Comparison of a Semi-Physical Model and an Artificial Neural Network for Simulating Past Groundwater Levels, in a Karst Aquifer in Southern France

Sayane Jouanneau-Couvret[1)], Jose-Manuel Tunqui Neira[2)], Sébastien Morilhat[3)], Séverin Pistre[4)], Anne Johannet[5*)]

[1)] CEA, DES, IRESNE, Nuclear Technology Department, Cadarache, F-13108, Saint Paul les Durance, France; HSM, Univ. Montpellier, IMT Mines Ales, CNRS, IRD.Ales, France; sayane.jouanneau-couvret@cea.fr; 0009-0008-8344-7524
[2)] HSM, Univ. Montpellier, IMT Mines Ales, CNRS, IRD.Ales, France; jose-manuel.tunqui-neira@mines-ales.fr; 0000-0002-6812-9957
[3)] CEA, DES, IRESNE, Nuclear Technology Department, Cadarache, F-13108, Saint Paul les Durance, France; sebastien.morilhat@cea.fr
[4)] HSM, Univ. Montpellier, CNRS, IMT, IRD - 163 rue Auguste Broussonnet, 34090 Montpellier, France; severin.pistre@umontpellier.fr; 0000-0002-6301-8801
[5*)] HSM, Univ. Montpellier, IMT Mines Ales, CNRS, IRD.Ales, France; anne.johannet@mines-ales.fr; 0000-0001-5717-9871

https://doi.org/10.29227/IM-2026-02-02-014

Submission date: *2026-17-04* | Review date: *2026-30-04*

Abstract

Because of their strongly heterogeneous permeability, karst aquifers exhibit nonlinear relations between their outputs (discharge or groundwater level) and inputs (rain, temperature, etc.). Three main types of models are commonly used to simulate karst aquifers: physical models, pipe-flow models and conceptual models. The "Méthode d'Évaluation des Crues Karstiques" or Method for Assessing Karst Floods (MECK) was implemented to perform the simulation of Cadarache's fractured karst aquifer's level (south of France). It relies on a semi-physical model that conceptualizes the functioning of the aquifer. Even though it simulates a continuous daily time series, the model's optimization focuses on high levels for enhanced flood representation. Although MECK shows good performances on high groundwater levels, it encounters difficulties under specific conditions. The model struggles to accurately reproduce level variations caused by recent rainfall events. Despite having a large number of parameters, this semi-physical model is not able to reproduce the complex phenomena that govern the response of a karst aquifer. Artificial Neural Networks, however, are able to learn the system's behavior directly from the data, without the operator needing. Most neural networks developed in the field of hydro(geo)logy are used to simulate spring discharge based on rain. Fewer studies focus on simulating groundwater levels, and among those, only a small subset specifically addresses karst aquifers. The Cadarache site belonging to the French Atomic Energy and Alternative Energies Commission (CEA) in southeastern France is subject to rapid kinetic groundwater flooding (several tens of meters rise in a few hours). With the aim of estimating the 100-year flood level for enhancing the safety of nuclear facilities, the goal of this study is to compare the performance of MECK against a Neural Network in simulating past groundwater level of a fractured karst aquifer, with a particular emphasis on high levels related to flooding. Daily groundwater levels are simulated from 1960 to 2025, using meteorological data. The data are pre-processed based on cross correlation analysis and stationarity tests. The neural network, a Multilayer Perceptron, is then developed focusing on the high levels of the groundwater. Its architecture is selected through a trial-and-error approach, and the hyperparameters are selected by optimizing cross-validation metrics. Simulated groundwater levels of MECK and the developed Neural Network are finally compared using classical hydrological metrics (NSE, RMSE, etc.), both for the entire time series, and for high levels. Results show that the Neural Network performs overall better than the semi-physical model. Whereas the levels simulated by the semi-physical model are mostly higher than the measured ones, the Artificial Neural Network reproduces accurately the overall dynamic and shows good performances on high levels but struggles to reproduce intermediate levels.

Keywords: *Karst, Flooding, Modeling, Artificial Neural Network, Reservoir model*

1. Introduction

The term "karst" is used to define a "region made of carbonated, compact and soluble rocks, which exhibit characteristic superficial and underground forms" [1]. Karstification occurs when acid and unsaturated in carbonates water, flows through carbonated rocks [2], creating highly localized voids of varying sizes and arranged in specific pattern. Karst are thus characterized by an important spatial variation of permeability, as water flows through fractures, pipes and the rock matrix. Excluding ice-covered areas, karst aquifer can be found on 15,2% of continental surface (20 274 km²) [3], and supply around one quarter of the world's population of drinking water [4]. Yet, because of its heterogeneity, karst is particularly vulnerable to pollutions [5] and may be subject to groundwater flooding or slope movements [6]. Karst simulation is a major issue, to better understanding of its functioning, but also to prevent risks and ensure the sustainable management of water resources. Many models have been developed to simulate karst's processes. Commonly, we distinguish three main types of models: physical models, pipe-flow models and data-driven models. Physical models use meshes to represent the spatial and temporal variations of physical and measurable parameters. In hydrogeology, these parameters are mainly acquired from pumping-tests and permit to apply multi-porosity concepts to the model. These models aim to accurately represent the reality [7]. Pipe-flow models simulate the flow in pipes generated by controlled stochastics hydraulic laws such as Saint-Venant or Manning-Strickler [8,9]. Data-driven models include conceptual reservoirs models and nonparametric transfer functions such as Artificial Neural Networks (ANN). Most data-driven models transform an input in an output without

considering physical process that control the system. Some data-driven models may include physical information, such as Physics-Informed Neural Networks (PINNs). Choice of the model depends on its purpose, scale and data availability. However, the development and optimization of a karst model can be quite challenging, as relations between their inputs (discharge or groundwater level) and their outputs (rainfall, temperature, etc.) is nonlinear [10,11].

The Cadarache center of French Alternative Energies & Atomic Energy Commission (CEA) exhibits a karst Cretaceous fractured aquifer. Its fracturing is very heterogeneous, and the locations of faults and their impacts on groundwater variations at the local scale are poorly known. Moreover, this aquifer is subject to flooding by groundwater rising [12]. During November 2011, 250 mm of rain was measured in 48 hours, causing the rise of the Cretaceous aquifer's level of around 40 meters. Artesianism was locally observed, and the pumps of a few installations were activated. A semi-physical reservoir model is currently used at Cadarache. The model "Méthode d'Evaluation des Crues Karstiques" (Method for Assessing Karst Floods), named MECK [13], was developed to simulate the Cretaceous aquifer's past groundwater levels, with a focus on high levels. For regulatory purposes, the simulations are then used to calculate the level reached during a 100-year flood. Though the model simulates accurately flood peaks, it struggles to reproduce the level's variations following close rainfall events. Also, MECK's actual configuration does not permit to transcript the change of dynamic that occurs when water flows through ensembles with different hydrogeological properties.

Physical models usually need a complex parametrization to accurately simulate groundwater levels. However, ANN are able to learn the non-linear behavior of the system and generalize it, meaning they can accurately predict data not represented in the training set [14]. Most ANN developed in the field on hydrogeology rely on rainfall to simulate discharge, only a few simulate groundwater levels. Among them, a little number of studies specifically addresses karst aquifers [15,16], and a majority use ANN to predict future groundwater in order to prevent flooding [12,17].

This study aims to explore the capabilities of an ANN for simulating past groundwater levels, with a particular focus on high levels. This analysis is intended to estimate the 100-year flood level for enhancing the safety of nuclear facilities. An ANN, more specifically a Multilayer Perceptron (MLP), is developed and its simulations are compared with those of the semi-physical reservoir model used at Cadarache.

The second section of the paper presents the materials and methods, including the study site, the data used as inputs for the ANN and their pre-processing. The functioning of the semi-physical model and the development of the ANN are also described in this section. The third section presents the performances of the ANN for groundwater level simulation, and more specifically the simulation of high levels. The performances of the two models are then compared. The third section is followed by a discussion.

2. Materials and Methods

2.1 Study Site

The Cadarache center of French Alternative Energies & Atomic Energy Commission (CEA) is located in the south eastern of France (Saint Paul-Lez-Durance). It is an energy research and technology development center, focused on nuclear and new energies. In accordance with French regulations, the CEA must demonstrate the safety of its facilities during safety reviews conducted every 10 years. Several risks are taken into account during these reviews: earthquakes, soil liquefaction, flooding, etc. The combination of Mediterranean climate with a karst aquifer makes the CEA of Cadarache highly susceptible to groundwater flooding. As part of the periodic safety reviews, the risk of flooding by groundwater rising of the nuclear facilities is assessed. For this purpose, the groundwater level associated with a 100-year flood is calculated using a groundwater level time series, simulated from 1960. This time series is generated with a semi-physical reservoir model called MECK, based on meteorological data.

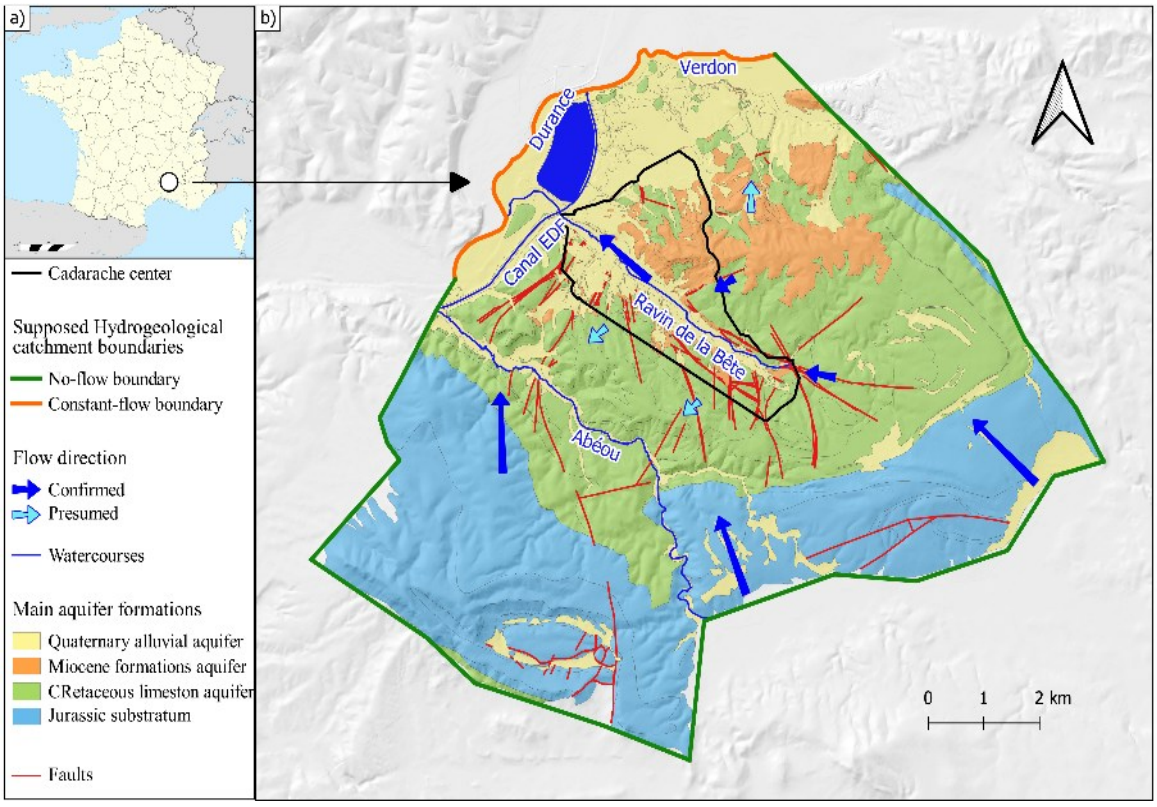

Fig. 1. (a) Localization and (b) map of the study hydrogeological catchment

Recently, a neural network was developed to predict future groundwater levels in the Cretaceous aquifer [12]. This model is used at the Cadarache Center to provide early warnings to nuclear facilities about potential flooding events.

Over the period 1960-2013, the mean annual rainfall was 651 mm for 83 rainy days. Strong rainfalls (10 $mm.h^{-1}$) are often observed, which may be very localized during storm events. Between 1966 and 2013, daily rainfall exceeded 40 mm for 79 days, and the 100-year daily rainfall is estimated at 180 mm. Recorded rainfall events thus vary in duration and intensity.

Fig. 1 presents the hydrogeological context of the Cadarache center of CEA.

The center is located in a Provençal geological context and belongs to the structural domain of the intermediary zone between external alpine zone and Pyrenean-Provençal domain. Cadarache's hydrogeological catchment measures 100 km² and is made up of three aquifers. Their outlet is the Durance River, at the north west of the catchment.

The Cretaceous limestone aquifer is a fractured aquifer and shows locally signs of karstification. Discordant Miocene formations' aquifer locally present above exhibits several geological facies, that ranges from sandy-sandstone formations to marl-clay formations. The above Quaternary aquifer is made of alluvial deposits, such as pebbles, gravel, sand or silt [18]. The three aquifers are hydraulically connected, and drainage or pressurization phenomena may occur depending on the hydrological regime (high or low water levels). Quaternary and Miocene aquifers exhibit moderate permeabilities, of 2.10^{-4}m/s et 10^{-5} m/s respectively. Those aquifers thus show cushioned and low amplitude variations. These aquifers are not very reactive to rainfall and are not subject to rapid flooding by groundwater rise. In contrast, the Cretaceous aquifer has a very low permeability of 10^{-13} m/s in its matrix, and high permeabilities in fractured and faulted areas, up to 10^{-4} m/s [18]. High and rapid level variations are thus measured in this aquifer following rainfall. Because of its dense network of fractures and its important reactivity, the Cretaceous aquifer is subject to rapid flooding by groundwater rise. The location of its faults, their width, and their connectivity are poorly known. It is thus difficult to establish the response of the aquifer to rainfall at a local scale. For instance, two nearby piezometers may record very different variations in the groundwater level depending on whether they are located in a fractured zone or in a matrix zone.

2.2 Dataset: sources and pre-processing

Data used for the ANN come from three sources: Cadarache's recordings, E-OBS European Program and SAFRAN-ISBA-MODCOU's (SIM) outputs. Rainfall and groundwater levels time series used for the ANN were recorded at Cadarache. Rainfall record was set up using manual measures from 1960 to 2006, and automatic 10 min record since 2007, summed daily from 6 to 6.

The site features a dense network of piezometers, with groundwater levels measured since the 2000s. Groundwater levels are expressed in meters Above Sea Level (MASL), referenced to the Marseille datum. The daily groundwater level record used for the ANN was constructed by extracting the daily maximum values recorded at 30 min intervals. Because they were not measured before 2006 at Cadarache, daily air temperature and relative air humidity records are extracted from the E-OBS dataset since 1960. Sea-level pressure, global radiation, and wind velocity are also extracted. E-OBS is a European daily meteorological dataset, constructed with kriging from meteorological stations' measures [19]. Data are available for mesh size 0.1° x 0.1°. Other hydrometeorological variables are extracted from Météo France's meteorological, land surface and hydrogeological model SIM [20]. Daily Potential Evapotranspiration (PET), Soil Wetness Index (SWI), and drainage are extracted since 1960. Data are available for mesh size 8 x 8 km. Both SIM and E-BOS data are extracted from the mesh located at longitude 5.75980847 and latitude 43.6893644, corresponding to the center of the hydrogeological catchment of Cadarache.

The piezometric data used in this study come from a piezometer located in the center of the catchment which records the levels of the Cretaceous aquifer**.** Its records have been available since 2002, with a gap during the 2008–2009 period. The piezometer is located at the bottom of a small valley and drains water from the surrounding hills. It thus shows a high reactivity to rainfall. Two artesianism events have been recorded since the 2000s, in November 2011 and November 2019 following heavy rainfall. During these events, the groundwater level reached the soil level, at 330 MASL. This piezometer also exhibits a "bottle-neck effect": the groundwater levels do not decline further 290 MASL, even during particularly dry periods. Gaps in the groundwater level record have been filled using the record of a close piezometer, for which a high Pearson correlation was calculated (0.98).

The relative air humidity and air temperature data from E-OBS and Cadarache are compared on 2007-2025 to ensure that the E-OBS kriging interpolations are consistent with the measured data in Cadarache. The stationarity of the time series is verified using several tests: Pettitt, Standard Normal Homogeneity Test and Buishand Range Test. Those tests detect a non-stationarity when there is a break in the mean of the time series. They allow to determine the precise date at which the break occurs. When the time series exhibits non-stationarity, the potential causes are identified where possible. If considered appropriate, the time series may be corrected (for the input data), or the portion preceding the break may be removed (for the output data). A break was detected in the groundwater level record in 2008. The mean daily level is lower during the 2002–2008 period than during 2009–2024. This difference can be attributed to particularly dry conditions between 2003 and 2007, with a mean annual rainfall of 467.28 mm, significantly below the average of 623 mm during 1960–2024. Given the relatively short length of the groundwater level record, the detected break in 2008 is likely linked to this dry period. Thus, no correction is applied to the groundwater record. A break was also detected in 2013, on the record made up of annual rainy days. That year, the 0.5 mm rain gauge was replaced with a 0.3 mm 2013. Daily tips were thus recorded, caused by morning dew rather that rainfall. Replacing these daily values of 0.3 by 0 allowed to correct the non-stationarity in the record. Breaks were also detected in the air temperature, humidity and SWI records, for which it is more difficult to attribute a specific cause. It is more likely that climate change has a major influence on the detection of these breaks, but other phenomenon may also interfere with these hydrometeorological data, such as deforestation or anthropic activities. No corrections were therefore applied to theses timeseries.

Correlation analysis applied to karst hydrosystems provide a better comprehension of their the global functioning [21]. Simple correlation is performed by comparing a time series to itself while applying time lags. It gives information about the hydrosystem's temporal dependency, naming its "memory effect". Cross correlation is performed by comparing an input's timeseries, such as rainfall, to an output's timeseries, such as groundwater level, while applying time lags. It gives information about the "response time" of the output to the input, and its memory effect, that is the amount of time before the series are no longer dependent. Cross correlations are performed between the groundwater level record of the study piezometer, and the other hydrometeorological available variables. They are performed over calendar years (from October to September), from 2010 to 2019. The results of the correlation analyses are not presented in this article for the sake of conciseness. The input variables for the ANN, along with their temporal windows, are selected based on the results of cross-correlation analyses. The variables with high maximum correlations will be tested as inputs for the neural network, while variables with low maximum correlations will not be used as they do not seem to give further information, and will probably introduce noise.

2.3 Semi-Physical Reservoir Model

"Méthode d'Evaluation Des Crues Karstiques" (MECK) is a semi-physical reservoir model, developed to reproduce the behavior of a karst aquifer with a quick kinetic, from meteorological data [13]. Though it was designed to simulate a continuous time series, its parameters optimization is made on the aquifer's high levels. MECK allows to conceptualize the relations between climatic variables and the aquifer's level through manually calibrated parameters. MECK is made of two reservoirs that represent the soil and the aquifer. Depending on the piezometer, it is also possible to add a third compartment that translates a different behavior of the aquifer under and below a given level. This situation may occur, for instance, when a mature epikarst is observed above a deep and less fractured zone.

MECK's parameters do not have a physical correspondence and cannot be measured. Therefore, it is not possible to verify them and ensure that the retained values are optimal. Thus, estimating the parameters requires a manual calibration procedure performed by the user. The calibration is done by visually comparing the simulated and the measured timeseries. Although the simulation is performed over a daily continuous time series, MECK's parameters are chosen to prioritize the accurate simulation of high levels. The variables used as inputs are rainfall, air temperature and relative humidity measured at Cadarache. As for the output, the reference piezometer used for the simulation must be representative of the observed high levels reached by the aquifer in the study zone. Simulated groundwater level record is generated by iteration through 3 steps repeated at each time: (1) calculation of the Potential Evapotransiration (PET), (2) calculation of the infiltration and (3) calculation of the groundwater level.

The version of MECK used in this study is the version that is used for operational studies realized by the CEA. The calibration is performed over the period 2002-2023, and the parameters obtained are used to reproduce the groundwater levels between 1960 and 2001. The rainfall, relative humidity and air temperature records used for operational studies are different from the ones used for the ANN. To ensure consistency in this study and for a more accurate comparison between the MECK and ANN simulations, the same inputs data are used for both models, except for the PET record. The PET is automatically calculated by MECK based on Thornthwaite method [22], which uses daily air temperature.

2.4 Neural Network development

2.4.1 Model's Architecture. A neuron is a mathematical operator that applies a nonlinear transformation, called an activation function, to a weighted sum of its inputs, which includes a bias term. An ANN is formed by combining several neurons. **Multilayer Perceptron (MLP), in its most basic form, is made of an entry layer, at least one hidden layer and an output layer (Fig. 1).** With just a single hidden layer, the MLP is capable of approximating any differentiable function **[23]**. We say it is a universal approximator. The MLP is also parsimonious: its number of parameters grows proportionally with the number of entries, unlike classical nonlinear statistical models for which this relation is exponential **[24]**.

In hydro(geo)logy, the dynamic of the systems is highly influenced by seasonality, and hydrometeorological variables at a given time are influenced by their previous values. It is thus necessary to address the time dimension. Static MLP do not consider the previous state of the system for their simulations (Fig. 2). However dynamic MLP such as Feed Forward Neural Network (FFNN) or Recurrent Neural network (RNN) are designed to deal with time series. In FFNN, all of the inputs are exogenous (external to the model) and some of them correspond to past values of rainfall or discharge. A RNN contains loops that redirect its outputs to its inputs, RNN can receive both exogenous and recurrent inputs.

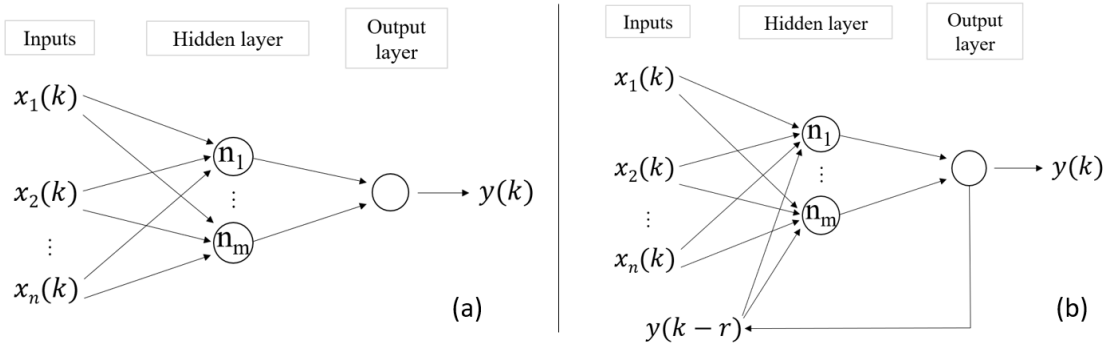

Fig. 1. Representation of a static Multilayer Perceptron (a) and a dynamic Multilayer Perceptron (b)

2.4.2 Training. The ANN is trained over a period for which both measured inputs and output data are available. The parameters found over the training set are then used to perform the simulation over another set, usually called test. The ability of the model to well predict the values of the test set is called generalization. The cost function defines the global error between the outputs of the ANN and the desired measured values associated as a function of the ANN parameters.

Learning rule aims to diminish the cost function. It defines the way the parameters are adjusted during the training through iterations. Levenberg-Marquardt (LM) learning rule [25,26] is frequently used to optimize MLP because it converges at a high rate, even on complex ANN. By using the Mean Squared Error (MSE) cost function, it is possible to apply the LM algorithm. The LM rule approximates the second derivatives of the cost function with respect to parameters using the first derivatives of the cost function. The parameters are then updated by solving an equation that integrates the curvature and the gradient of the first derivatives. The aptitude of the ANN to generalize is finally verified over the 'test' set.

2.4.3 Regularization Methods. Because they are trained by comparing their simulated outputs to measured ones, ANN are subject to under or overfitting. The model is said to overfit when it not only learns the relations between the inputs and outputs, but also the noise included in data [27]. When overfitting occurs, the model is highly performant on training set but does not perform well on test set. Variance (variability of predictions on test set) is then higher than bias (the mean difference between predicted and measured outputs). An ANN with good generalization shows good performance both on training and on test sets, different from the training data. Regularization methods are used to limit overfitting, and enhance ANN generalization. These methods are particularly important in hydrogeology, where there are many unknowns, leading to significant uncertainties.

Cross-validation is a regularization method, used to estimate the mean generalization error [28]. The training set is divided into K subsets. *K* training runs are performed, each on *K*-1 subsets, followed by a generalization test on the remaining subset, known as the validation set. Each subset is used once as the validation subset. The cross-validation score is finally calculated as the median of the scores obtained for each validation subset.

Early stopping [29] is one of the most used regularization methods. In addition to the training and test sets, a stop set is extracted from the database. During training, the cost function is calculated on the training and stop sets. The model is able to generalize as long as the cost function diminishes for both sets. Training is stopped when the cost function of the stop set increases, meaning an overfitting of the model.

Ensemble models [30] aggregate the results of multiple models into one model, whose generalization capabilities are enhanced. Because the simulations of an MLP depend on its parameters' initialization, creating an ensemble model that combines model with different initializations allows a better generalization. The ensemble ANN is there assembled by calculating the median of ten different models at each time step.

2.4.4 Quality criteria. Two scores are used to measure the quality of the models.

1) Nash-Sutcliffe Efficiency (NSE) criterion [31]. It is widely used in hydrology, and relates the Mean Squared Error (MSE) to the variance of the observed measures:

$$NSE = 1 - \frac{\sum_{k=1}^{n}\left(y_p(k) - y(k)\right)^2}{\sum_{k=1}^{n}\left(y_p(k) - \overline{y_p}\right)^2} \quad (1)$$

With $y_p(k)$ the measured groundwater level at time k; $y(k)$ the groundwater level simulated by the model at time k; $\overline{y_p}$ the mean of the measured groundwater levels of the data set; n the length of the data vector on which the criterion is calculated. The value of the NSE is 1 when the simulated values are equal to the measured ones. The value of the NSE is 0 when the model does not provide better performances than the mean of the measures. In hydrology, we consider the model is good when the NSE is over 0.8.

2) Mean Root Square Error (RMSE) (2):

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}\left(y(k) - y_p(k)\right)^2}{n}} \quad (2)$$

With the same notations as previously. The RMSE is expressed in the same unity as the study variable. The larger the RMSE value, the greater the error is.

2.4.5 Variable Selection. Variables are selected through a trial-and-error approach: variables pre-selected with correlation analysis are applied to the model. At each step of the selection, only the variables that improve the cross-validation score are kept in order to adjust the model complexity.

To specifically select the model architecture that accurately simulates high groundwater levels, the Nash criterion is calculated only for groundwater levels that exceed 291 MASL. This threshold ensures that only variations in groundwater levels linked to rainfall events are considered, excluding the 'bottleneck effect' observed during dry periods.

The necessary timeline (temporal window) of the selected variables is determined thanks to correlation analysis and cross-validation scores. Variables selected are daily rainfall, cumulative rainfall over 30 days, daily PET, mean monthly temperature and daily Soil Wetness Index (SWI). The other hyperparameters (number of neurons of the hidden layer, slope of the activation functions, bias and recurrence) are also selected by maximizing the score obtained in cross-validation. Cross validation, test and stop ensemble are delimited over two hydrological years (from September to August) so that they capture long-term trends. The period 2012-2013 is chosen as the test ensemble, because the highest groundwater levels of the aquifer were recorded in November 2011 (included in hydrological year 2012). The stop ensemble is chosen based on the results of cross-validation, and corresponds to the ensemble with the highest criteria, here the 2002-2003 period. Training is thus realized over 2004-2007, 2010-2011, 2014-2023. Ultimately, the forecast is obtained using an ensemble model, each member of which is identical in architecture to the selected model but differs in terms of parameter initialization. This improves the robustness of the resulting model. The final simulations correspond to the median, at each time-step of the output of each member. The architecture of the ANN (one member) and its hyperparameters are presented in Fig. 3. The average early stopping epoch was reached at around 57 epochs.

2.5 Comparison of the Performance of the Semi-Physical Model and the ANN

The simulations of the semi-physical model and the Artificial Neural Network are compared both over the whole dataset, and over high groundwater levels. To this purpose, two criteria are used: the NSE criteria, and the Root Mean Squared Error (section 2.4.5).

The RMSE and NSE are calculated for several datasets filtered by a threshold: the criteria are applied to all the groundwater levels exceeding the chosen measured threshold. This allows for the comparison of model performance based on hydrological conditions: overall, and specifically for intermediate and high groundwater levels. While low thresholds give information about the overall dynamic of the aquifer, high thresholds provide insight into the aquifer's responses to high rainfall and the model's ability to reproduce the amplitude of the loading following major rainfall events.

Cumulative frequency curves of the three records enable the identification of discrepancies between simulated and measured

time series. More specifically, by visually comparing the measured and simulated frequency curves, it is possible to determine whether the model underestimates or overestimates a specific range of groundwater levels.

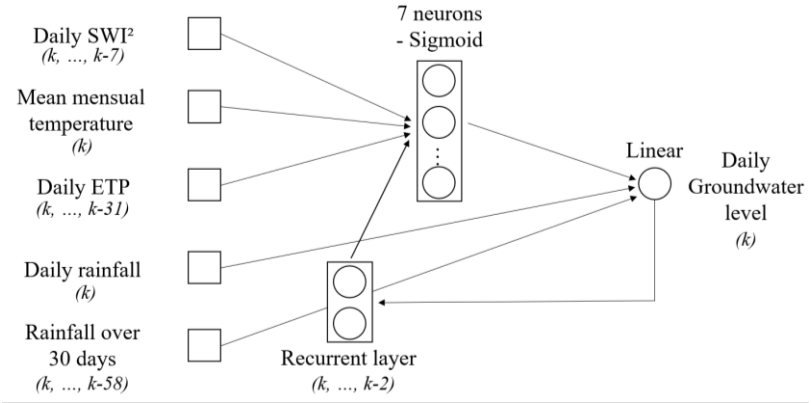

Fig. 2. Representation of the developed ANN's architecture

3. Results

3.1 Simulations of the Artificial Neural Network

Fig. 4 shows the groundwater levels simulated by the ANN over a subset of the training set for improved readability, though this subset is representative of the training set. Overall, the model reproduces the behavior of the aquifer with satisfactory accuracy. The dynamic is precisely captured: the loading of the water table following rainfall occurs on time, even though it tends to underestimate the threshold observed at 290 MASL caused by the bottle-neck effect. As for the amplitude of the peaks, the simulated peaks are close to the measured ones, except for low loadings. With groundwater levels ranging between 290 and 330 MASL, the model struggles to reproduce the low loadings between 290 and 312 MASL. During high loadings, the major phenomenon that controls the variations of the aquifer is the infiltration of rainfall, whereas during low loadings following intermediary rainfall event, other phenomena have a strong influence on the variations of the levels, such as exchanges between aquifers, contributions from a fault, etc.

Two artesianisms events are observed over the period 2002-2023, during for which the groundwater level reaches the soil (at 330 MASL). These events are overestimated by the ANN. This is a normal behavior of the ANN, since it is not provided with this information. The ANN cannot detect the presence of an overflow threshold, especially since this type of phenomenon occurs very rarely. These events should therefore not be interpreted as simulation errors.

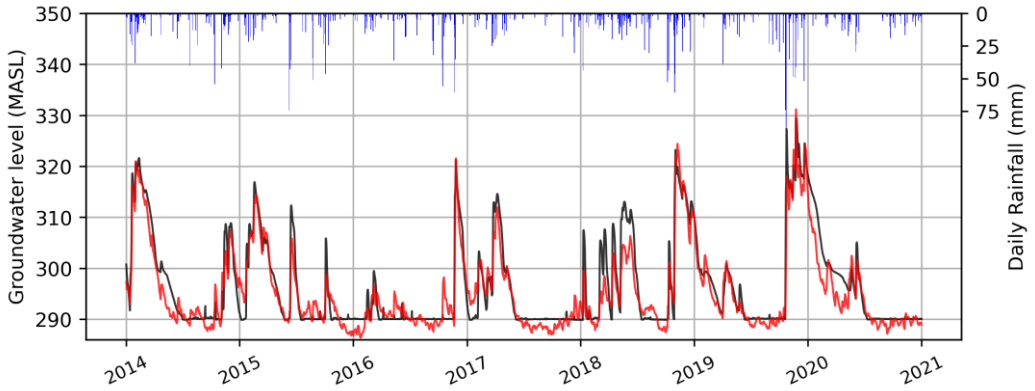

Fig. 4. Simulations of the ANN on a subset (2014-2020) of the training set

The simulations of the ANN are accurate on all data sets (Table 1). Notably, the high scores obtained on the validation set which is very similar to the test and training scores demonstrate that the model has a good generalization capability and is able to simulate data that are not represented in the training set. This is remarkable on the test, which includes an artesianism event.

Tab. 1. NSE obtained for the partitioned sets of the ANN

	Training	**Test**	**Stop**	**Validation**	**Whole period**
NSE ANN	0.86	0.87	0.89	0.82	0.87

3.2 Comparison of the Simulations of the Artificial Neural Network with the Simulations of MECK

Fig. 5 compares the groundwater levels simulated by MECK and the ANN with the measured levels, during 2010-2012 (test ensemble of the ANN). The dynamics reproduced by the ANN are closer to measurements than those of MECK, with NSE scores of 0.87 and 0.60 respectively. In fact, the recession curves simulated by MECK follow a decreasing exponential trend, which is adjustable using a single parameter. Since MECK was not designed to simulate recessions, it shows low performances in reproducing these processes. Even though the ANN is optimized over high levels, it stills captures the overall dynamic of the aquifer.

Also, MECK includes a parameter that defines the minimum groundwater level of the aquifer. This explains the excellent simulations obtained for levels around 290 MASL. Similar to the lack of information concerning the soil level at 330 MASL, the ANN is not provided with the information of threshold at 290 MASL. Consequently, the simulations of the ANN for these low groundwater levels are not as accurate at those generated by MECK. These thresholds could be applied as a post-processing step to the ANN model.

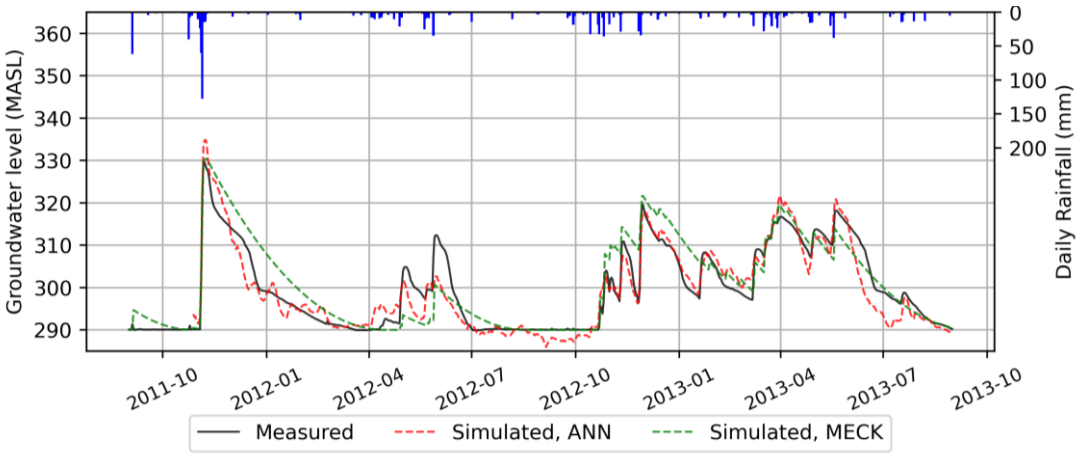

Fig. 5. Comparison of the Groundwater levels simulated by MECK and by the ANN. Test Ensemble of the ANN

Fig. 6 (a) displays the violin plots of the three timeseries. The violin plots show that the distributions of the three time series are similar. Although most of groundwater level values are close to 290 MASL for the three time series, the median of the simulated levels by MECK is higher than that of the ANN, which is higher than the measured levels. As for high groundwater levels (above 300 MASL), their distributions are very similar for the measured and the ANN time series, but MECK simulates proportionally more high ground water levels. Fig. 6 (b) compares the cumulative frequencies of the groundwater levels simulated by MECK and the ANN with the measured ones, over 2002-2023. It also appears on cumulative frequencies that the groundwater levels simulated by MECK are most of the time higher than the ones observed, whereas levels simulated by the ANN are, as said in the previous section, closer to the measured levels excepted for low levels (around 290 MASL).

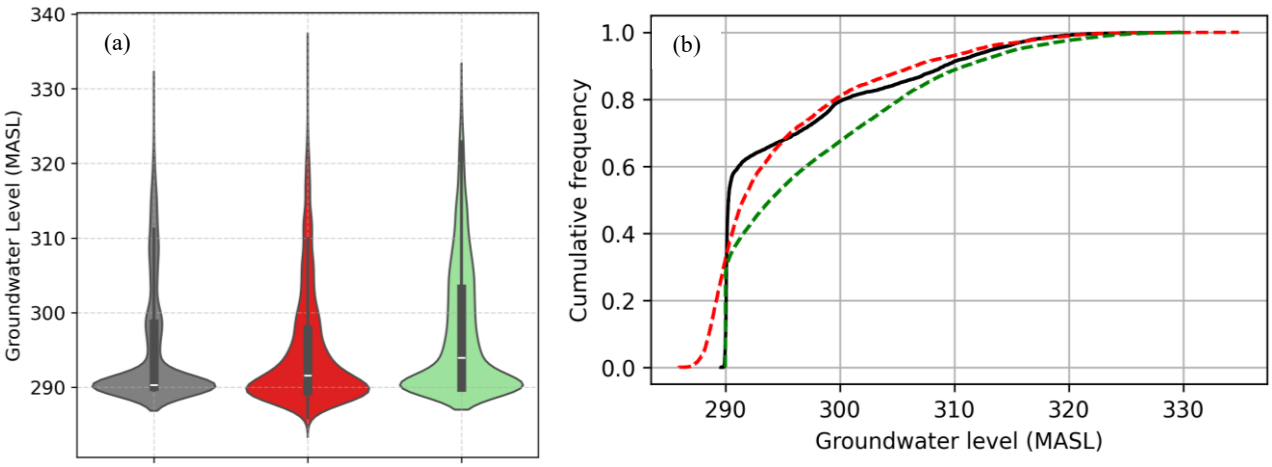

Fig. 6. (a) Violin plots showing the median (white line), and (b) Cumulative frequency of the groundwater levels, for the measured record (black), simulated by ANN record (red) and simulated by MECK record (green). Violin plots provide a representation of the statistical properties of a dataset. They visualize the data density, represented by the width of the violin at a given groundwater level.

Finally, Fig. 7 (a) and (b) display respectively the NSE and RMSE obtained for the groundwater level exceeding thresholds ranging from 290 MASL (bottle-neck effect) to 323 MASL (7 meters below the soil level). The scores obtained with the time series simulated by the ANN are generally better than those obtained with MECK, except for measured groundwater levels exceeding 318 MASL for which MECK reproduces more accurately the measured groundwater levels. These results are not surprising, since MECK is specifically calibrated on high levels measured during the artesianism events of 2011 and 2019. Additionally, as discussed in Section 3.1, the levels simulated by the ANN during artesianism events are higher than the measured ones. This overestimation contributes to degrade the scores of the ANN for high groundwater levels.

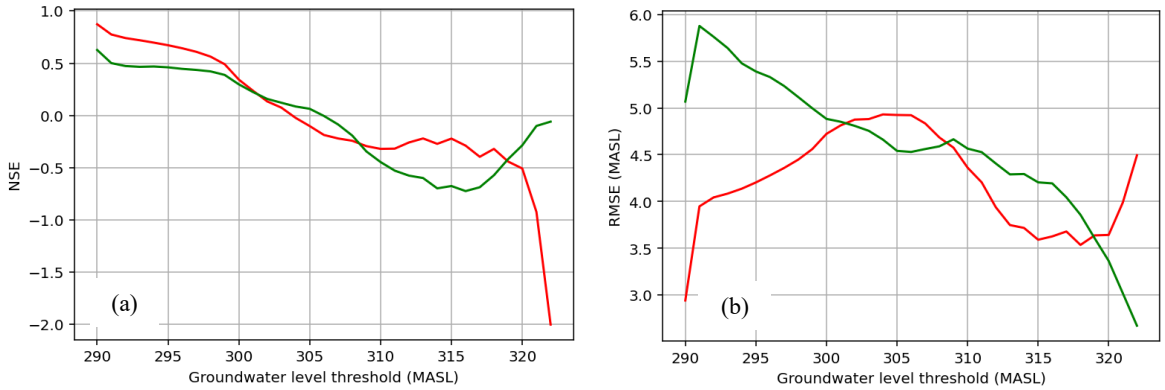

Fig. 7. (a) NSE and (b) RMSE calculated over groundwater levels simulated by ANN record (red) and simulated by MECK record (green), exceeding a range of thresholds

4. Discussion

The results show that an ANN outperform the semi-physical model in simulating groundwater dynamics, including during flood events, especially when the study aquifer is complex, and when little is known about its local hydrogeological properties.
However, it should be noted that the version of MECK used in this study could be improved. Its parameters are selected based on a manual trial-and-error approach. Yet, the simulations depend heavily on the parameters chosen, which may cancel each other out. Therefore, the simulations of MECK may vary depending on the operator who performs the calibration. The parameters retained for MECK for operational studies tend to overestimate the groundwater levels. In the future the calibration could be optimized by developing a methodology for automatically calibrating parameters, for instance by using Bayesian inferences. Adding additional parameters to simulate a piezometer with N thresholds instead of just two would also allow for a more detailed representation of groundwater behavior.

The ANN also requires further improvements: peaks in the groundwater levels with low to intermediate loadings are not accurately simulated. During high loadings, the primary phenomenon controlling level variations is rainfall. However, during low loadings, multiple phenomena can strongly influence the levels variations, such as contributions from faults or from overlying or underlying aquifers. The developed ANN relies on rainfall, Soil Water Index, air temperature and Potential Evapotranspiration to simulate the groundwater levels. These variables make it possible, in fairly general terms, to balance the water budget:

$$Q = P - R - PET \pm \Delta S \tag{3}$$

With Q the discharge, P the rainfall, R the runoff, PET the Potential Evapotranspiration and ΔS the variations of the resulting water supply. The balance cannot be closed if changes in the level are also influenced by other inflows.

Discharge is the most commonly used variable for studying karst systems, as it is an integrative physical parameter, that reflects the behavior of the whole aquifer. However, discharge is not simply measurable at Cadarache because the outlet corresponds to the Durance River, which drains the aquifer. Groundwater levels, on the other hand, are more influenced by local phenomena such as the presence of faults, karst cavities, etc.

Graph Neural Networks (GNN) are a type of ANN, for which the data are represented not as matrices but as graphs: neurons are represented as nodes, and their connection can vary to capture dependencies between entities. GNNs thus extract the spatiotemporal patterns from complex systems. A few studies have already applied GNNs to groundwater level simulation [32,33], but none specifically addresses karst aquifers. Because there is a very dense network of piezometers at Cadarache (around 300 piezometers within the catchment), developing a Graph Neural Network, for which the nodes correspond to several piezometers, would allow to take into account the relations between different aquifers, faulted zones, etc.

Finally, the aquifer's response is also influenced by drainage systems and pumping from installations, which are activated when groundwater levels are high. However, information regarding their activation times and discharge rates is not available.

5. Conclusion

The permeability of the fractured karst aquifer of the CEA Cadarache is very heterogeneous, and little is known about the location of its faults and their connectivity. The semi-physical model MECK is currently used to simulate past groundwater levels, but it has difficulty reproducing the variations of the levels caused by complex phenomena such as close rainfall events, or the vertical distribution of zones with varying permeability. While physical models require a large amount of data for accurate groundwater level simulation, ANNs are able to learn non-linear relations between the inputs and outputs and can generalize and predict unknown data effectively.

With the aim of estimating the 100-year flood level for enhancing the safety of nuclear facilities, this study compared the results of an Artificial Neural Network and a semi-physical model, for the simulations of past groundwater levels with a focus on high levels. A Multilayer Perceptron is developed. Its architecture and the selection of its hyperparameters is achieved based on correlation analyses and through a trial-and-error approach. The groundwater levels simulated by the Artificial Neural Network, simulated by the semi-physical model, and measured at Cadarache are then compared. NSE and RMSE over several groundwater level thresholds are used to compare the three timeseries.

Results show that, although the study piezometer accounts for complex phenomena (such as the bottle-neck effect), the developed Multilayer Perceptron manages to reproduce fairly accurately the variations of the groundwater level.

The simulations of the ANN level (overall NSE of 0.87 on test set) are more accurate than those of the semi-physical model level (overall NSE of 0.60 on the same ANN test set). Indeed, the actual version of the semi-physical model tends to overestimate the groundwater levels. Besides, the semi-physical model does not permit precisely simulate complex phenomena such as nearby rain events of several zones with different permeabilities within one piezometer.

Further improvements of the ANN are nevertheless needed for a better simulation of intermediate amplitude loadings. That could possibly be achieved by developing a Graph Neural Network that would integrate the exchanges between aquifers and zones with different hydrogeological properties.

Acknowledgments

The authors wish to thank Manon Erguy, Sarah Gautier, Youness Hrour and Julien Yise Peniel Adounkpe for their helpful discussions. We would also like to thank Dominique Bertin for developing the software used to develop the Artificial Neural Network.

Reference

1. Gèze B. Lexique des termes français de spéléologie physique et de karstologie. Annales de Spéléologie. 1973;28(0001):1–20.
2. Salomon J-N. Le facteur temps dans la karstification/The factor time in karstification. 1999 [cited 2024 Nov 19]; doi: 10.3406/morfo.1999.988.
3. Goldscheider N, Chen Z, Auler AS, et al. Global distribution of carbonate rocks and karst water resources. Hydrogeol J. 2020;28(5):1661–1677.
4. Ford D, Williams PD. Karst Hydrogeology and Geomorphology. John Wiley & Sons; 2007.
5. Doerfliger N, Jeannin P-Y, Zwahlen F. Water vulnerability assessment in karst environments: a new method of defining protection areas using a multi-attribute approach and GIS tools (EPIK method). Environmental Geology. 1999;39(2):165–176.
6. De Waele J, Gutiérrez F, Parise M, et al. Geomorphology and natural hazards in karst areas: A review. Geomorphology. 2011;134(1):1–8.
7. Eisenlohr L, Király L, Bouzelboudjen M, et al. Numerical simulation as a tool for checking the interpretation of karst spring hydrographs. Journal of Hydrology. 1997;193(1):306–315.
8. Jeannin P-Y. Modeling flow in phreatic and epiphreatic Karst conduits in the Hölloch Cave (Muotatal, Switzerland). Water Resources Research. 2001;37(2):191–200.
9. Liedl R, Sauter M, Hückinghaus D, et al. Simulation of the development of karst aquifers using a coupled continuum pipe flow model. Water Resources Research [Internet]. 2003 [cited 2025 Sep 3];39(3).
10. Duran L, Fournier M, Massei N, et al. Assessing the Nonlinearity of Karst Response Function under Variable Boundary Conditions. Groundwater. 2016;54(1):46–54.
11. Kong a Siou, L., Johannet, A., Borrell, V., Pistre, S., 2011. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). Journal of Hydrology 403, 367–380. https://doi.org/10.1016/j.jhydrol.2011.04.015.
12. Erguy M. Apport des réseaux de neurones artificiels dans la prévision du risque de crue karstique sur le site du Commissariat à l'Energie Atomique de Cadarache [Internet] [thesis]. IMT Mines Alès; 2024 [cited 2026 Apr 16]. Available from: https://theses.fr/2024EMAL0009.
13. Najib K. Phénomènes de remontées de nappe extrêmes en terrain carbonaté fracturé et karstifié : évaluation de l'aléa et Prévention du risque dans le cadre de la protection des bâtiments [Internet] [These de doctorat]. Montpellier 2; 2007 [cited 2024 Dec 10]. Available from: https://theses.fr/2007MON20218.
14. Khashei M, Bijari M. An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with Applications. 2010;37(1):479–489.
15. Kong-A-Siou, L., Fleury, P., Johannet, A., Borrell Estupina, V., Pistre, S., Dörfliger, N., 2014. Performance and complementarity of two systemic models (reservoir and neural networks) used to simulate spring discharge and piezometry for a karst aquifer. Journal of Hydrology 519, 3178–3192. https://doi.org/10.1016/j.jhydrol.2014.10.041.
16. Trichakis IC, Nikolos IK, Karatzas GP. Artificial Neural Network (ANN) Based Modeling for Karstic Groundwater Level Simulation. Water Resour Manage. 2011;25(4):1143–1152.
17. Kong-A-Siou, L., Johannet, A., Borrell Estupina, V., Pistre, S. Neural networks for karst groundwater management: case of the Lez spring (Southern France). Environmental Earth Sciences, 2015, 74 (12), pp.7617--7632. 10.1007/s12665-015-4708-9
18. Guerin R. Synthèse des connaissances géologiques et hydrogéologiques. CEA Cadarache; 2001. Report No.: R.T.CAD/ST/GPI PAT/GEOL RAP 012. .
19. Cornes RC, van der Schrier G, van den Besselaar EJM, et al. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. Journal of Geophysical Research: Atmospheres. 2018;123(17):9391–9409.
20. Habets F, Boone A, Champeaux JL, et al. The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France. Journal of Geophysical Research: Atmospheres [Internet]. 2008 [cited 2025 Dec 12];113(D6).

21. Mangin A. Contribution à l'étude hydrodynamique des aquifères karstiques [Internet] [phdthesis]. Université de Dijon; 1975 [cited 2024 Nov 19]. Available from: https://hal.science/tel-01575806.

22. Thornthwait CW. An Approach toward a Rational Classification of Climate. Geographical Review. 1948;Vol. 38(No. 1):55–94.

23. Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signal Systems. 1989;2(4):303–314.

24. Barron AR. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information Theory. 1993;39(3):930–945.

25. Levenberg K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics. 1944;2(2):164–168.

26. Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics. 1963;11(2):431–441.

27. Dietterich T. Overfitting and undercomputing in machine learning. ACM Comput Surv. 1995;27(3):326–327.

28. Stone M. Cross-Validation and Multinomial Prediction. Biometrika. 1974;61(3):509–515.

29. Sjöberg J, Zhang Q, Ljung L, et al. Nonlinear black-box modeling in system identification: a unified overview. Automatica. 1995;31(12):1691–1724.

30. Darras T, Johannet A, Vayssade B, et al. Ensemble model to enhance robustness of flash flood forecasting using an Artificial Neural Network: case-study on the Gardon Basin (south-eastern France). Boletin Geologico y Minero. 2018;129(3):565.

31. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology. 1970;10(3):282–290.

32. Tao B, Pejman T. Graph neural network for groundwater level forecasting. Journal of Hydrology. 2023;616:128792.

33. Taccari ML, Wang H, Nuttall J, et al. Spatial-temporal graph neural networks for groundwater data. Sci Rep. 2024;14(1):24564.