water

MDPI

Article

A Fully Connected Neural Network (FCNN) Model to Simulate Karst Spring Flowrates in the Umbria Region (Central Italy)

Francesco Maria De Filippi [1,*], Matteo Ginesi [2] and Giuseppe Sappa [1]

[1] Department of Civil, Environmental and Construction Engineering (DICEA), Sapienza University of Rome, Via Eudossiana 18, 00185 Rome, Italy; giuseppe.sappa@uniroma1.it
[2] Mediterranean Institute of Fundamental Physics (MIFP), Via Appia Nuova 31, 00047 Marino, Italy; mat.ginesi@gmail.com
* Correspondence: francescomaria.defilippi@uniroma1.it

Abstract: In the last decades, climate change has led to increasingly frequent drought events within the Mediterranean area, creating an urgent need of a more sustainable management of groundwater resources exploited for drinking and agricultural purposes. One of the most challenging issues is to provide reliable simulations and forecasts of karst spring discharges, whose reduced information, as well as the hydrological processes involving their feeding aquifers, is often a big issue for water service managers and researchers. In order to plan a sustainable water resource exploitation that could face future shortages, the groundwater availability should be assessed by continuously monitoring spring discharge during the hydrological year, using collected data to better understand the past behaviour and, possibly, forecast the future one in case of severe droughts. The aim of this paper is to understand the factors that govern different spring discharge patterns according to rainfall inputs and to present a model, based on artificial neural network (ANN) data training and cross-correlation analyses, to evaluate the discharge of some karst spring in the Umbria region (Central Italy). The model used is a fully connected neural network (FCNN) and has been used both for filling gaps in the spring discharge time series and for simulating the response of six springs to rainfall seasonal patterns from a 20-year continuous daily record, collected and provided by the Regional Environmental Protection Agency (ARPA) of the Umbria region.

Keywords: artificial neural network; karst spring; machine learning; karst modelling; groundwater management

Citation: De Filippi, F.M.; Ginesi, M.; Sappa, G. A Fully Connected Neural Network (FCNN) Model to Simulate Karst Spring Flowrates in the Umbria Region (Central Italy). *Water* **2024**, *16*, 2580. https://doi.org/10.3390/w16182580

Academic Editor: Zhi-jun Dai

Received: 21 August 2024
Revised: 7 September 2024
Accepted: 10 September 2024
Published: 12 September 2024

1. Introduction

Karst regions represent about 7–12% of the global continental area and host large aquifers, representing one of the most important freshwater resources for agricultural and drinking purposes all over the world. An estimated 10% of the global drinking water supply comes from these resources [1,2].

The difficulty in the exploitation, management and protection of these regions, due to the high heterogeneity in their hydraulic properties, makes them unique and represents a continuous challenge for researchers, technicians and water managers to find new solutions when they face issues in karst settings [3–6].

In this framework, the variability in the response of spring flowrates to rainfall patterns is certainly one most challenging topics for karst modelers, because of the strong non-linear characteristics in the inflow-outflow process within the aquifers. For this reason, but also taking into account the climate change effects on karst groundwater, long-term exploitation from these resources requires increasingly high performance models able to understand and simulate complex behaviors related to multiple factors [7–11].

In Europe, karstified carbonate rocks constitute about 21.6% of the total land surface, hosting significant groundwater resources that supply freshwater for most of the largest

cities in the Mediterranean area [12–14]. However, especially in this region, climate change issues and overexploitation are increasingly threatening karst aquifers, influencing human freshwater dependency on these vulnerable resources [15–19].

A sustainable groundwater exploitation solution from karst springs would first need quantitative evaluations coming from the water budget and long-term analyses, continuously monitoring karst spring discharge with high-frequency recording to assess the response to daily or even sub-daily rainfall events. However, many case studies and practical experiences have shown that this has not always been carried out, even to this day, for several reasons ranging from low efficiency of institutional water management to practical issues related to karst heterogeneity itself. If it is not possible to perform measurements of karst spring flow and there is a lack of available data, the discharge estimation is the only way to quantitatively assess the behavior of karst springs over time [20]. In the field of hydrological simulation models, physically based, statistical and black box models have been useful tools for experts in recent years, representing very different ways to obtain the same useful results, sometimes improved by being used in combination. Models based on physical laws are strictly related to the nature of the processes occurring and the present equations that can be solved by analytical or numerical procedures. Data-driven, machine learning and artificial neural network approaches, instead, are considered black box models because they do not depend on the physical knowledge of the process, but they reconstruct a purely empirical model related to relationships found between input and output variables. In the last decades, many approaches have been utilized to analyze the relationship between the rainfall time series, mostly over the recharge area, and the spring flows [21–25]. Recession curve analysis is the oldest and most known physically based approach for studying the behavior of a karst spring. It allows the understanding of variations in spring discharge related to recharge events. The analysis of the recession curves focused on the recession coefficient (α), which is qualitatively related to the degree of aquifer karstification and to the size of the karst system too. Maillet [26] used this analytical model for the first time to describe flow recession using a simple exponential equation. Later, new models were proposed starting from this basis and considering the different flow mechanisms related to conduits, fractures and voids [27,28].

Regarding the statistical models, flow duration curve (FDC) analysis is a measure of the range and variability of spring discharge. FDC analysis represents the percentage of time during which the flow rates of a given spring exceed a specified threshold. The discharges were ordered and plotted from the highest to the lowest values, without considering their chronological sequence [29,30]. Still in statistical methods, an interesting proposal came from the DISHMET model (Discharge Hydro-Climatological Model) that was designed to predict monthly spring discharges by only utilizing a statistical approach that focuses on precipitation and climate variability [23].

Other kinds of analyses, coming from the theory of signals, include assessments of the coherence and gain functions. The first measures the degree of linear correlation between two signals at each frequency. It is the normalized version of the cross spectrum, providing a value between 0 and 1, where 1 indicates perfect correlation and 0 indicates no correlation. The gain function, on the other hand, can indicate the degree of the amplification or attenuation of aquifer discharge based on different frequency components of rainfall [31–33].

Studies focusing on correlation analysis include autocorrelation and cross-correlation analysis, as well as auto-spectral and cross-spectral analysis [34–37]. The autocorrelation function (ACF) can evaluate the linear dependency of successive values of a single parameter for a defined time series. Auto-spectral analysis explains the distribution of variance in the signal x(t) as a function of frequency. The first is helpful in identifying the time delay between related time series and understanding the temporal relationship between them, whereas cross-spectral analysis is a frequency-domain method used to examine the relationship between two time series [38].

In the field of modelling rainfall spring discharge or rainfall groundwater level transfer processes, more focus has been recently put on continuous and discrete wavelet analysis,

cross-correlation analysis or machine learning models [39–42]. Other studies, specifically concerning karst springs, have employed time-series analysis studying the transfer function between rainfall and spring discharge, obtained by black box models or artificial neural networks [24,41].

In fact, with the recent application of big data and artificial intelligence (AI) to a variety of disciplines—even daily routine activities—the tables have turned. The number of papers related to AI models and neural networks has increased in multiple fields, including the field of groundwater hydrology and hydrogeology.

An artificial neural network (ANN) is a machine learning model that makes decisions in a way similar to the human brain, using processes like biological neurons which work together to identify phenomena, weigh options and reach conclusions. Each neural network consists of layers of nodes, or artificial neurons: an input layer, one or more hidden layers and an output layer. Each node connects to others and has its own associated weight and threshold. The neuron applies a linear transformation to the input vector through a weight matrix.

ANN models that simulate hydrological processes learn from a significant amount of data to capture complex nonlinear relationships between measured rainfall and runoff values. In these kinds of applications, models are usually multi-layer feed-forward networks. However, obtaining the number of hidden layers and the number of neurons in each hidden layer is not straightforward, and no rules are available to determine the exact numbers. These numbers are usually determined by a trial-and-error procedure [24].

As the variability and heterogeneity of karst aquifers is a sensitive topic for water service managers and researchers, the use of ANN models has been spread out to provide reliable simulations of spring discharges [43–51]. A nonlinear autoregressive exogenous (NARX) neural network model has been applied in the same study area of this work, with satisfactory results and promising developments [41]. However, long short-term memory (LSTM) models were shown to be particularly effective for karst spring discharge forecasting and seem to be the most used method in these kinds of applications [43,51].

LSTM networks are a type of recurrent neural network (RNN) that may learn order dependence in sequence prediction problems. The current RNN step uses the output from the previous step as its input.

Other kinds of ANNs and data-driven methods have successfully modelled spring flows at the daily scale, being able to also solve issues related to karst flood forecasting [49,50]. The use of these kinds of models can provide information not only about future values of karst spring discharges, but also about past behavior in data-scarce karst springs [52].

In this study, the discharge values of six karst springs located in the Umbria region of Central Italy have been assessed and simulated by using a black box model, implemented in the Phyton coding language. The goal of this research is to determine the primary factors that govern studied spring discharge patterns and behavior with an artificial neural network technique able to fill gaps in data series and estimate spring discharge. More specifically, the neural network used in this study is a fully connected neural network (FCNN). It consists of a series of fully connected layers that link each neuron in one layer to each neuron in the other layer. A transformation is then applied to the product through the activation function.

The major advantage of fully connected networks is that there are no special assumptions needed to be made about the input. It is usually considered less suitable for long-term predictions compared to other types of models like RNN or LSTM. However, it can be used for short-term predictions with appropriate feature engineering and data pre-processing, though the results might not be optimal, due to its natural lack of temporal dependencies.

2. Geographical, Geological and Hydrogeological Framework

The study area is in the Umbria region, in the central part of Italy, being the only peninsular region that is not bordered by the sea. However, it is full of groundwater due to the presence of the Apennines, a mountain chain characterized by the presence of

thick limestone sequences. They outcrop in the eastern part of the region, where several springs are present and there is a strong interaction between groundwater and surface water (Figure 1). In fact, it has been assessed that almost 80% of the groundwater resource in the hydrogeological basins of the study area comes out as linear springs, with the majority of streams mostly fed by groundwater due to the interaction with saturated zones [53].

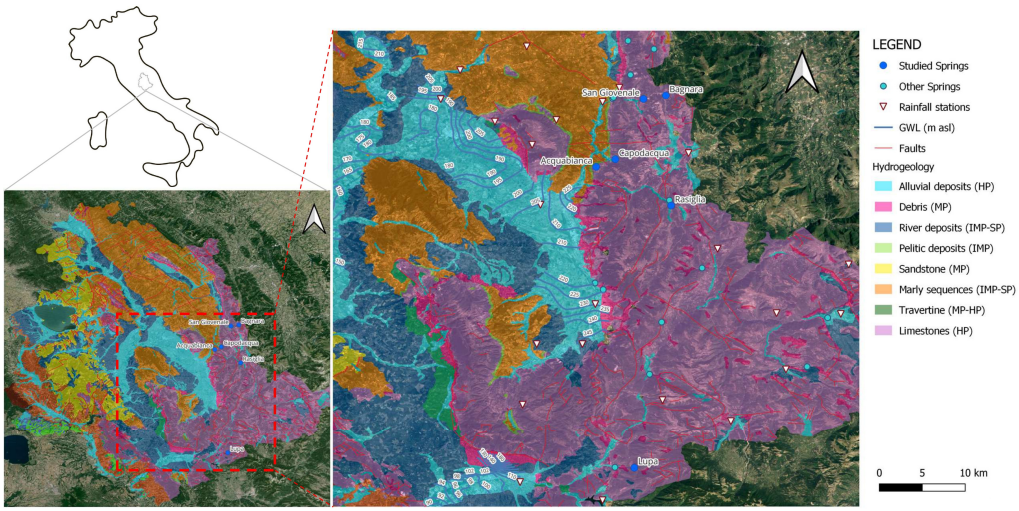


Figure 1. Simplified hydrogeological map (IMP: impermeable, SP: semi-permeable, HP: highly permeable).

The geological formations hosting the aquifers that feed the selected studied springs are mainly composed of calcareous and marly lithotypes (from the Triassic to Cretaceous age). They belong to the so-called Umbria-Marche succession, taking this name from the two regions (Umbria and Marche) and are present in many other Apennine areas of central Italy. From a structural point of view, the Umbria–Marche Apennines are a thin-skinned thrust belt with a hierarchy of multiple, superimposed detachments over a main basal detachment at the level of Triassic evaporites. In terms of hydrogeology, this means that the latter acts as an aquiclude, overlaid by the carbonatic sequence which presents multiple aquifers characterized by different permeabilities and more or less of a presence of fractures, joints and karst conduits according to the karstification degree [54,55]. This sequence, also present in other regions in central Italy such as the Lazio and Abruzzo regions, has been grouped in the same formation, where different limestone complexes are separated by less thick and permeable layers of marl, considered to be almost impermeable (dark purple area in Figure 1).

Different types of karst aquifers feed the most important springs in the area, including the six considered in this study. As the outcropping geology of the reliefs is composed of carbonate hydrostructures, groundwater recharge mainly occurs in the eastern inner areas, thanks to the material having a high capability of infiltration [53]. The recharge area of the springs under study was not taken into account for this specific research because of the "blind" nature of the technique used, i.e., a black box model. However, several previous studies have already obtained the extension of the recharge area of each spring based on lithological, structural and hydrogeochemical studies [53,56–62].

3. Materials and Methods

3.1. Data Source

For setting up the model input and output data, rainfall and spring flowrates, as well as coordinates and elevations of all the monitoring points, were taken from websites related to institutional regional agencies of Umbria. The rainfall data were collected by automated rain gauges and provided by the Hydrographic Service of the Umbria region

at the following website link: "www.regione.umbria.it/ambiente/servizio-idrografico (accessed date on 10 June 2024)".

In this specific study, input data from a twenty-four year period were considered (from 1 January 2000 to 1 January 2024) for a total of more than 350,000 values.

Regarding the output, daily discharge values of six selected karst springs (Rasiglia, San Giovenale, Nocera, Lupa, Acquabianca and Bagnara), monitored by the Regional Protection Agency of the Umbria region (ARPA Umbria), were analyzed and elaborated to build up the model, for a total of more than 43,800 values. These springs were selected because of their long time series of data, freely available at the website "www.arpa.umbria.it (accessed date on 10 June 2024)". Unfortunately, they were not downloadable as a text or .csv file, so it was necessary to manually copy and paste the values to create the dataset for the same time window of the input data, constituting a huge amount of effort in terms of time and data pre-treatment. The name, location in the WGS84 reference system and elevation of both the springs and the rainfall stations are shown in Tables 1 and 2, respectively.

Table 1. Name, coordinates (WGS84) and elevation of the studied springs.

Spring	Latitude	Longitude	Elevation (m asl)
Nocera	43.167	12.849	632
San Giovenale	43.103	12.811	456
Lupa	42.585	12.813	375
AcquaBianca	43.029	12.741	391
Bagnara	43.109	12.855	623
Rasiglia	42.983	12.852	665

Table 2. Name, coordinates (WGS84) and elevation of the rainfall stations.

Rainfall Station	Latitude	Longitude	Elevation (m asl)
Nocera Umbra	43.11889	12.79111	535
Colfiorito	43.02639	12.88917	759
Gualdo Tadino	43.24083	12.78139	595
Armenzano	43.07333	12.70167	716
La Bolsella	43.03817	12.66732	923
Assisi	43.07098	12.61462	424
Casa Castalda	43.17750	12.65972	718
Branca	43.26028	12.68083	350
Torre dell'Olmo	43.31889	12.69500	550
Pianello	43.14389	12.56528	234
Nocera Scalo	43.09889	12.76722	392
Petrignano	43.10278	12.53778	244
Foligno	42.95314	12.67908	224
Spoleto	42.75583	12.73861	357
Azzano	42.81250	12.75694	240
Sellano	42.89083	12.93028	608
Forsivo	42.79972	13.01389	968
Norcia	42.79861	13.10500	700
Forca Canapine	42.76056	13.18889	1654
Sorgenti Pescia	42.67667	13.16444	1179
Castelluccio di Norcia	42.82933	13.21402	1452
Campi Altopiano	42.86861	13.11611	1141
Cascia	42.72004	13.02722	604
Monteleone di Spoleto	42.64667	12.94917	935
San Vito	42.67639	12.85222	1006
Castagnacupa	42.67760	12.65389	778
Piediluco	42.53417	12.76722	370
S. Silvestro	42.75583	12.67389	383
Terni	42.55972	12.65028	130

3.2. Data Structure and Normalization

The dataset is structured as a single list of numerical data divided into the 5 available features at the time, and as the output, the flowrate value for each spring. The dates were divided into day, month and year components to increase the entropy, improve the quality of the dataset and provide more data points for training the model.

In an attempt to capture some spatial relationships, feature engineering and inclusion of spatial variables (latitude, longitude and elevation) were used. FCNNs are not inherently suited for capturing spatial dependencies. Hence, incorporating them directly from the dataset can help the model learn spatial relationships to some extent.

In summary, the dataset consists of multiple time series data, with each series representing a different spring. Each data point includes values for latitude, longitude, elevation, day, month, year and the corresponding karst spring flow. This approach allows the model to capture temporal and spatial patterns more effectively, leading to better performances. To mitigate issues arising from the wide range of feature values, normalization was applied to be sure that each feature contributed equally to the training process. Preventing features with larger numerical ranges from dominating the training process was fundamental, because it could lead to slower convergence, thereby worsening the overall model performance. Normalization was performed using the standard normalization (z-score normalization) formula:

$$x' = (x - \mu)/\sigma$$

where μ is the mean and σ is the standard deviation.

3.3. Data Splitting and Model Architecture

To ensure an unbiased evaluation of the model performance on unseen data, it is common practice to split the dataset into two subsets:

- ✓ "training" used for model training
- ✓ "validation" used as "values unknown by the model" in order to have unbiased performance evaluation.

The training and validation ratio was about 80/20. This choice represents an optimal balance between the need to have a sufficiently large sample for model training and the need to have an adequate validation set to independently evaluate the model performance [63,64]. This is a crucial compromise to reduce the risk of overfitting and to ensure a good generalization of the model on the validation data not considered during training. Among the empirical studies that confirm this, Kohavi (1995) showed that this splitting ratio is effective in providing reliable estimates of the model performance, minimizing the impact of fluctuations in the data; in particular, for moderately sized datasets such as the one under consideration, an 80/20 split allows a sufficiently representative validation set to be maintained without excessively compromising the amount of data available for training [65].

The data's temporal dependence was maintained by choosing the samples for the two subsets sequentially from the complete dataset: temporally we had data prior to a certain date in the training set, and data following a certain date in the validation set.

The sequential FCNN model was trained on all available springs in the training set and tailored to avoid overfitting issues due to the small size of the dataset; in total, the architecture included 10 fully connected layers with the ReLu activation function, interspersed with 6 batch normalization layers. ReLu is a function defined as follows:

$$\begin{aligned} Y(x) &= 0, \text{ if } x \leq 0 \\ Y(x) &= x, \text{ if } x > 0 \end{aligned} \quad (1)$$

The structure ended with a single fully connected layer of neuronal output using a simple linear activation function. In particular, the maximum number of neurons in the hidden

layers reached 1024, allowing the network to capture complex patterns while managing the risk of overfitting (Table 3). The code was implemented in Python version 3.12.3.

Table 3. Detailed structure of the FCNN model code.

Layer	Output Shape	Activation Function
Dense Fully Connected (Input)	64	ReLu
Batch Normalization	64	
Dense Fully Connected	128	ReLu
Dense Fully Connected	256	ReLu
Batch Normalization	256	
Dense Fully Connected	512	ReLu
Batch Normalization	512	
Dense Fully Connected	1024	ReLu
Batch Normalization	1024	
Dense Fully Connected	512	ReLu
Batch Normalization	512	
Dense Fully Connected	128	ReLu
Dense Fully Connected	64	ReLu
Dense Fully Connected	32	ReLu
Batch Normalization	32	
Dense Fully Connected	16	ReLu
Dense Fully Connected (Output)	1	Linear

The structure of the layers was:

1. Input Layer:
   - Purpose: Matches the number of input features (5 in this case)
   - Activation function: ReLu
2. Hidden Layers:
   - Purposes: capturing features, spatial and temporal relations, hidden variables
   - Activation Function: Rectified Linear Unit (ReLu), which allows the network to capture complex, non-linear patterns: ReLu (x) = max(0,x)
   - Batch Normalization: Applied to stabilize the learning process, accelerate training and reduce internal covariate shift
3. Output Layer:
   - Purpose: A single neuron with linear activation, suitable for regression tasks where the output is a continuous value.
   - Activation function: Linear

4. Results and Discussion

The results refer to the cross correlation analyses of the discharge time series of the springs under study and the application of the FCNN model. Regarding the latter, the raw dataset was initially pre-processed by the model, using the discharge values of other springs to infer the missing values of a single spring. This was done to obtain a new dataset in which there were no longer any gaps present. Subsequently, the same model was applied to all of the dataset, this time considering the rainfall values provided by the 30 meteorological stations as input.

4.1. Cross Correlation Analyses

After creating the dataset for the model setting, some simple cross correlation analyses on the six spring discharge time series were performed. The correlation coefficients (ρ_{xy}) between the daily discharge values of the springs have been calculated (Table 4).

Table 4. Cross-correlation coefficients for the discharge time series from the six karst springs (2000–2024).

	Nocera	San Giovenale	Lupa	Acquabianca	Bagnara	Rasiglia
Nocera	-	0.55	0.74	0.57	0.81	0.60
San Giovenale	0.55	-	0.67	0.61	0.60	0.67
Lupa	0.74	0.67	-	0.82	0.74	0.84
AcquaBianca	0.57	0.61	0.82	-	0.72	0.78
Bagnara	0.81	0.60	0.74	0.72	-	0.74
Rasiglia	0.60	0.67	0.84	0.78	0.74	-

The results show that all of the correlation coefficients obtained are higher than 0.5, due to the same seasonal pattern of rainfall, but also due to the karst nature of the aquifers, with a reduced lag time between inputs and outputs. However, some similarities related to the specific karstification degree and the main type of groundwater circulation in the feeding aquifers could be noticed by the higher value of the correlation coefficient obtained for specific springs. That is the case of the following couples: Bagnara–Nocera, Lupa–Rasiglia and Lupa–Acquabianca. In all of these three cases, the correlation coefficient ρ_{xy} is more than 0.8, indicating a strong correlation (almost linear) between the discharges of the two springs. A clear example is the case of the Acquabianca–Nocera couple. Although they are very close (Figure 1), they actually show a rather low correlation coefficient (equal to 0.57), highlighting that in a very small area the two springs could drain from different karst systems, or in correspondence with a specific threshold value, an overflow from one system to another can occur.

4.2. FCNN Model Results

4.2.1. Filling Gaps in the Spring Discharge Time Series

In some cases, the measurement instruments were out of service during the 24 years under study. This leads to a lack of continuity, with serious consequences for the training of the model and potentially generating divergence and accuracy loss in the process. In the historical measured data of all six springs analyzed, a total of 2408 missing or not determined (n.d.) values were present. In particular, they were distributed among the springs as follows: 793 for San Giovenale, 371 for Acquabianca, 151 for Lupa, 208 for Bagnara, 356 for Rasiglia and 608 for Nocera. This required a tool to pre-process the data and obtain a continuous homogeneous dataset to be used for the model. Raw data were pre-processed using the FCNN model, using all the available information and discharge values of the other springs to infer the missing values of a single spring on the selected time interval (Figure 2).

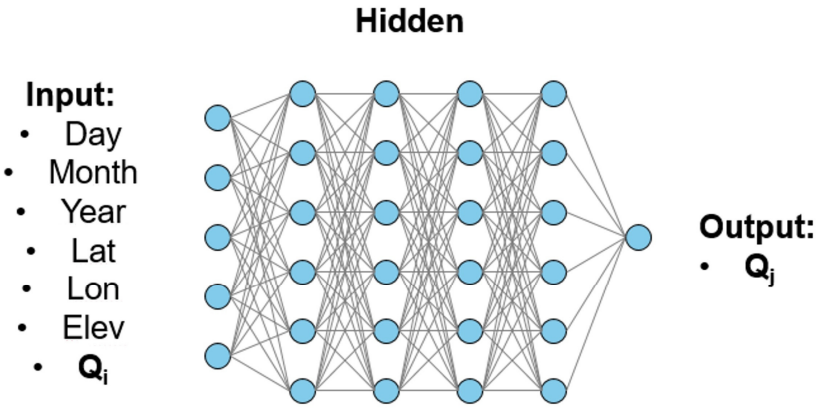
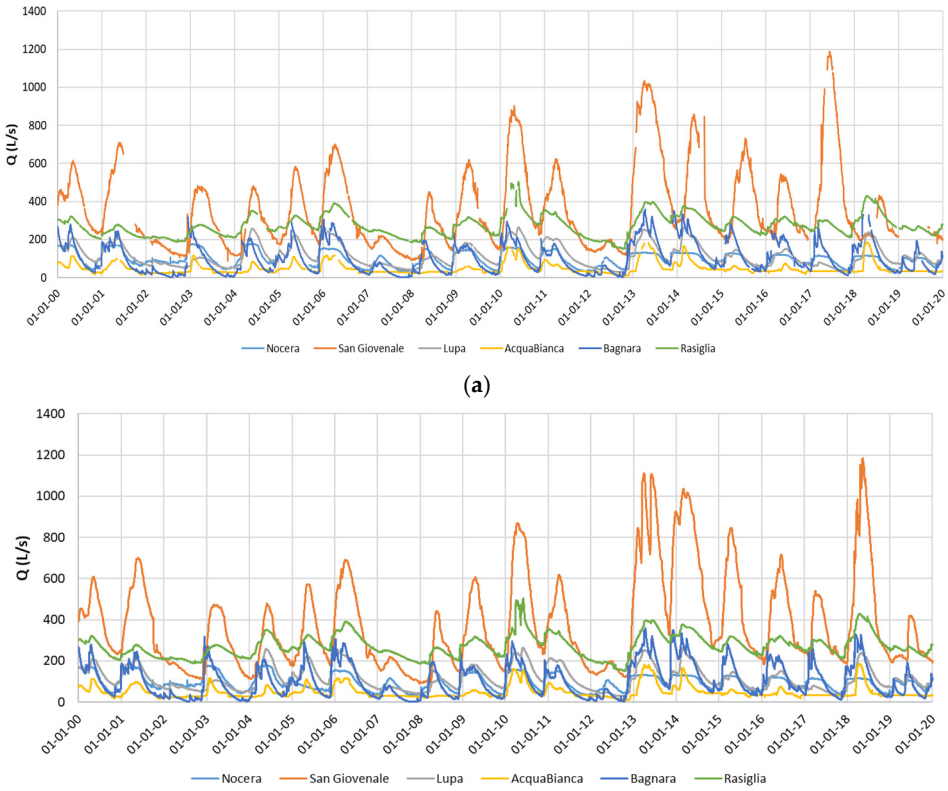


Figure 2. Conceptual structure of FCNN for filling gaps within the discharge time series.

The results are shown in Figure 3, where raw data (Figure 3a) and processed data with gaps filled (Figure 3a) are presented for all six springs considered in this study. A

few spikes and oscillations are visible in specific time intervals, such as the case of San Giovenale in the winter of 2013, but except this the dataset is continuously filled with reasonable discharge values, following the usual behavior of karst springs to both seasonal and high intensity rainfall patterns.



(**a**)

(**b**)

Figure 3. Karst spring discharge time series: (**a**) raw data with gaps and (**b**) post-processed data after filling gaps.

4.2.2. Simulating Karst Spring Flowrate Behavior

For the simulation of the responses of the six karst springs to rainfall inputs, the same model was applied to the new dataset obtained after the gap-filling procedure. This time, in the input vector there was no more information about the discharge, but the rainfall values of the 30 meteorological stations in the study area were considered to directly train the model on the output vector (Figure 4).

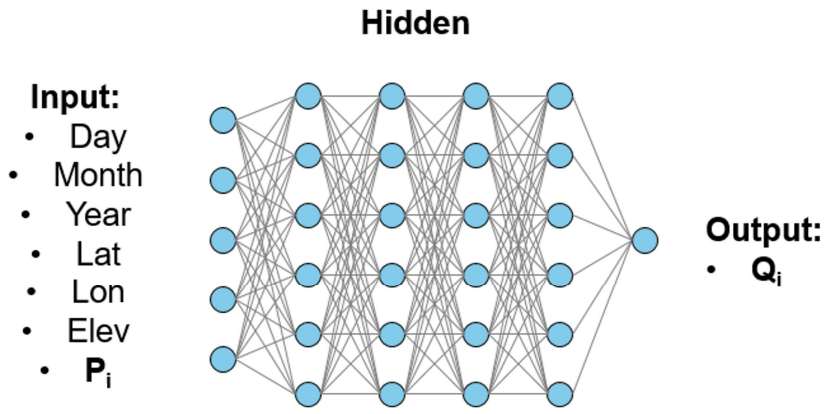


Figure 4. Conceptual structure of FCNN for simulating karst spring discharge behavior.

The results for the entire dataset related to the six springs is presented in Figure 5. The loss function was minimized after 332 epochs, even if an initial rapid training was clearly visible and no divergence occurred. The final scatter plot, indicating the accuracy of results compared to the measured discharge values, confirmed the good performance of the proposed model.

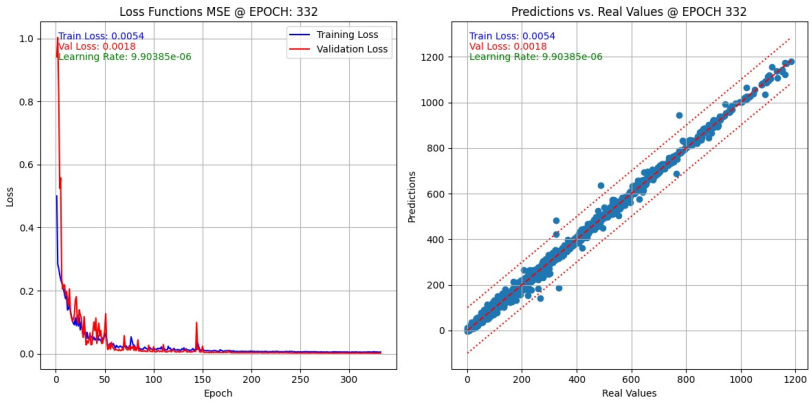


Figure 5. Final plots of loss function and simulated vs. measured discharge values for the entire dataset of the six selected springs (Phyton 3.12.3). Red dotted lines in the scatter plot defines the 90% confidence interval.

The obtained results showing the measured and simulated values of flowrates for each spring are presented in Figure 6 for the Rasiglia (Figure 6a), Nocera (Figure 6b) and San Giovenale (Figure 6c) springs and in Figure 7 for the Lupa (Figure 7a), Bagnara (Figure 7b) and Acquabianca (Figure 7c) springs.

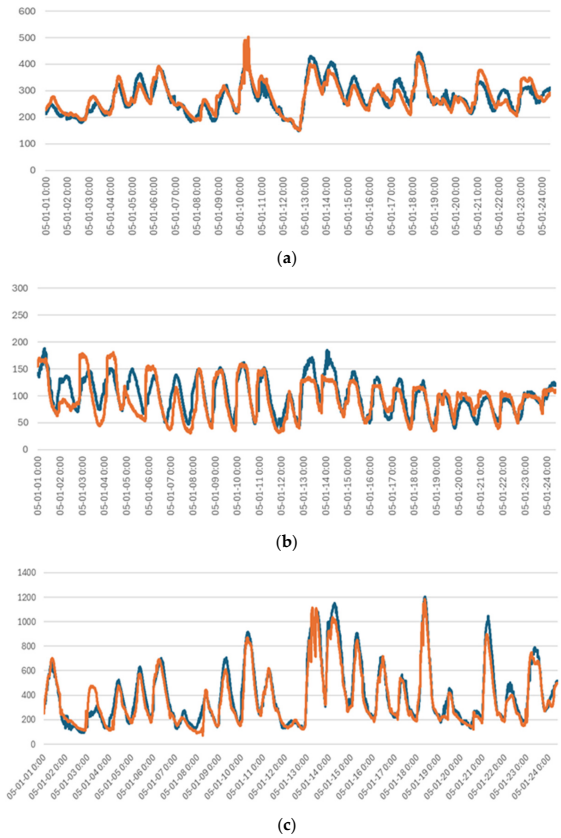


Figure 6. Comparison between measured (orange) and simulated (blue) spring flowrates (in L/s) in the time series 2000–2024: (**a**) Rasiglia; (**b**) Nocera; (**c**) San Giovenale.

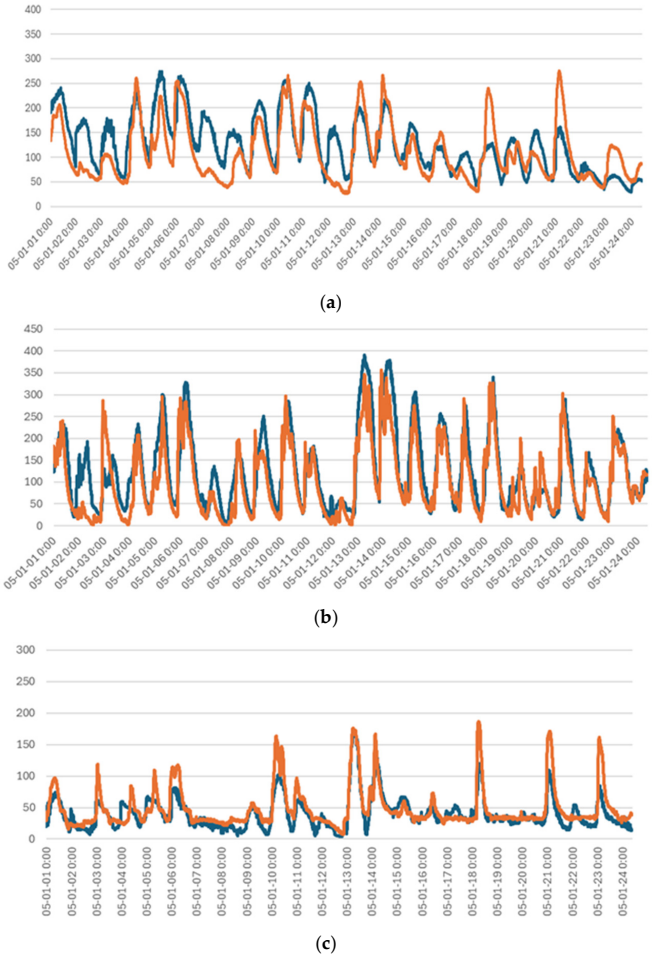


Figure 7. Comparison between measured (orange) and simulated (blue) spring flowrates (in L/s) in the time series 2000–2024: (**a**) Lupa; (**b**) Bagnara; (**c**) Acquabianca.

The model's performance on simulating spring discharge was assessed by calculating the R-squared coefficient (R^2) for the simulated time series according to the measured time series (Table 5). The best results were found for the Rasiglia and San Giovenale springs, which were also the springs with higher values of average annual flowrates, higher storage capacity in the feeding aquifer and shorter lag time in the spring response to seasonal rainfall inputs. On the contrary, in the case of Lupa the model showed the worst performance, with some of the maximum discharge during the driest years not being matched by the model simulation. However, most of the minimum values (the ones of interest for groundwater management during shortages) are well simulated and, even in this case, the overall results showed a value of R^2 equal to 0.89, which is still high. Regarding the mean absolute error (MAE) and root mean squared error (RMSE), the highest obtained values were those related to San Giovenale, whereas the lowest were for Nocera. However, these results refer to springs with large differences in the average discharge; for example, San Giovenale produces several hundreds of liters per second, while Acquabianca only produces a few tens.

This case study highlights that the smallest and more karstified springs are the most difficult to be modelled (Bagnara, Lupa and Acquabianca), but very good results could be obtained with those characterized by higher average discharge (San Giovenale) and lower variability (Rasiglia). Anyway, the performance results confirm the suitability of the FCNN model to provide useful insights into the six karst springs present in the study area of Umbria, in Central Italy. Focusing on their dependence on each other in terms of time and space (coordinates and altitude) and using daily recorded rainfall values of 30 stations spread all over the area, the model has been able to successfully simulate the response of

each spring, obtaining results potentially useful for the groundwater management of the local water agencies.

Table 5. Evaluation metrics for spring discharge simulations obtained by FCNN model.

Spring	R^2 (-)	MAE (L/s)	RMSE (L/s)
Nocera	0.95	17.5	23.1
San Giovenale	0.98	49.8	68.6
Lupa	0.89	46.1	56.5
AcquaBianca	0.91	19.7	25.1
Bagnara	0.92	35.6	45.4
Rasiglia	0.99	19.8	24.8

5. Conclusions

The proposed study aimed to understand and evaluate the behavior of six karst springs located in the Umbria region (central Italy) using cross correlation analyses and a fully connected neural network (FCNN) implemented in the Phyton coding language. Cross correlation analyses between the different karst spring discharge time series came out to be helpful in recognizing similarities related to the hydraulic properties of the feeding aquifers, focusing on the groundwater circulation type.

The FCNN model used dates as input (day, month and year), as well as the coordinates and elevation data of the springs and meteorological stations, considering rainfall and training the model on spring discharge values as output. The proportions of the dataset was 80% for training and 20% for validation, randomly chosen in the entire time series.

Results showed that the FCNN model was effective in simulating karst spring flowrates with the amount of data available in this study and can be used for short-term predictions with appropriate feature engineering and data pre-processing. However, predictions beyond the historical data of the dataset more than a few days push the model to divergence. The feasibility study of the model, with the small quantitative data available, is useful for preliminary analysis and the understanding the phenomenon. The model has given positive feedback and results both as a tool for filling gaps within the data set, and by providing insights into the computational and modelling characteristics of the phenomenon. Hopefully the proposed model will drive the analyses to more complex forecasting models such as RNN and LSTMs that are the future steps of this project, with the aim of improving results and adaptability.

Neural networks and artificial intelligence models could be useful and important tools for water management in karst regions, but one must not forget that no matter how good the model is, it will be always a black box: it does not care about the physics. For this reason, some "hidden" relationships that the ANN might find could be related to the phenomena, even if one is not able to see them with ordinary physical and analytical approaches. Furthermore, many other relationships might not be found if the right amount and nature of input is not considered in the ANN model.

Author Contributions: Conceptualization, F.M.D.F.; methodology, F.M.D.F. and M.G.; software, M.G.; validation, F.M.D.F.; formal analysis, G.S.; data curation, F.M.D.F. and M.G.; writing—original draft preparation, F.M.D.F. and M.G.; writing—review and editing, F.M.D.F. and G.S.; visualization, F.M.D.F.; supervision, F.M.D.F. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The input rainfall data and measured spring discharges data are available, respectively, from the following resources available in the public domain: www.regione.umbria.it/ambiente/servizio-idrografico (accessed date on 10 June 2024); www.arpa.umbria.it (accessed date on 10 June 2024). The raw data of simulated spring discharges will be made available by the authors on request.

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

References

1. Hartmann, A.; Goldscheider, N.; Wagener, T.; Lange, J.; Weiler, M. Karst water resources in a changing world. *Rev. Geophys.* **2013**, *2013*, 218–242. [CrossRef]
2. Stevanović, Z. Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.* **2019**, *78*, 662. [CrossRef]
3. Hartmann, A. The karst and the furious—Ways to keep calm when dealing with karst hydrology. In Proceedings of the EGU General Assembly 2021, Online, 19–30 April 2021. EGU21-1353. [CrossRef]
4. Fan, X.; Goeppert, N.; Goldscheider, N. Quantifying the historic and future response of karst spring discharge to climate variability and change at a snow-influenced temperate catchment in central Europe. *Hydrogeol. J.* **2023**, *31*, 2213–2229. [CrossRef]
5. Fiorillo, F.; Malik, P. Hydraulic behavior of karst aquifers. *Water* **2019**, *11*, 1563. [CrossRef]
6. De Filippi, F.M.; Iacurto, S.; Grelle, G.; Sappa, G. Magnesium as Environmental Tracer for Karst Spring Baseflow/Overflow Assessment—A Case Study of the Pertuso Karst Spring (Latium Region, Italy). *Water* **2021**, *13*, 93. [CrossRef]
7. Guo, Y.; Qin, D.; Li, L.; Sun, J.; Li, F.; Huang, J. A complicated karst spring system: Identified by karst springs using water level, hydrogeochemical, and isotopic data in Jinan, China. *Water* **2019**, *11*, 947. [CrossRef]
8. Zeng, S.; Liu, Z.; Goldscheider, N.; Frank, S.; Goeppert, N.; Kaufmann, G. Comparisons on the effects of temperature, runoff, and land-cover on carbonate weathering in different karst catchments: Insights into the future global carbon cycle. *Hydrogeol. J.* **2020**, *29*, 331–345. [CrossRef]
9. Rudolph, M.G.; Collenteur, R.A.; Kavousi, A.; Giese, M.; Wöhling, T.; Birk, S.; Hartmann, A.; Reimann, T. A data-driven approach for modelling Karst spring discharge using transfer function noise models. *Environ. Earth Sci.* **2023**, *82*, 339. [CrossRef]
10. Citrini, A.; Camera, C.; Beretta, G. Pietro Nossana Spring (Northern Italy) under Climate Change: Projections of future discharge rates and water availability. *Water* **2020**, *12*, 387. [CrossRef]
11. Hao, Y.; Yeh, T.C.J.; Wang, Y.; Zhao, Y. Analysis of karst aquifer spring flows with a gray system decomposition model. *Ground Water* **2007**, *45*, 46–52. [CrossRef]
12. Bakalowicz, M. Karst and karst groundwater resources in the Mediterranean. *Environ. Earth Sci* **2015**, *74*, 5–14. [CrossRef]
13. Xanke, J.; Goldscheider, N.; Bakalowicz, M.; Barbera, J.A.; Broda, S.; Chen, Z.; Ghanmi, M.; Gunther, A.; Hartmann, A.; Jourde, H.; et al. Mediterranean Karst Aquifer Map (MEDKAM), 1:5,000,000. Berlin, Karlsruhe, Paris. Available online: https://doi.org/10.25928/MEDKAM.1 (accessed on 10 June 2024).
14. Bakalowicz, M. Coastal Karst Groundwater in the Mediterranean: A Resource to Be Preferably Exploited Onshore, Not from Karst Submarine Springs. *Geosciences* **2018**, *8*, 258. [CrossRef]
15. Nerantzaki, S.D.; Nikolaidis, N.P. The response of three Mediterranean karst springs to drought and the impact of climate change. *J. Hydrol.* **2020**, *591*, 125296. [CrossRef]
16. Sappa, G.; De Filippi, F.M.; Ferranti, F.; Iacurto, S. Environmental Issues and Anthropic Pressures in Coastal Aquifers: A Case Study in Southern Latium Region. *Acque Sotter. Ital. J. Groundw.* **2019**, *8*. [CrossRef]
17. Sivelle, V.; Jourde, H.; Bittner, D.; Mazzilli, N.; Tramblay, Y. Assessment of the relative impacts of climate changes and anthropogenic forcing on spring discharge of a Mediterranean karst system. *J. Hydrol.* **2021**, *598*, 126396. [CrossRef]
18. Jodar, J.; Herms, I.; Lamban, L.J.; Martos-Rosillo, S.; Herrera-Lameli, C.; Urrutia, J.; Soler, A.; Custodio, E. Isotopic content in high mountain karst aquifers as a proxy for climate change impact in Mediterranean zones: The Port del Comte karst aquifer (SE Pyrenees, Catalonia, Spain). *Sci. Total Environ.* **2021**, *790*, 148036. [CrossRef]
19. D'Oria, M.; Balacco, G.; Todaro, V.; Alfio, M.R.; Tanda, M.G. Assessing the impact of climate change on a coastal karst aquifer in a semi-arid area. *Groundw. Sustain. Dev.* **2024**, *25*, 101131. [CrossRef]
20. Sappa, G.; De Filippi, F.M.; Iacurto, S.; Grelle, G. Evaluation of Minimum Karst Spring Discharge Using a Simple Rainfall-Input Model: The Case Study of Capodacqua di Spigno Spring (Central Italy). *Water* **2019**, *11*, 807. [CrossRef]
21. Romano, E.; Del Bon, A.; Petrangeli, E.; Preziosi, E. Generating synthetic time series of springs discharge in relation to standardized precipitation indices. Case study in Central Italy. *J. Hydrol.* **2013**, *507*, 86–99. [CrossRef]
22. Fiorillo, F. The Recession of Spring Hydrographs, Focused on Karst Aquifers. *Water Resour. Manag.* **2014**, *28*, 1781–1805. [CrossRef]
23. Diodato, N.; Guerriero, L.; Fiorillo, F.; Esposito, L.; Revellino, P.; Grelle, G.; Guadagno, F.M. Predicting Monthly Spring Discharges Using a Simple Statistical Model. *Water Resour. Manag.* **2014**, *28*, 969–978. [CrossRef]
24. Xi, C.; Cai, C.; Quingquing, H.; Zhicai, Z.; Peng, S. Simulation of rainfall-underground outflow responses of a karstic watershed in Southwest China with an artificial neural network. *Water Sci. Eng.* **2008**, *1*, 1–9.
25. Chang, W.; Chen, X. Monthly Rainfall-Runoff Modeling at Watershed Scale: A Comparative Study of Data-Driven and Theory-Driven Approches. *Water* **2018**, *10*, 1116. [CrossRef]
26. Maillet, E. *Essais D'hydraulique Souterraine et Fluviale*; Hermann, A., Ed.; Librairie Sci.: Paris, France, 1905; 218p.
27. Bonacci, O. Karst Springs Hydrographs as Indicators of Karst Aquifers. *Hydrol. Sci. J.* **1993**, *38*, 51–62. [CrossRef]
28. Ford, D.; Williams, P. *Karst Hydrogeology and Geomorphology*; John Wiley: Chichester, UK, 2007; p. 562. [CrossRef]
29. Torresan, F.; Fabbri, P.; Piccinini, L.; Dalla Libera, N.; Pola, M.; Zampieri, D. Defining the hydrogeological behavior of karst springs through an integrated analysis: A case study in the Berici Mountains area (Vicenza, NE Italy). *Hydrogeol. J.* **2020**, *28*, 1229–1247. [CrossRef]

30. Medici, G.; Lorenzi, V.; Sbarbati, C.; Manetta, M.; Petitta, M. Structural Classification, Discharge Statistics, and Recession Analysis from the Springs of the Gran Sasso (Italy) Carbonate Aquifer; Comparison with Selected Analogues Worldwide. *Sustainability* **2023**, *15*, 10125. [CrossRef]
31. Padilla, A.; Pulido-Bosch, A. Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis. *J. Hydrol.* **1995**, *168*, 73–79. [CrossRef]
32. Petalas, C.P.; Akratos, C.S.; Tsihrintzis, V.A. Hydrogeological Investigation of a Karst Aquifer System. *Environ. Process.* **2018**, *5*, 155–181. [CrossRef]
33. Duran, L.; Massei, N.; Lecoq, N.; Fournier, M.; Labat, D. Analyzing multi-scale hydrodynamic processes in karst with a coupled conceptual modeling and signal decomposition approach. *J. Hydrol.* **2020**, *583*, 124625. [CrossRef]
34. Fiorillo, F.; Doglioni, A. The relation between karst spring discharge and rainfall by cross-correlation analysis (Campania, southern Italy). *Hydrogeol. J.* **2010**, *18*, 1881–1895. [CrossRef]
35. Lo Russo, S.; Suozzi, E.; Gizzi, M.; Taddia, G. SOURCE: A semi-automatic tool for spring-monitoring data analysis and aquifer characterisation. *Environ. Earth Sci.* **2021**, *80*, 710. [CrossRef]
36. Guo, Y.; Wang, F.; Qin, D.; Zhao, Z.; Gan, F.; Yan, B.; Bai, J.; Muhammed, H. Hydrodynamic characteristics of a typical karst spring system based on time series analysis in northern China. *China Geol.* **2021**, *4*, 433–445. [CrossRef]
37. Denić-Jukić, V.; Lozić, A.; Jukić, D. An Application of Correlation and Spectral Analysis in Hydrological Study of Neighboring Karst Springs. *Water* **2020**, *12*, 3570. [CrossRef]
38. Pavlić, K.; Parlov, J. Cross-Correlation and Cross-Spectral Analysis of the Hydrographs in the Northern Part of the Dinaric Karst of Croatia. *Geosciences* **2019**, *9*, 86. [CrossRef]
39. Granata, F.; Saroli, M.; De Marinis, G.; Gargano, R. Machine Learning Models for Spring Discharge Forecasting. *Geofluids* **2018**, *2018*, 8328167. [CrossRef]
40. Giustolisi, O.; Doglioni, A.; Savic, D.A.; di Pierro, F. An evolutionary multiobjective strategy for the effective management of groundwater resources. *Water Resour. Res.* **2008**, *44*, W01403. [CrossRef]
41. Di Nunno, F.; Granata, F.; Gargano, R.; de Marinis, G. Prediction of Spring Flows Using Nonlinear Autoregressive Exogenous (NARX) Neural Network Models. *Environ. Monit. Assess.* **2021**, *193*, 350. [CrossRef]
42. Sezen, C.; Bezak, N.; Bai, Y.; Šraj, M. Hydrological Modelling of Karst Catchment Using Lumped Conceptual and Data Mining Models. *J. Hydrol.* **2019**, *576*, 98–110. [CrossRef]
43. An, L.; Hao, Y.; Yeh, T.C.J.; Liu, Y.; Liu, W.; Zhang, B. Simulation of Karst Spring Discharge Using a Combination of Time–Frequency Analysis Methods and Long Short-Term Memory Neural Networks. *J. Hydrol.* **2020**, *589*, 125320. [CrossRef]
44. Song, X.; Hao, H.; Liu, W.; Wang, Q.; An, L.; Jim Yeh, T.C.; Hao, Y. Spatial-Temporal Behavior of Precipitation Driven Karst Spring Discharge in a Mountain Terrain. *J. Hydrol.* **2022**, *612*, 128116. [CrossRef]
45. Zhou, R.; Zhang, Y. Linear and Nonlinear Ensemble Deep Learning Models for Karst Spring Discharge Forecasting. *J. Hydrol.* **2023**, *627*, 130394. [CrossRef]
46. Wunsch, A.; Liesch, T.; Cinkus, G.; Ravbar, N.; Chen, Z.; Mazzilli, N.; Jourde, H.; Goldscheider, N. Karst Spring Discharge Modeling Based on Deep Learning Using Spatially Distributed Input Data. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 2405–2430. [CrossRef]
47. Cinkus, G.; Wunsch, A.; Mazzilli, N.; Liesch, T.; Chen, Z.; Ravbar, N.; Doummar, J.; Fernández-Ortega, J.; Barberá, J.A.; Andreo, B.; et al. Comparison of Artificial Neural Networks and Reservoir Models for Simulating Karst Spring Discharge on Five Test Sites in the Alpine and Mediterranean Regions. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 1961–1985. [CrossRef]
48. Paleologos, E.K.; Skitzi, I.; Katsifarakis, K.; Darivianakis, N. Neural Network Simulation of Spring Flow in Karst Environments. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 1829–1837. [CrossRef]
49. Siou, L.K.A.; Johannet, A.; Borrell, V.; Pistre, S. Complexity Selection of a Neural Network Model for Karst Flood Forecasting: The Case of the Lez Basin (Southern France). *J. Hydrol.* **2011**, *403*, 367–380. [CrossRef]
50. Rahbar, A.; Mirarabi, A.; Nakhaei, M.; Talkhabi, M.; Jamali, M. A Comparative Analysis of Data-Driven Models (SVR, ANFIS, and ANNs) for Daily Karst Spring Discharge Prediction. *Water Resour. Manag.* **2022**, *36*, 589–609. [CrossRef]
51. Pölz, A.; Blaschke, A.P.; Komma, J.; Farnleitner, A.H.; Derx, J. Transformer Versus LSTM: A Comparison of Deep Learning Models for Karst Spring Discharge Forecasting. *Water Resour. Res.* **2024**, *60*, e2022WR032602. [CrossRef]
52. Wen, C.; Li, J.; Sun, D.; Zhang, Y.; Zhao, N.; Hu, L. Reconstruction of Past Water Levels in Data-Deficient Karst Springs. *Water* **2024**, *16*, 1150. [CrossRef]
53. Tamburini, A.; Menichetti, M. Groundwater Circulation in Fractured and Karstic Aquifers of the Umbria-Marche Apennine. *Water* **2020**, *12*, 1039. [CrossRef]
54. Centamore, E.; Deiana, G. La Geologia delle Marche. *Studi Geol. Camerti* **1986**, *Special Volume*, 1–145.
55. Barchi, M.; De Feyter, A.; Magnani, M.B.; Minelli, G.; Pialli, G.; Sotera, B.M. The structural style of the Umbria-Marche fold and thrust belt. *Soc. Geol. Ital. Mem.* **1998**, *52*, 557–578.
56. Capaccioni, B.; Didero, M.; Paletta, C.; Salvadori, P. Hydrogeochemistry of groundwaters from carbonate formations with basal gypsiferous layers: An example from Mt Catria-Mt Nerone ridge (Northern Apennines, Italy). *J. Hydrol.* **2001**, *253*, 14–26. [CrossRef]
57. Nanni, T.; Vivalda, P. The aquifers of the Umbria-Marche Adriatic region: Relationship between structural setting and groundwater chemistry. *Boll. Soc. Geol. Ital.* **2005**, *124*, 523–542.

58. Mastrorillo, L.; Baldoni, T.; Banzato, F.; Boscherini, A.; Cascone, D.; Checcucci, R.; Petitta, M.; Boni, C. Quantitative hydrogeological analysis of the carbonate domain in the Umbria region. *Ital. Eng. Geol. Environ.* **2009**, *1*, 137–155.
59. Di Matteo, L.; Dragoni, W.; Valigi, D. Update on Knowledge of Water Resources of Amelia Mountains (central Italy). *Ital. J. Eng. Geol. Environ.* **2009**, *1*, 83–96.
60. Mastrorillo, L.; Petitta, M. Hydrogeological conceptual model of the Upper River basin aquifers (Umbria-Marche Apennines). *Ital. J. Geosci.* **2014**, *133*, 396–408. [CrossRef]
61. Di Matteo, L.; Capoccioni, A.; Porreca, M.; Pauselli, C. Groundwater-Surface Water Interaction in the Nera River Basin (Central Italy): New Insights after the 2016 Seismic Sequence. *Hydrology* **2021**, *8*, 97. [CrossRef]
62. Preziosi, E.; Guyennon, N.; Petrangeli, A.B.; Romano, E.; Di Salvo, C. A Stepwise Modelling Approach to Identifying Structural Features That Control Groundwater Flow in a Folded Carbonate Aquifer System. *Water* **2022**, *14*, 2475. [CrossRef]
63. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
64. Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
65. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Appear. Int. Jt. Conf. Articial Intell.* **1995**, *14*, 1137–1143.