the 13 climate models.

3. Results

3.1. Performance evaluation

The Mean Squared Error (MSE) was used to evaluate the performance of the developed machine learning models during the training, validation, and test phases:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N_d}(\widehat{W}_i - W_i)^2}{N_d} \tag{23}$$

where W_i is the actual value and $\widehat{W}_i$ is the corresponding value estimated by the neural networks and N_d is the number of observations in either the training, validation, testing or the whole dataset. In addition, for each well, the Root Mean-Square Error (RMSE) between predicted and observed groundwater levels in the test period was evaluated.

For each neural network implemented in this work, hyperparameters and activation functions were manually adjusted in order to identify the architecture with the minimum value of the Loss function and, at the same time, to limit the computational cost during the training procedure.

Table 4 reports the MSEs computed on the whole dataset and for the training, validation and testing subsets for all the AI methods proposed. For the NARX, the overall performance of the network is good, the MSE computed on the whole dataset is equal to 0.17 m^2. The learning phase highlights a training MSE equal to 0.04 m^2 and a validation MSE equal to 0.54 m^2, while the test phase is characterized by a MSE equal to 0.82 m^2. Therefore, the validation and, in particular, the test performances are poor, while the training performance is very good. This suggests that the neural network is probably affected by overfitting problems; it can map input to output during the training phase, but it is not accurate in estimating groundwater levels outside the training range.

Regarding the LSTM, the overall performance, expressed in terms of MSE, is 0.14 m^2. The training value of the performance index is 0.12 m^2, a MSE of 0.23 m^2 characterizes the validation phase and the test phase provides a MSE of 0.30 m^2. The metrics highlight good performance for the LSTM in all phases. Unlike the NARX, the Dropout Layer allows the LSTM to learn slightly less during the training phase, so as not to lose the ability to generalize. In fact, although the performance value in the training phase is higher than the NARX, the test value is lower denoting a good behavior of the LSTM network to forecast groundwater levels.

Table 4
MSE (m^2) between the output of the neural networks (NARX, LSTM, CNN) and the observed groundwater levels.

	NARX	LSTM	CNN
Training	0.04	0.12	0.05
Validation	0.54	0.23	0.20
Test	0.82	0.30	0.31
Whole dataset	0.17	0.14	0.10

Table 5
RMSE (m) between the output of the neural networks (NARX, LSTM, CNN) and the observed groundwater levels in the periods 2005–2018 (learning) and 2019–2020 (testing).

	NARX		LSTM		CNN	
Well	Learning	Testing	Learning	Testing	Learning	Testing
Paganico	0.10	0.37	0.11	0. 17	0.15	0. 30
Cugnia	0.07	0.35	0.08	0.10	0.16	0.25
Sat 1	0.11	0.36	0.09	0.11	0.14	0.26
Via Barsanti	0.10	0.27	0.11	0.14	0.10	0.29
Via Romboni	0.30	1.46	0.42	0.83	0.48	0.98
Percorso Vita	0.05	0.17	0.07	0.12	0.06	0.23
Nozzano	0.18	0.93	0.23	0.36	0.23	0.66
S. Alessio	0.06	0.59	0.11	0.18	0.17	0.27
Salicchi	0.07	0.43	0.10	0.17	0.11	0.38
Diecimo	0.13	0.57	0.21	0.28	0.16	0.39

The overall performance of the CNN is 0.1 m^2, while the MSE for the training subset is equal to 0.05 m^2. The validation procedure provides a MSE of 0.2 m^2 and the test phase is characterized by a performance index equal to 0.31 m^2. Overall, the CNN performance is comparable to that of the LSTM.

Table 5 shows the comparison between observed and predicted groundwater levels, in terms of RMSE values, for the 10 wells investigated and related to the learning and testing phases, obtained with the three AI models implemented. Overall, the models exhibit satisfactory performance metrics (RMSE less than 0.4 m) for both the learning and testing phases, except for the well Via Romboni, which presents the higher errors (greater than 0.8 m) across the three networks during the testing procedure. Always referring to the test phase, the well Nozzano shows a good performance only for the LSTM network, while the wells S. Alessio and Diecimo present unsatisfactory metrics for the NARX. In general, the LSTM seems more suitable and capable of predicting groundwater levels. As an example, Fig. 5 compares the predicted and observed groundwater levels in the test period 2019–2020 for the well Paganico. It can be noticed that the LSTM neural network better describes the actual groundwater levels in the test period.

3.2. Future projections

The trained networks were used in combination with the climate model projections to compute groundwater levels under two different scenarios (RCP4.5 and RCP8.5), from 1976 until the end of this century.

To highlight the evolving trends in piezometric levels over time, the results for the Paganico well in April are reported. April was selected because it typically experiences minimum anthropogenic impacts as it precedes the irrigation withdrawals period. The predicted groundwater levels for the entire simulated period (1976–2095) are shown in terms of 10-year moving average, which highlights the effects of climate change over the natural variability. The results are reported in terms of median value and interquartile range of the ensemble of climate models.

Fig. 6 shows the results obtained with the NARX. For both RCP scenarios, according to the median values, a decrease in groundwater levels over time is evident. The reduction is expected to be more pronounced for the RCP8.5 in the long-term; this can be attributed to the highest greenhouse gas concentrations associated with this scenario, which lead to a more severe increase in temperature and subsequent high evapotranspiration rates that impact recharge processes. In contrast, for the RCP4.5 scenario, mitigation measures are anticipated to be implemented to reduce the emission of greenhouse gases. The interquartile range of the ensemble of climate models highlights the uncertainties of future predictions given by the different RCMs.

To highlight changes over time, the results were also analyzed considering four periods of the climate models: historical (1976–2005) and short- (2006–2035), medium- (2036–2065) and long-term

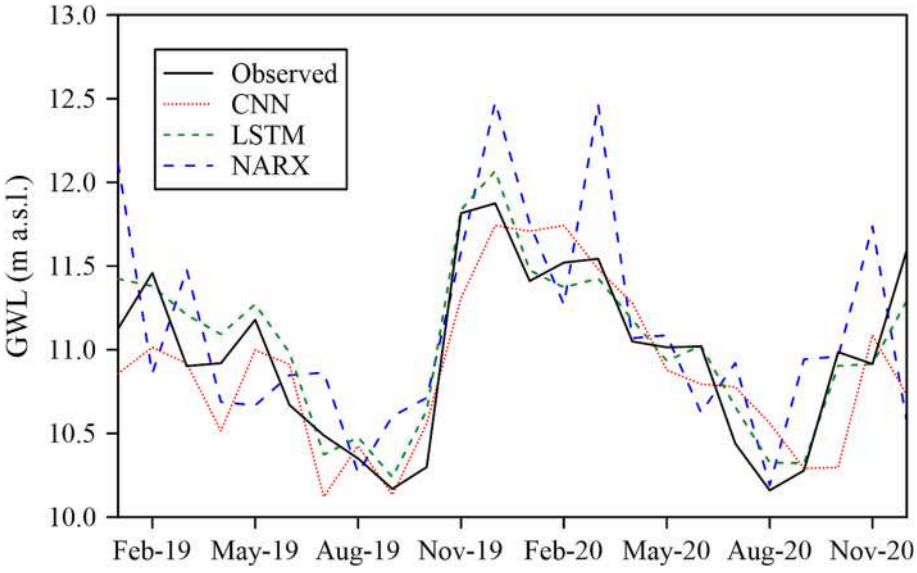


Fig. 5. Observed and predicted groundwater levels for the testing phase (period 2019–2020) for the well Paganico.

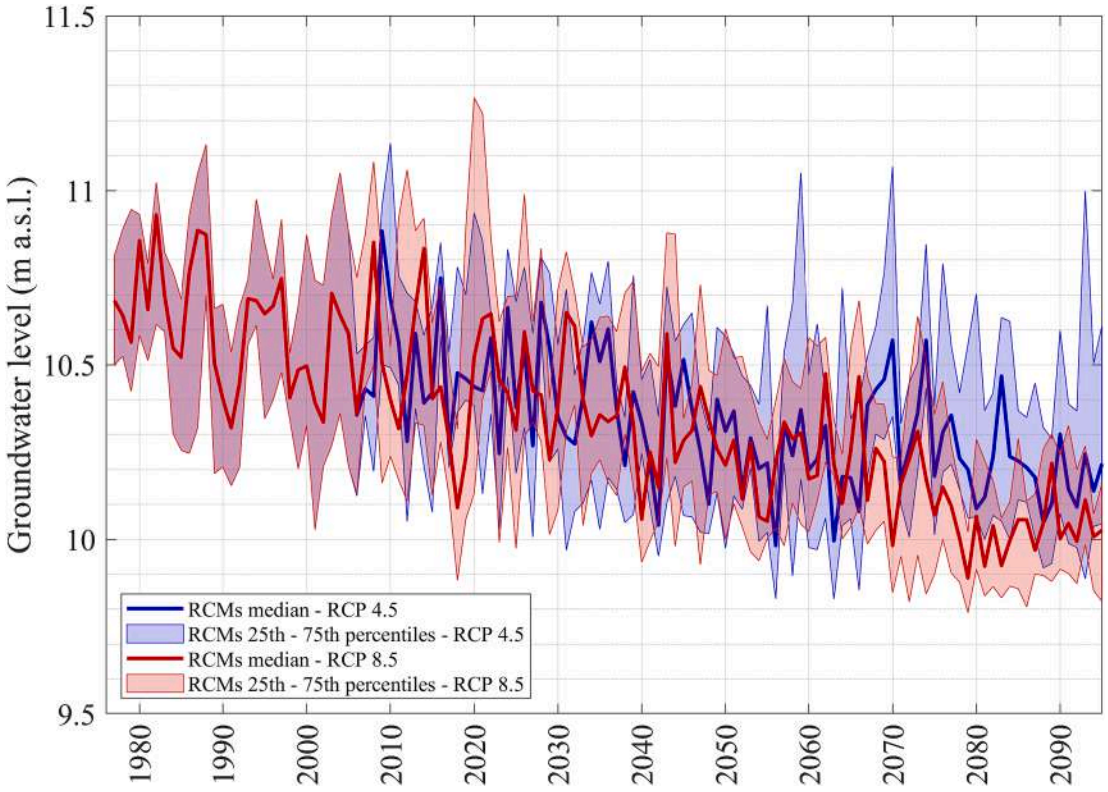


Fig. 6. Predicted groundwater levels with the NARX in April for the Paganico well in terms of 10-year moving average under the RCP4.5 and RCP8.5 scenarios.

(2066–2095).

Fig. 7 shows the empirical cumulative distribution functions (CDFs) of the standardized groundwater levels predicted over the four projection periods for the Paganico well and for the RCP4.5 and RCP8.5 scenarios. The envelope curves were obtained considering the individual climate models, in order to highlight the uncertainty of the prediction. In addition, the data of the 13 climate models were considered as statistical realizations of the same stochastic process, hence they were assembled in order to create a single dataset, hereinafter referred to as "whole RCM ensemble". For both scenarios, a clear reduction in the standardized groundwater levels is detectable for the future compared to the historical period, especially in the medium- and long-term. In fact, the leftward shift of the CDFs from the historical one highlights the increase in the probability or frequency of the lower values of the distribution. The RCP4.5 shows similar results in the medium- and long-term, while the RCP8.5 indicates a more pronounced increase in negative standardized groundwater levels over the long-term.

Fig. 8 and Fig. 9 show the results obtained with the LSTM. Looking at the 10-year moving average of the predicted groundwater levels in April (Fig. 8), for both RCP scenarios, the median values highlight a moderate decrease in the piezometric levels over time, more accentuated for the RCP8.5 at the end of the simulated period. The variability between the

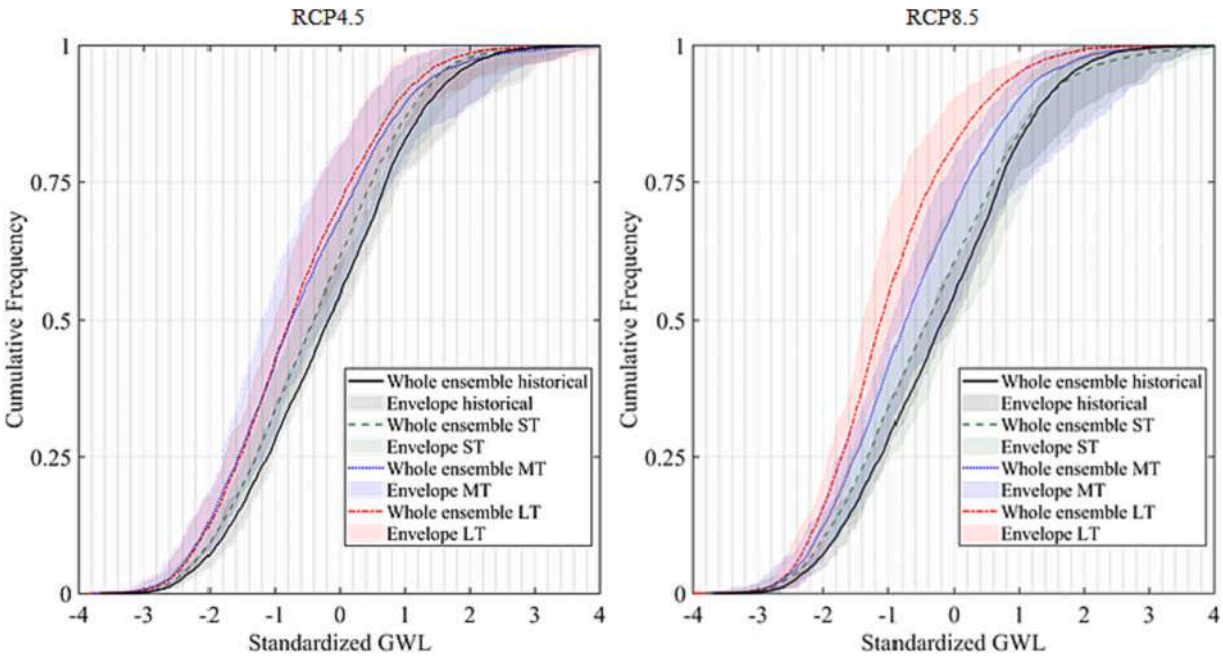


Fig. 7. Cumulative distribution probability functions of the standardized GWLs according to the whole RCM ensemble obtained with the NARX for the Paganico well for the historical period and at short- (ST), medium- (MT) and long-term (LT) under the RCP4.5 (left) and RCP8.5 (right) scenarios together with the envelope curves provided by the 13 RCMs.

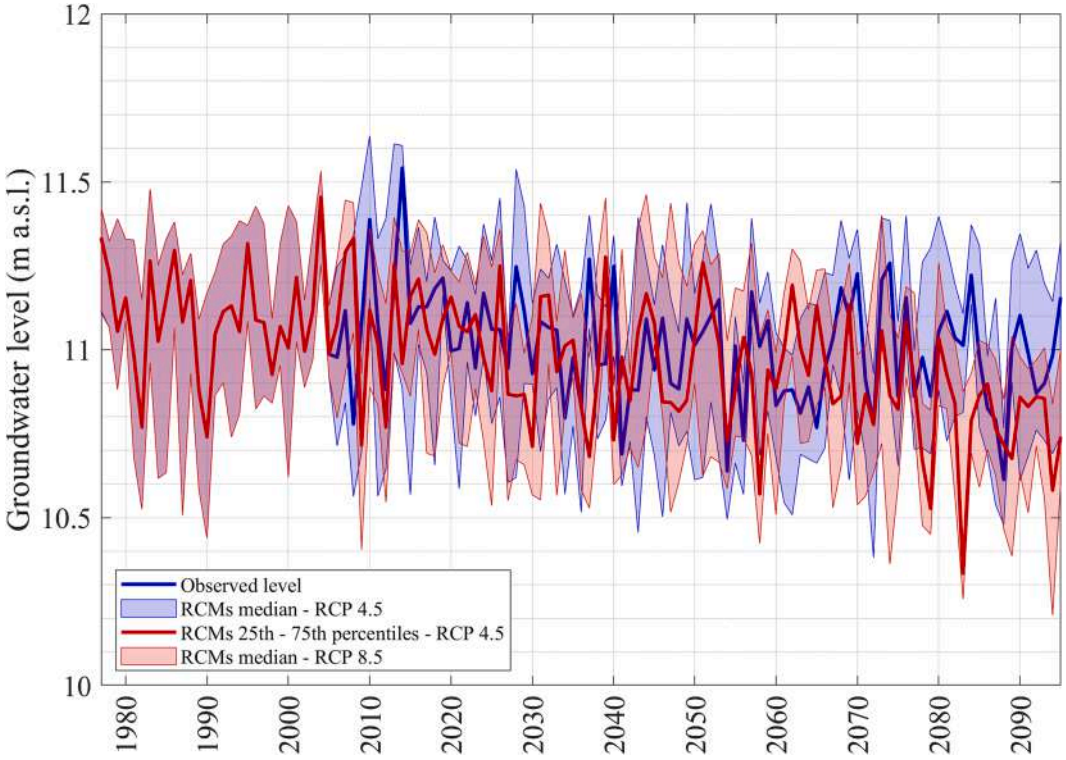


Fig. 8. Predicted groundwater levels with the LSTM in April for the Paganico well in terms of 10-year moving average under the RCP4.5 and RCP8.5 scenarios.

climate models (interquartile range) is similar to that obtained with the NARX.

The CDFs of the predicted standardized GWLs for the Paganico well (Fig. 9) indicate that an increment in negative standardized GWLs is expected for the future compared to the historical period. For the RCP4.5 similar results are obtained in the medium- and long-term; very few differences are expected in the short-term compared to the historical period. The RCP8.5 shows a progressive reduction of standardized GWLs over time, higher in the long-term than that expected with the RCP4.5. Furthermore, the envelope curves show a lower uncertainty on the estimated future GWLs using the LSTM with respect to the NARX. Overall, the LSTM estimates a future reduction in terms of standardized groundwater levels lower than the NARX.

Fig. 10 and Fig. 11 show the results obtained with the CNN. From

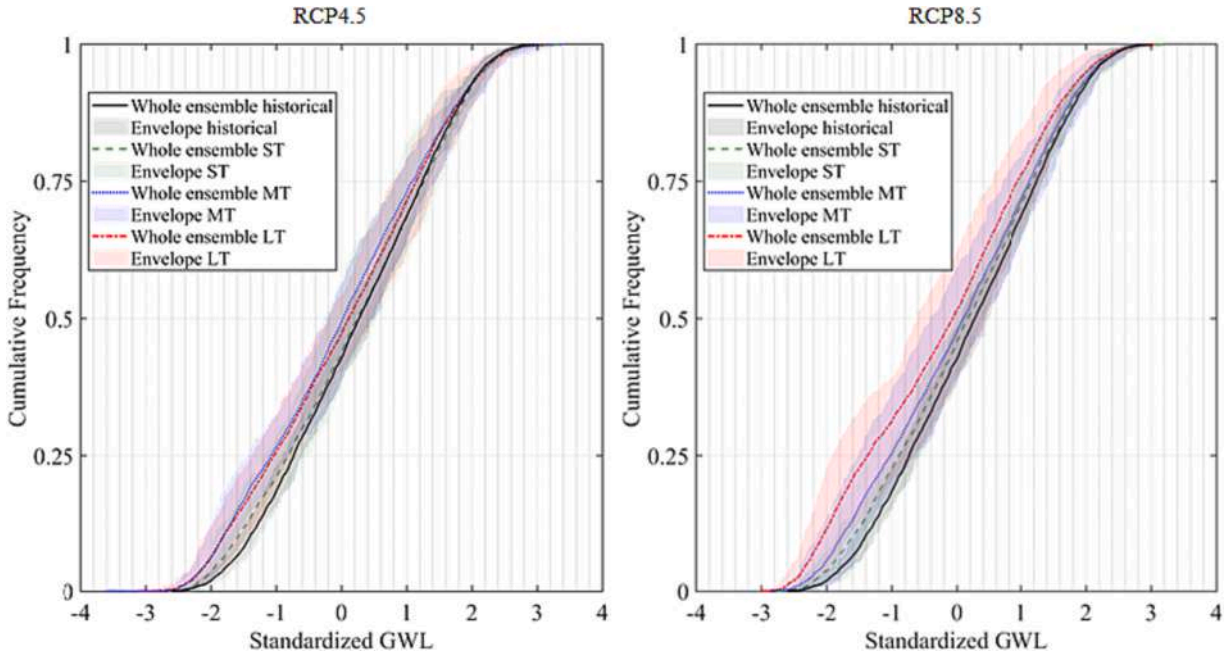


Fig. 9. Cumulative distribution probability functions of the standardized GWLs according to the whole RCM ensemble obtained with the LSTM for the Paganico well for the historical period and at short- (ST), medium- (MT) and long-term (LT) under the RCP4.5 (left) and RCP8.5 (right) scenarios together with the envelope curves provided by the 13 RCMs.

Fig. 10, no systematic trends can be detected for both RCP scenarios. The variability between RCMs is lower than that obtained with NARX and LSTM.

According to Fig. 11, the CDFs of the future GWLs obtained for the Paganico well do not show appreciable changes compared to that of the historical period, for both RCP scenarios. Only a slight decrease in standardized GWLs at long-term for the RCP8.5 scenario is detectable. This highlights a very different behavior of the CNN compared to the other two AI models.

Finally, to quantify the results for all wells and to have a comparison with the finding of Secci et al. (2021), Table 6 and Table 7 show the differences in terms of standardized GWLs between the future periods at short-, medium- and long-term for the RCP4.5 and RCP8.5 and the historical period. The 25th, 50th and 75th percentiles of the whole RCM

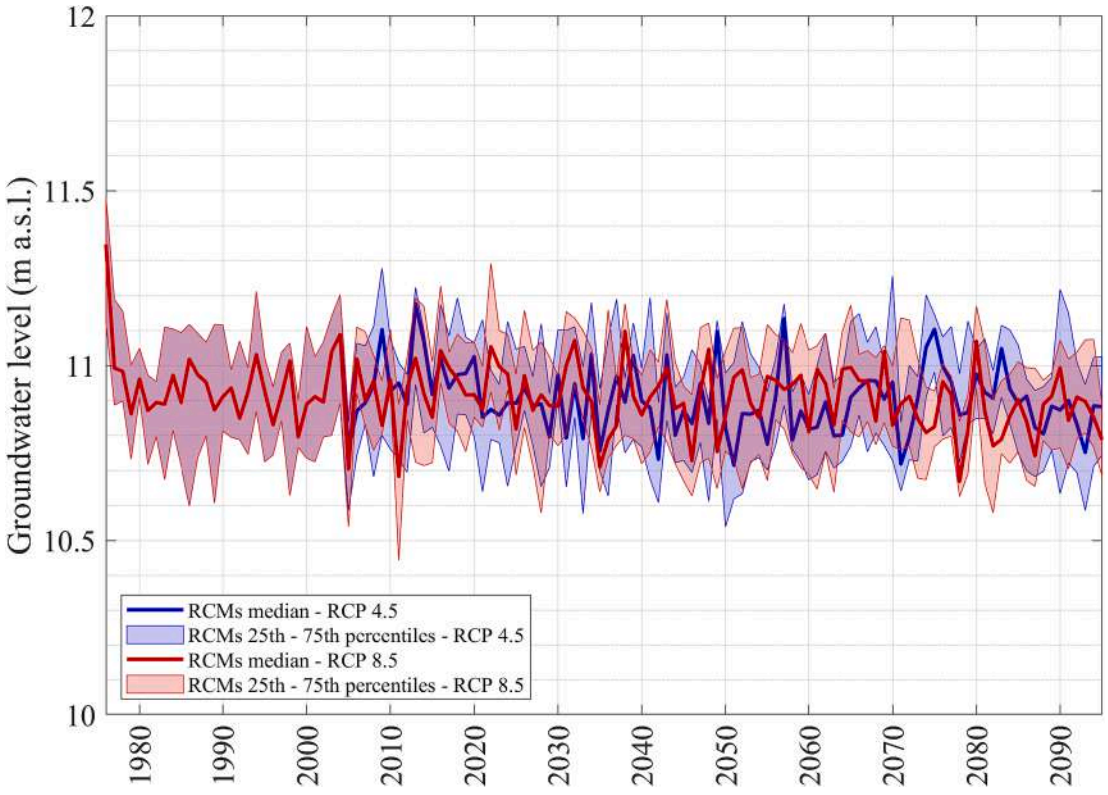


Fig. 10. Predicted groundwater levels with the CNN in April for the Paganico well in terms of 10-year moving average under the RCP4.5 and RCP8.5 scenarios.

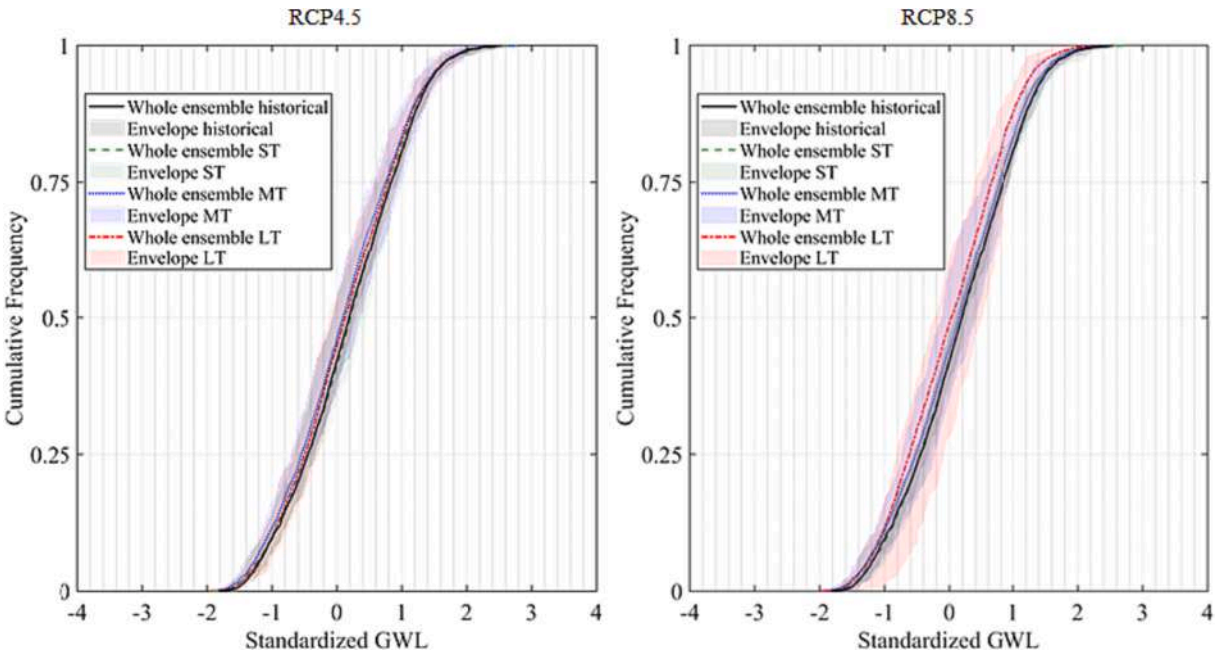


Fig. 11. Cumulative distribution probability functions of the standardized GWLs according to the whole RCM ensemble obtained with the CNN for the Paganico well for the historical period and at short- (ST), medium- (MT) and long-term (LT) under the RCP4.5 (left) and RCP8.5 (right) scenarios together with the envelope curves provided by the 13 RCMs.

ensemble were selected to compute the differences. The color intensity highlights the magnitude of the differences between future and historical standardized GWLs. All models detect a reduction in groundwater levels for the future, however, on the whole, the CNN is characterized by the smaller differences if compared with the other two AI models implemented in the present study. For example, for the Paganico well and the RCP8.5 the difference between the standardized GWL median value at long-term and that of the historical period is −0.93 for the NARX, −0.34 for the LSTM and only −0,07 for the CNN (Table 7). According to the CNN, the largest differences are observed for the wells Via Romboni and Nozzano; however, they present high values of RMSE between predicted and observed groundwater levels in the test period (Table 5). In general, the regression model implemented by Secci et al. (2021) highlights slightly higher reductions of the groundwater levels with respect to those predicted by the CNN. On the contrary, the LSTM, with very few exceptions, shows more pronounced decreases in future levels with respect to the regression model. The NARX provides the most severe declines in future groundwater levels; however, it presents the highest MSEs in the test phase (Table 4).

4. Discussion

The approach proposed in this study aims to evaluate the effects of future climate on groundwater while assuming that other driving forces of the hydrological processes will remain unchanged. This assumption may introduce uncertainties in the predictions, as other factors, such as precipitation intensity, soil moisture and vegetation type, play a role in aquifer recharge. Furthermore, also anthropogenic activities can influence the groundwater condition. However, some of these factors are inherently taken into account during the training phase of the models when establishing the relationship between the exogenous input (precipitation and temperature) and the desired output (groundwater level). For instance, the models can indirectly capture the influence of changes in pumping rates on the groundwater levels, as these rates are closely correlated with changes in precipitation and temperature. An increase in temperature often results in an increased demand for irrigation to compensate for higher evapotranspiration rates; this leads to higher groundwater extraction rates as agricultural sectors may rely on groundwater resources. A decrease in precipitation and prolonged period of droughts can lead to a reduced recharge and at the same time to an increase in groundwater extraction rates as surface water supplies become limited. Hence, despite the models used do not explicitly include factors beyond precipitation and temperature, some of these factors can be indirectly considered as they depend on climate variables used as input. Moreover, precipitation and temperature are among the most straightforward variables to measure and collect and they are readily accessible from climate models. This justifies their use as input features for the machine learning models to forecast future groundwater levels.

Another aspect worthy of discussion is the comparison between a straightforward statistical approach (Secci et al., 2021) and artificial intelligence methods to infer future projections of groundwater levels in the context of climate change. One of the advantages of the proposed methods compared to that of Secci et al. (2021) is that the climate variables (precipitation and temperature) are used directly instead of computing meteorological indices. In particular, to evaluate the SPEI both precipitation and evapotranspiration are needed and, within the scientific literature, different equations are proposed to evaluate evapotranspiration based on temperature and other variables. This leads to different values of SPEI and can represent a further source of uncertainty, which is avoided by directly using temperature as input variable to the models. Furthermore, in this work a single network is able to simultaneously map all climate data with groundwater levels at all wells, while the method proposed by Secci et al. (2021) requires the development of a regression model for each well. Another main advantage of the AI models is their ability to capture non-linear relationships. The process of water infiltration and aquifer recharge is non-linear; therefore, machine-learning algorithms better represent the mutual dependences among groundwater levels and climate variables than simple linear functions.

Another aspect to discuss is the influence of the characteristics of the training dataset on the model performance and the design of the network architecture. The amount of data used to train a neural network is a crucial factor that can significantly impact the model performance. Generally, a larger dataset provides the model with more information and can lead to better performance and generalization. However, the variability of the data within the dataset is also important: a diverse

Table 6
Differences between the 25th, 50th and 75th percentiles of the future standardized GWLs at short- (ST), medium- (MT) and long-term (LT) and the historical ones under the RCP4.5 scenario. Results obtained with the AI models proposed in this study and the regression model presented by Secci et al. (2021).

		RCP4.5 - Differences with the historical period											
		REGRESSION			NARX			LSTM			CNN		
		ST	MT	LT	ST	MT	LT	ST	MT	LT	ST	MT	LT
Paganico	25th	-0.04	-0.28	-0.23	-0.19	-0.46	-0.43	-0.08	-0.35	-0.30	0.05	-0.01	0.05
	50th	-0.08	-0.38	-0.32	-0.22	-0.58	-0.59	-0.03	-0.24	-0.16	0.03	-0.08	-0.02
	75th	-0.08	-0.36	-0.29	-0.22	-0.43	-0.55	0.02	-0.16	-0.10	0.01	-0.06	-0.04
Cugnia	25th	0.01	-0.13	-0.11	0.14	0.01	0.03	-0.08	-0.32	-0.21	0.01	-0.10	-0.04
	50th	-0.03	-0.20	-0.17	0.15	0.05	0.06	-0.04	-0.19	-0.12	0.00	-0.11	-0.07
	75th	-0.07	-0.22	-0.19	0.25	0.19	0.22	0.02	-0.14	-0.05	-0.03	-0.11	-0.08
SAT 1	25th	0.01	-0.13	-0.11	-0.27	-0.50	-0.46	-0.23	-0.57	-0.61	0.04	-0.08	-0.02
	50th	-0.03	-0.20	-0.17	-0.18	-0.40	-0.46	-0.11	-0.34	-0.33	0.02	-0.08	-0.04
	75th	-0.07	-0.22	-0.19	-0.12	-0.29	-0.37	-0.05	-0.24	-0.21	-0.04	-0.10	-0.08
Via Barsanti	25th	0.03	-0.16	-0.11	-0.18	-0.45	-0.41	-0.12	-0.38	-0.33	0.00	-0.10	-0.05
	50th	-0.02	-0.21	-0.16	-0.09	-0.27	-0.27	-0.05	-0.27	-0.19	0.02	-0.10	-0.06
	75th	-0.05	-0.19	-0.16	-0.02	-0.12	-0.12	0.01	-0.18	-0.12	0.00	-0.09	-0.05
Via Romboni	25th	0.02	-0.20	-0.16	-0.17	-0.57	-0.61	-0.11	-0.38	-0.30	-0.04	-0.20	-0.17
	50th	-0.04	-0.25	-0.20	-0.09	-0.51	-0.51	-0.04	-0.21	-0.17	-0.02	-0.18	-0.14
	75th	-0.06	-0.20	-0.17	0.04	-0.25	-0.25	-0.02	-0.12	-0.10	-0.04	-0.10	-0.07
Percorso Vita	25th	0.01	-0.13	-0.11	-0.19	-0.51	-0.45	-0.10	-0.39	-0.35	0.03	-0.04	0.02
	50th	-0.02	-0.17	-0.14	-0.13	-0.36	-0.35	-0.06	-0.26	-0.19	0.02	-0.09	-0.03
	75th	-0.07	-0.20	-0.19	-0.14	-0.25	-0.32	0.02	-0.19	-0.12	-0.04	-0.11	-0.07
Nozzano	25th	0.01	-0.18	-0.14	-0.45	-0.90	-0.94	-0.01	-0.12	-0.02	-0.06	-0.12	-0.13
	50th	-0.04	-0.25	-0.20	-0.41	-0.82	-0.91	0.01	-0.16	-0.01	-0.05	-0.18	-0.16
	75th	-0.07	-0.24	-0.20	-0.30	-0.66	-0.75	0.10	-0.02	0.06	-0.05	-0.12	-0.12
S. Alessio	25th	-0.01	-0.16	-0.13	-0.39	-0.74	-0.67	-0.01	-0.15	-0.05	0.01	0.00	0.05
	50th	-0.04	-0.21	-0.18	-0.43	-0.74	-0.73	0.03	-0.14	0.01	0.02	-0.06	-0.01
	75th	-0.08	-0.23	-0.20	-0.49	-0.81	-0.78	0.06	0.01	0.07	0.03	0.02	0.05
Salicchi	25th	0.02	-0.16	-0.12	-0.85	-1.39	-1.45	0.01	-0.14	-0.04	-0.05	-0.11	-0.10
	50th	-0.04	-0.22	-0.18	-0.84	-1.50	-1.65	0.00	-0.18	-0.05	-0.03	-0.17	-0.14
	75th	-0.06	-0.20	-0.16	-0.82	-1.38	-1.59	0.08	-0.02	0.05	-0.03	-0.11	-0.10
Diecimo	25th	0.12	-0.10	-0.05	-0.29	-0.69	-0.77	-0.10	-0.33	-0.22	-0.05	-0.16	-0.13
	50th	0.07	-0.18	-0.12	-0.22	-0.64	-0.71	-0.02	-0.20	-0.12	-0.01	-0.14	-0.09
	75th	0.05	-0.13	-0.09	-0.02	-0.30	-0.36	0.00	-0.11	-0.07	0.00	-0.02	0.01

Table 7
Differences between the 25th, 50th and 75th percentiles of the future standardized GWLs at short- (ST), medium- (MT) and long-term (LT) and the historical ones under the RCP8.5 scenario. Results obtained with the AI models proposed in this study and the regression model presented by Secci et al. (2021).

		RCP8.5 - Differences with the historical period											
		REGRESSION			NARX			LSTM			CNN		
		ST	MT	LT	ST	MT	LT	ST	MT	LT	ST	MT	LT
Paganico	25th	-0.04	-0.22	-0.29	-0.20	-0.40	-0.60	-0.16	-0.28	-0.64	0.05	0.03	0.06
	50th	-0.11	-0.32	-0.35	-0.23	-0.57	-0.93	-0.12	-0.18	-0.34	0.03	-0.02	-0.07
	75th	-0.06	-0.27	-0.29	-0.11	-0.52	-1.04	-0.08	-0.11	-0.30	0.02	-0.05	-0.14
Cugnia	25th	-0.01	-0.11	-0.13	0.15	0.04	-0.35	-0.13	-0.21	-0.40	0.01	-0.05	-0.16
	50th	-0.06	-0.17	-0.18	0.20	0.08	-0.38	-0.07	-0.11	-0.24	0.00	-0.06	-0.14
	75th	-0.08	-0.20	-0.21	0.36	0.20	-0.30	-0.06	-0.07	-0.19	-0.02	-0.08	-0.14
SAT 1	25th	-0.01	-0.11	-0.13	-0.23	-0.46	-0.52	-0.31	-0.61	-1.11	0.04	-0.01	-0.11
	50th	-0.06	-0.16	-0.17	-0.15	-0.43	-0.64	-0.17	-0.33	-0.63	0.03	-0.03	-0.13
	75th	-0.08	-0.20	-0.21	-0.06	-0.33	-0.61	-0.13	-0.24	-0.47	-0.01	-0.09	-0.19
Via Barsanti	25th	0.00	-0.12	-0.15	-0.17	-0.37	-0.51	-0.17	-0.36	-0.68	0.00	-0.07	-0.17
	50th	-0.06	-0.14	-0.17	-0.07	-0.25	-0.35	-0.13	-0.23	-0.43	0.00	-0.05	-0.16
	75th	-0.05	-0.15	-0.17	0.02	-0.13	-0.23	-0.07	-0.14	-0.34	0.00	-0.06	-0.18
Via Romboni	25th	-0.01	-0.16	-0.21	-0.26	-0.54	-0.93	-0.15	-0.31	-0.51	-0.05	-0.20	-0.34
	50th	-0.07	-0.19	-0.22	-0.15	-0.50	-0.97	-0.08	-0.19	-0.37	-0.06	-0.14	-0.27
	75th	-0.06	-0.15	-0.18	-0.04	-0.32	-0.77	-0.08	-0.14	-0.22	-0.01	-0.10	-0.19
Percorso Vita	25th	0.00	-0.10	-0.14	-0.20	-0.44	-0.68	-0.16	-0.34	-0.70	0.05	0.01	-0.03
	50th	-0.05	-0.12	-0.16	-0.09	-0.31	-0.53	-0.14	-0.22	-0.39	0.00	-0.03	-0.08
	75th	-0.08	-0.19	-0.21	-0.03	-0.32	-0.61	-0.07	-0.13	-0.32	-0.02	-0.06	-0.16
Nozzano	25th	-0.01	-0.14	-0.20	-0.45	-0.92	-1.28	-0.04	-0.02	-0.10	-0.03	-0.16	-0.21
	50th	-0.07	-0.19	-0.24	-0.38	-0.90	-1.32	-0.02	-0.07	-0.17	-0.07	-0.17	-0.27
	75th	-0.06	-0.21	-0.24	-0.20	-0.75	-1.16	0.05	0.03	-0.06	-0.02	-0.12	-0.22
S. Alessio	25th	-0.02	-0.13	-0.18	-0.49	-0.61	-0.74	-0.04	-0.04	-0.12	0.06	0.02	0.05
	50th	-0.06	-0.16	-0.21	-0.50	-0.70	-0.70	0.00	-0.06	-0.14	0.02	-0.01	-0.05
	75th	-0.08	-0.20	-0.25	-0.52	-0.81	-0.81	0.04	0.02	-0.01	0.05	0.04	0.04
Salicchi	25th	-0.01	-0.12	-0.19	-0.87	-1.43	-1.70	-0.02	-0.03	-0.13	-0.03	-0.14	-0.14
	50th	-0.06	-0.17	-0.22	-0.87	-1.64	-2.15	-0.04	-0.10	-0.22	-0.05	-0.16	-0.23
	75th	-0.06	-0.17	-0.21	-0.73	-1.53	-2.34	0.04	0.02	-0.09	-0.02	-0.11	-0.19
Diecimo	25th	0.10	-0.06	-0.11	-0.33	-0.73	-1.16	-0.12	-0.25	-0.39	-0.05	-0.13	-0.24
	50th	0.04	-0.12	-0.16	-0.25	-0.69	-1.14	-0.06	-0.15	-0.31	-0.02	-0.08	-0.17
	75th	0.07	-0.10	-0.13	-0.11	-0.40	-0.74	-0.07	-0.10	-0.19	0.01	-0.04	-0.04

dataset that encompasses a wide range of variations in the input data can help the neural network learning to recognize different patterns and making more accurate predictions. On the other hand, if the dataset contains only a limited range of examples, the model may overfit to the training data and perform poorly on unseen data. In cases where datasets have limited amounts of data, the challenge is to train a model effectively with limited information. Thus, the objective is to find strategies to optimize the use of the available data. Techniques based on physics constrains, regularization and the use of auxiliary variables can help in enhancing the model performance when facing data scarcity (e.g. Bierkens et al., 2001; Varouchakis and Hristopulos, 2013). In the context of machine learning, there is a prevalent preference for simpler models over complex ones. This is attributed to several advantages associated with simpler models, including ease of interpretation, a lower number of parameters, and potentially improved generalization performance. Simple models are more likely to capture the essential patterns, are less sensitive to individual data points and generalize well to unseen data; this is particularly advantageous when dealing with small datasets. Additionally, simpler models are often computationally more efficient, requiring less time and computational resources for training and inference. However, it is not always the case that a simpler model is inherently more suitable when working with limited data. It is important to note that the choice between simple and complex models depends on the specific problem and available data. In some cases, complex models may be necessary to capture intricate patterns. The trade-off between model simplicity and performance should be carefully considered based on the requirements of the task at hand.

The results of this study pointed out that different neural networks provide different outputs. The selection of the best model for estimating future groundwater levels is a topic worthy of discussion. Although there is no strict way to define which AI model is best suited to solve the problem at hand, the comparison of different methods can help the interpretation of the results. The training and testing performance metrics of the neural networks allow to evaluate the reliability of the developed models. However, metrics alone are not sufficient to determine which model will perform best in predicting future values. As an example, in the present work, the CNN presents good metrics for the training and test phases, but they are not completely able to extrapolate future groundwater levels. It is well known that CNNs are able to efficiently extract features from the training dataset, thereby reducing the dimensionality of the initial information and relating it to the desired output. However, they may fail when extrapolating values beyond the range of variation observed in the training set. This may explain why the CNN model does not show remarkable changes in the predicted future groundwater levels, although the future temperature shows a significant upward trend in the investigated area (Secci et al., 2021). CNNs are typically better suited for tasks that involve spatial relationships as image processes; when applied for processing time series data, by treating them as a 1D signal and using 1D convolutions, they may not be as effective in capturing temporal dependencies and patterns. Additionally, CNNs require a fixed input size, which can be challenging when dealing with time series data with variable length or missing values. On the contrary, NARXs and LSTMs are designed to capture temporal dependencies and patterns in time series data. The LSTM network outperforms the other approaches, showing the best performance during both the training and testing phases. The main advantage of the LSTM is its ability to selectively retain, through its memory cell, and use relevant information from the input sequence, while filtering out the irrelevant or redundant one. Furthermore, unlike recurrent neural networks, LSTMs are not affected by the vanishing gradient problem. This empowers the model to capture complex temporal patterns and relationships within the data, which can present challenges for other type of networks. Finally, NARXs use a simple feedforward neural network architecture that incorporates a looping structure by using the output as subsequent input. This design allows the model to identify potential trends in the reference output but may not be as effective at capturing long-term dependencies in the data. They are not classified as deep learning methods and do not have the capability to automatically extract features. This is why antecedent groundwater levels serve as essential auxiliary information for the NARX.

To summarize, the LSTM is specifically tailored to model sequences and capture long-term temporal dependencies, which make them well suited for time series prediction. Based on these considerations, as described by Yang and Zhang (2022), an interesting approach could be a combined CNN-LSTM system for forecasting problems. CNNs could be used as feature extractors to reduce the dimensionality of the dataset and select only the most relevant characteristics, which represents a more sophisticated alternative to clustering algorithms. Then, LSTMs would act as a time-sequence detector on a more representative sample.

5. Conclusions

In this study, three different artificial intelligence models (NARX, LSTM and CNN) were developed in order to assess the effects of climate change on groundwater resources. The applicability of the models was demonstrated for an Italian area in northern Tuscany, but other groundwater systems can be analyzed following the same procedure, assuming that enough data are available. The models were trained using historical precipitation and temperature data of 18 climate stations as input in order to obtain groundwater levels for 10 monitoring wells as output. Once trained, the AI models were used to predict future groundwater levels using, as input, precipitation and temperature projections provided by an ensemble of 13 RCMs under two emission pathways (RCP4.5 and RCP8.5). The NARX and the LSTM show a remarkable decrease in future groundwater levels, in particular for the RCP8.5 at long-term, denoting that the impact of climate change on groundwater resources could be significant; on the contrary, the CNN does not show significant changes.

Data-driven surrogate models, such as neural networks, can be a valid alternative to complete numerical models with an advantage in terms of computational burden. Although there are no well-defined methods capable of evaluating which surrogate model may perform better in the future, non-linear artificial intelligence models seem more suitable for future prediction of groundwater levels than statistical approaches. LSTMs are the best structured networks to manage dependencies in long time series, also confirmed by the evaluation metrics. However, a comparison between various methodologies and a general analysis of the historical and future climate is essential in order to rigorously analyze the results.

New climate scenarios, the Shared Socioeconomic Pathways (SSPs), have been proposed to describe different plausible future pathways of human societies and their interactions with the natural environment, based on assumptions about socio-economic development, demographic trends, technological change, energy consumption, land use, and other factors that affect greenhouse gas emissions. The SSPs are designed to work in combination with an updated version of the RCPs and are used in phase 6 of the Coupled Model Intercomparison Project (CMIP6). In the sixth Assessment Report of the Intergovernmental Panel on Climate Change (AR6, IPCC, 2021), GCMs were used to simulate long-term climate projections covering the period 2015–2100 under different SSP scenarios (from SSP1-1.9 to SSP5-8.5). RCM experiments, useful for vulnerability, impact and adaptation studies at regional and local scale, are still under development. Future works will focus on the new SSP scenarios.

CRediT authorship contribution statement

Daniele Secci: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Maria Giovanna Tanda:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Marco D'Oria:** Conceptualization, Methodology, Writing – review & editing. **Valeria Todaro:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was developed under the scope of the InTheMED project. InTheMED is part of the PRIMA programme supported by the European Union's HORIZON 2020 research and innovation programme under grant agreement No 1923. The authors are grateful to GAIA S.p.A. for the help during the data collection phase. Valeria Todaro acknowledges financial support from PNRR MUR project ECS_00000033_ECOSISTER.

References

Afrifa, S., Zhang, T., Appiahene, P., Varadarajan, V., 2022. Mathematical and Machine Learning Models for Groundwater Level Changes: A Systematic Review and Bibliographic Analysis. Futur. Internet 14, 259. https://doi.org/10.3390/fi14090259.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration.

ARPAE EMILIA ROMAGNA – Agenzia Regionale Prevenzione Ambiente Energia Emilia Romagna, 2021. URL https://www.arpae.it (accessed 9.01.22).

Asher, M. J., B.F.W.C., A. J. Jakeman, and L.J.M.P., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51, 5957–5973. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x.

Azizi, H., Ebrahimi, H., Samani, H.M.V., Khaki, V., 2021. Evaluating the effects of climate change on groundwater level in the Varamin plain. Water Science and Technology: Water Supply 21, 1372–1384. https://doi.org/10.2166/ws.2021.007.

Bierkens, M.F.P., Knotters, M., Hoogland, T., 2001. Space-time modeling of water table depth using a regionalized time series model and the Kalman Filter. Water Resources Research 37, 1277–1290. https://doi.org/10.1029/2000WR900353.

Bloomfield, J.P., Marchant, B.P., 2013. Analysis of groundwater drought building on the standardised precipitation index approach. Hydrology and Earth System Sciences 17, 4769–4787. https://doi.org/10.5194/hess-17-4769-2013.

Brouyère, S., Carabin, G., Dassargues, A., 2004. Climate change impacts on groundwater resources: Modelled deficits in a chalky aquifer, Geer basin. Belgium. Hydrogeol. J. 12, 123–134. https://doi.org/10.1007/s10040-003-0293-1.

Chang, J., Wang, G., Mao, T., 2015. Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. Journal of Hydrology 529, 1211–1220. https://doi.org/10.1016/j.jhydrol.2015.09.038.

Chen, H., Wang, S., Gao, Z., Hu, Y., 2010. Artificial neural network approach for quantifying climate change and human activities impacts on shallow groundwater level - A case study of Wuqiao in north China plain. 2010 18th Int. Conf. Geoinformatics, Geoinformatics 2010. https://doi.org/10.1109/GEOINFORMATICS.2010.5567678.

Citrini, A., Camera, C., Beretta, G.P., 2020. Nossana Spring (Northern Italy) under Climate Change: Projections of future discharge rates and water availability. Water (switzerland) 12. https://doi.org/10.3390/w12020387.

Coppola, E., Szidarovszky, F., Poulton, M., Charles, E., 2003. Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System under Variable State, Pumping, and Climate Conditions. Journal of Hydrologic Engineering 8, 348–360. https://doi.org/10.1061/(asce)1084-0699(2003)8:6(348).

Croley, T.E., Luukkonen, C.L., 2003. Potential effects of climate change on ground water in Lansing. Michigan. J. Am. Water Resour. Assoc. 39, 149–163. https://doi.org/10.1111/j.1752-1688.2003.tb01568.x.

D'Oria, M., Ferraresi, M., Tanda, M.G., 2017. Historical trends and high-resolution future climate projections in northern Tuscany (Italy). Journal of Hydrology 555, 708–723. https://doi.org/10.1016/j.jhydrol.2017.10.054.

D'Oria, M., Cozzi, C., Tanda, M.G., 2018. Future precipitation and temperature changes over the Taro, Parma and Enza River basins in Northern Italy. Ital. J. Eng. Geol. Environ. 2018, 49–63. https://doi.org/10.4408/IJEGE.2018-01.S-05.

D'Oria, M., Ferraresi, M., Tanda, M.G., 2019. Quantifying the impacts of climate change on water resources in northern Tuscany, Italy, using high-resolution regional projections. Hydrological Processes 33, 978–993. https://doi.org/10.1002/hyp.13378.

Emamgholizadeh, S., Moslemi, K., Karami, G., 2014. Prediction the Groundwater Level of Bastam Plain (Iran) by Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS). Water Resources Management 28, 5433–5446. https://doi.org/10.1007/s11269-014-0810-0.

Fallah-Mehdipour, E., Bozorg Haddad, O., Mariño, M.A., 2013. Prediction and simulation of monthly groundwater levels by genetic programming. J. Hydro-Environment Res. 7, 253–260. https://doi.org/10.1016/j.jher.2013.03.005.

Gharehbaghi, A., Ghasemlounia, R., Ahmadi, F., Albaji, M., 2022. Groundwater level prediction with meteorologically sensitive Gated Recurrent Unit (GRU) neural networks. Journal of Hydrology 612, 128262. https://doi.org/10.1016/j.jhydrol.2022.128262.

Ghazi, B., Jeihouni, E., Kalantari, Z., 2021. Predicting groundwater level fluctuations under climate change scenarios for Tasuj plain. Iran. Arab. J. Geosci. 14 https://doi.org/10.1007/s12517-021-06508-6.

Ghose, D., Das, U., Roy, P., 2018. Modeling response of runoff and evapotranspiration for predicting water table depth in arid region using dynamic recurrent neural network. Groundwater for Sustainable Development 6, 263–269. https://doi.org/10.1016/j.gsd.2018.01.007.

Gonzalez, R.Q., Arsanjani, J.J., 2021. Prediction of groundwater level variations in a changing climate: A danish case study. ISPRS Int. J. Geo-Information 10. https://doi.org/10.3390/ijgi10110792.

Guo, M., Yue, W., Wang, T., Zheng, N., Wu, L., 2021. Assessing the use of standardized groundwater index for quantifying groundwater drought over the conterminous US. Journal of Hydrology 598, 126227. https://doi.org/10.1016/j.jhydrol.2021.126227.

Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. IEEE Transactions on Neural Networks 5, 989–993. https://doi.org/10.1109/72.329697.

Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen Netzen. Inst. Für Inform. Tech. Univ. Munchen 1–71.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Idrizovic, D., Pocuca, V., Vujadinovic Mandic, M., Djurovic, N., Matovic, G., Gregoric, E., 2020. Impact of climate change on water resource availability in a mountainous catchment: A case study of the Toplica River catchment. Serbia. J. Hydrol. 587, 124992 https://doi.org/10.1016/j.jhydrol.2020.124992.

IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3-24. https://doi.org/10.1017/9781009157940.001.

IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. EURO-CORDEX: New high-resolution climate change projections for European impact research. Regional Environmental Change 14, 563–578. https://doi.org/10.1007/s10113-013-0499-2.

Javadinejad, S., Dara, R., Jafary, F., 2020. How groundwater level can predict under the effect of climate change by using artificial neural networks of NARX. Resour. Environ. Inf. Eng. 2, 90–99. https://doi.org/10.25082/reie.2020.01.005.

Jeihouni, E., Mohammadi, M., Eslamian, S., Zareian, M.J., 2019. Potential impacts of climate change on groundwater level through hybrid soft-computing methods: a case study—Shabestar Plain. Iran. Environ. Monit. Assess. 191 https://doi.org/10.1007/s10661-019-7784-6.

Jiménez Cisneros, B.E., Oki, T., Arnell, N.W., Benito, G., Cogley, J.G., Döll, P., Jiang, T., Mwakalila, S.S., Kundzewicz, Z., Nishijima, A., 2015. Freshwater resources. Clim. Chang. 2014 Impacts, Adapt. Vulnerability Part A Glob. Sect. Asp. 229–270. https://doi.org/10.1017/CBO9781107415379.008.

Karthikeyan, L., Kumar, D.N., Graillot, D., Gaur, S., 2013. Prediction of Ground Water Levels in the Uplands of a Tropical Coastal Riparian Wetland using Artificial Neural Networks. Water Resources Management 27, 871–883. https://doi.org/10.1007/s11269-012-0220-0.

Khan, S., Gabriel, H.F., Rana, T., 2008. Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. Irrigation and Drainage Systems 22, 159–177. https://doi.org/10.1007/s10795-008-9049-3.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–15.

Kumar, R., Musuuza, J.L., Van Loon, A.F., Teuling, A.J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., Attinger, S., 2016. Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator. Hydrology and Earth System Sciences 20, 1117–1131. https://doi.org/10.5194/hess-20-1117-2016.

Lallahem, S., Mania, J., Hani, A., Najjar, Y., 2005. On the use of neural networks to evaluate groundwater levels in fractured media. Journal of Hydrology 307, 92–111. https://doi.org/10.1016/j.jhydrol.2004.10.005.

Lee, S., Lee, K.K., Yoon, H., 2019. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeology Journal 27, 567–579. https://doi.org/10.1007/s10040-018-1866-3.

Leelaruban, N., Padmanabhan, G., Oduor, P., 2017. Examining the relationship between drought indices and groundwater levels. Water (switzerland) 9. https://doi.org/10.3390/w9020082.

Malcolm, R., Soulsby, C., 2000. Modelling the potential impact of climate change on a shallow coastal aquifer in northern Scotland. Geological Society - Special Publications 182, 191–204. https://doi.org/10.1144/GSL.SP.2000.182.01.18.

Mohanty, S., Jha, M.K., Kumar, A., Sudheer, K.P., 2010. Artificial neural network modeling for groundwater level forecasting in a river island of eastern India. Water Resources Management 24, 1845–1865. https://doi.org/10.1007/s11269-009-9527-x.

Mohanty, S., Jha, M.K., Raul, S.K., Panda, R.K., Sudheer, K.P., 2015. Using Artificial Neural Network Approach for Simultaneous Forecasting of Weekly Groundwater Levels at Multiple Sites. Water Resources Management 29, 5521–5532. https://doi.org/10.1007/s11269-015-1132-6.

Nourani, V., Khodkar, K., Paknezhad, N.J., Laux, P., 2022. Deep learning-based uncertainty quantification of groundwater level predictions. Stoch. Environ. Res. Risk Assess. 36, 3081–3107. https://doi.org/10.1007/s00477-022-02181-7.

Omirl, 2021. Osservatorio Meteo Idrologico della Regione Liguria. accessed 9.01.22. https://omirl.regione.liguria.it.

Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. Journal of Hydrology 572, 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resources Research 48. https://doi.org/10.1029/2011WR011527.

Sahoo, S., Jha, M.K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: A comparative assessment. Hydrogeology Journal 21, 1865–1887. https://doi.org/10.1007/s10040-013-1029-5.

Secci, D., Tanda, M.G., D'Oria, M., Todaro, V., Fagandini, C., 2021. Impacts of climate change on groundwater droughts by means of standardized indices and regional climate models. Journal of Hydrology 603, 127154. https://doi.org/10.1016/j.jhydrol.2021.127154.

Shakiba, A.R., Cheshmi, A., 2013. Evaluation of the effect of climate change on groundwater resources of Ramhormoz plain using NARX neural network. Journal of Researches in Earth Sciences 2 (5), 46–57.

Shiri, J., Kisi, O., Yoon, H., Lee, K.K., Hossein Nazemi, A., 2013. Predicting groundwater level fluctuations with meteorological effect implications-A comparative study among soft computing techniques. Computational Geosciences 56, 32–44. https://doi.org/10.1016/j.cageo.2013.01.007.

Sir, 2021. Servizio Idrologico della Regione Toscana. accessed 9.01.22. https://www.sir.toscana.it.

Soleimani Motlagh, M., Ghasemieh, H., Talebi, A., Abdollahi, K., 2017. Identification and Analysis of Drought Propagation of Groundwater During Past and Future Periods. Water Resources Management 31, 109–125. https://doi.org/10.1007/s11269-016-1513-5.

Suryanarayana, C., Sudheer, C., Mahammood, V., Panigrahi, B.K., 2014. An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing 145, 324–335. https://doi.org/10.1016/j.neucom.2014.05.026.

Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddam, S., Sungwon, K., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., Allawi, M.F., Abba, S. I., Zain, J.M., Falah, M.W., Jamei, M., Bokde, N.D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S.K., Tiyasha, T., Khedher, K.M., Al-Ansari, N., Shahid, S., Yaseen, Z.M., 2022. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing 489, 271–308. https://doi.org/10.1016/j.neucom.2022.03.014.

Taormina, R., Chau, K.W., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence 25, 1670–1676. https://doi.org/10.1016/j.engappai.2012.02.009.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society. https://doi.org/10.1175/BAMS-D-11-00094.1.

Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology 456–457, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052.

Todaro, V., D'Oria, M., Secci, D., Zanini, A., Tanda, M.G., 2022. Climate Change over the Mediterranean Region: Local Temperature and Precipitation Variations at Five Pilot Sites. Water (switzerland) 14. https://doi.org/10.3390/w14162499.

Trichakis, I.C., Nikolos, I.K., Karatzas, G.P., 2011. Artificial Neural Network (ANN) Based Modeling for Karstic Groundwater Level Simulation. Water Resources Management 25, 1143–1152. https://doi.org/10.1007/s11269-010-9628-6.

Uddameri, V., Singaraju, S., Hernandez, E.A., 2019. Is Standardized Precipitation Index (SPI) a Useful Indicator to Forecast Groundwater Droughts? — Insights from a Karst Aquifer. Journal of the American Water Resources Association 55, 70–88. https://doi.org/10.1111/1752-1688.12698.

Van Loon, A.F., Kumar, R., Mishra, V., 2017. Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time. Hydrology and Earth System Sciences 21, 1947–1971. https://doi.org/10.5194/hess-21-1947-2017.

Varouchakis, E.A., Hristopulos, D.T., 2013. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. Advances in Water Resources 52, 34–49. https://doi.org/10.1016/j.advwatres.2012.08.002.

Wunsch, A., Liesch, T., Broda, S., 2022. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. Nature Communications 13, 1–13. https://doi.org/10.1038/s41467-022-28770-2.

Yan, Q., Ma, C., 2016. Application of integrated ARIMA and RBF network for groundwater level forecasting. Environment and Earth Science 75, 1–13. https://doi.org/10.1007/s12665-015-5198-5.

Yang, X., Zhang, Z., 2022. A CNN-LSTM Model Based on a Meta-Learning Algorithm to Predict Groundwater Level in the Middle and Lower Reaches of the Heihe River. China. Water (switzerland) 14. https://doi.org/10.3390/w14152377.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., Lee, K.K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology 396, 128–138. https://doi.org/10.1016/j.jhydrol.2010.11.002.