water

IMPACT FACTOR 3.0

CITESCORE 6.0

Article

Modeling Water Table Response in Apulia (Southern Italy) with Global and Local LSTM-Based Groundwater Forecasting

Lorenzo Di Taranto, Antonio Fiorentino, Angelo Doglioni and Vincenzo Simeone

MDPI

https://doi.org/10.3390/w17152268

water

MDPI

Article

Modeling Water Table Response in Apulia (Southern Italy) with Global and Local LSTM-Based Groundwater Forecasting

Lorenzo Di Taranto *, Antonio Fiorentino, Angelo Doglioni and Vincenzo Simeone

Politecnico di Bari, Department of Civil, Environmental, Construction Engineering and Chemistry (DICATECh), via Edoardo Orabona 4, 70125 Bari, Italy; a.fiorentino4@phd.poliba.it (A.F.); angelo.doglioni@poliba.it (A.D.); vincenzo.simeone@poliba.it (V.S.)

* Correspondence: l.ditaranto@phd.poliba.it

Abstract

For effective groundwater resource management, it is essential to model the dynamic behaviour of aquifers in response to rainfall. Here, a methodological approach using a recurrent neural network, specifically a Long Short-Term Memory (LSTM) network, is used to model groundwater levels of the shallow porous aquifer in Southern Italy. This aquifer is recharged by local rainfall, which exhibits minimal variation across the catchment in terms of volume and temporal distribution. To gain a deeper understanding of the complex interactions between precipitation and groundwater levels within the aquifer, we used water level data from six wells. Although these wells were not directly correlated in terms of individual measurements, they were geographically located within the same shallow aquifer and exhibited a similar hydrogeological response. The trained model uses two variables, rainfall and groundwater levels, which are usually easily available. This approach allowed the model, during the training phase, to capture the general relationships and common dynamics present across the different time series of wells. This methodology was employed despite the geographical distinctions between the wells within the aquifer and the variable duration of their observed time series (ranging from 27 to 45 years). The results obtained were significant: the global model, trained with the simultaneous integration of data from all six wells, not only led to superior performance metrics but also highlighted its remarkable generalization capability in representing the hydrogeological system.

Keywords: groundwater; porous aquifer; karst aquifer; rainfall and groundwater; data-driven modeling; multi LSTM model; regional modeling

check for updates

Academic Editor: Alessandra Feo

Received: 20 June 2025
Revised: 26 July 2025
Accepted: 28 July 2025
Published: 30 July 2025

Citation: Di Taranto, L.; Fiorentino, A.; Doglioni, A.; Simeone, V. Modeling Water Table Response in Apulia (Southern Italy) with Global and Local LSTM-Based Groundwater Forecasting. *Water* **2025**, *17*, 2268. https://doi.org/10.3390/w17152268

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Modeling hydrogeological systems is a complex task due to their intrinsic natural complexity and the highly nonlinear processes that govern their behavior and response to external inputs. This complexity makes the control and management of these systems challenging, even when robust monitoring data are available. To effectively model their response as a function of precipitation, an initial conceptualization defining the system is developed, followed by the calibration of model parameters based on direct observations [1]. For these reasons, and for the increasing number of technologies and data available, during the last years there has been a growing interest in the use of data-driven systems for the study of Ground Water Level (GWL) prediction, since the evolution of Machine Learning (ML) and Deep Learning (DL) models has provided efficient and robust tools to analyze complex and nonlinear data; however, without a sufficient increase in available data, these

techniques may not yield satisfactory results. These techniques can be a valid alternative for overcoming the limitations of conventional numerical methods, which require detailed physically based models of aquifers, which are complex to be defined, calibrated and validated. The increasing availability of global and regional datasets of remote sensing data such as GRACE, InSAR and GNSS, and products of open models such as DEM and climate data, has also enabled a deeper understanding of GWL fluctuations even with the inclusion of more complex features than rainfall, significantly improving the accuracy of forecasts in complex climate contexts [2].

Evolutionary Polynomial Regression (EPR-MOGA) is a data-driven modeling technique that has found wide and successful application in the study of natural systems. Its success is due to the method's intrinsic ability to provide explicit polynomial equations that describe the underlying physical phenomenon [1,2]. The application of EPR has also been frequently utilized for modeling aquifer systems, allowing for a deeper understanding of the intrinsic phenomenological dynamics of the hydrogeological system [1,3]. Within the field of neural networks, Long Short-Term Memory (LSTM) [4] networks belong to the family of Recurrent Neural Networks (RNNs). They possess a notable ability to capture long-term dependencies, making them especially suitable for groundwater level forecasting.

In recent years, Long Short-Term Memory (LSTM) networks have increasingly been applied in hydrological analyses, particularly for the prediction and reconstruction of groundwater levels. Another study [5] proposed and evaluated a Long Short-Term Memory (LSTM) model to predict daily fluctuations for a lowland river. Zhong et al. [6] evaluated the use of Long Short-Term Memory (LSTM) neural networks to reconstruct groundwater levels, fill gaps, and extend existing time series. A further study [7] compared the performance of different artificial neural network architectures, specifically Nonlinear Autoregressive Networks with Exogenous Inputs (NARX), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs), in forecasting groundwater levels, utilizing widely available meteorological data such as precipitation, temperature, and relative humidity, proving the effectiveness of these networks. Furthermore, most previous studies did not demonstrate the ability of a local LSTM model to learn effectively as the number of wells included in the training increases.

The main aim of this work is the development and evaluation of the performance of a "global" LSTM model that leverages data from multiple wells to improve forecasting accuracy, inside a methodological framework. The use of this specific case study does not intend to produce methodology tailored to this aquifer. The proposed approach aims at proposing and discussing the potentialities that can be used for other monitored aquifers where at least precipitation and groundwater levels are available. Moreover, the rationale for this approach is that although the wells are not directly correlated in terms of individual measurements, they exhibit a similar hydrogeological response. They are all geographically located within the same relatively large shallow porous aquifer. The shallow Brindisi aquifer is a large porous system in Salento (Southern Italy). Its hydraulic conductivity, related to its porosity, combined with its exclusive recharge by rainfall, makes it an ideal location for investigating the correlation between precipitation and groundwater level fluctuations, as already shown by past research [3,8]. For these reasons, precipitation is used as the only forcing variable to the model. We introduce a novel comparative analysis between this global model and traditional single-well LSTM models. Furthermore, we conduct a sensitivity analysis to systematically assess how the number of wells included in the training impacts the model's predictive performance, addressing a gap in past research.

2. Study Area

West of Brindisi lies a large, irregularly shaped flat area known as the "Messapian Threshold" (Figure 1). This area is a small tectonic depression situated between the Murge and the Salento peninsula. The underlying Mesozoic carbonate rock has subsided, allowing for the deposition of Plio-Pleistocene sedimentary layers.

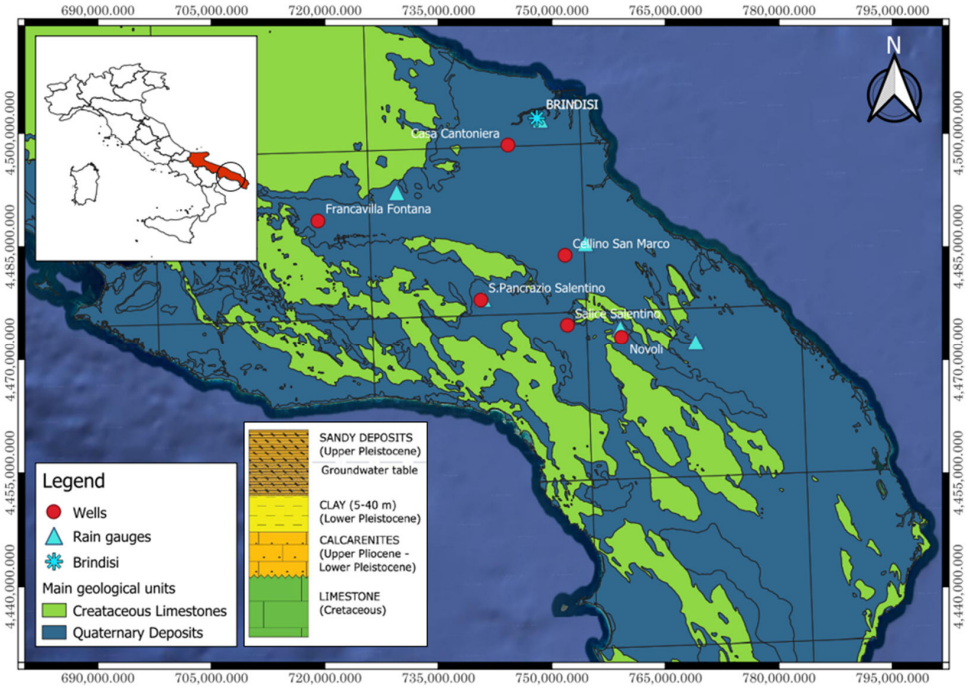

Figure 1. Salento peninsula main geology, location of the sampling wells, rain gauges, and stratigraphic.

The geological sequence has clayey layers resembling sub-Apennine blue-grey clays at its bottom. These clays support more permeable Quaternary deposits, which consist of various materials like sands, calcarenites, conglomerates, terraced levels, and both alluvial and colluvial deposits. These surface layers host a shallow aquifer, which has been thoroughly described in other studies [9]. This aquifer exhibits primary porosity, allowing water to flow through the spaces within the rock and sediment. Its permeability ranges from 8×10^{-6} m/s to 1.4×10^{-4} m/s, and its groundwater circulation relies solely on direct recharge from rainfall. The region is characterized by a well-developed but shallow hydrographic network with only a few deeper channels.

Since this aquifer is permeable due to its porosity and is recharged exclusively by rainfall, it is an excellent location for studying the relationship between rainfall and water table fluctuations. The Brindisi shallow aquifer has a flat hydrogeological catchment area of approximately 1000 km^2 and does not receive water from other catchment areas.

The analysis of the aquifer's water table oscillations is based on monthly data collected by the former National Hydrographic Bureau over a 50-year span, specifically from January 1953 to December 1996. Rainfall data, presented as total monthly precipitation, are available for the same timeframe as the water table levels. Rain-gauge stations are situated relatively close to both monitoring wells, supporting the assumption that the aquifer's primary recharge is local.

3. Materials and Methods

3.1. Data Description

Figure 2 shows the average monthly total rainfall and the corresponding water table levels of the Brindisi shallow aquifer. The main recharge of these aquifers primarily

occurs during the first three months of the year. Although autumn months show higher precipitation than winter months, their contribution to aquifer recharge is not significant. Indeed, autumn precipitation restores water stored by shallow soil layers, lost during the summer dry season. Instead, during autumn, the infiltrating water is first retained by the shallow, unsaturated soil layers to restore their field capacity (their ability to hold water). Water table fluctuations are lagged in relation to precipitation, meaning there's a delay. The peak water table levels typically occur in March and April, with the lowest levels observed during the summer. This summer minimum happens because infiltration is almost absent, and evapotranspiration (water loss from the soil and plants) is very intense [3].

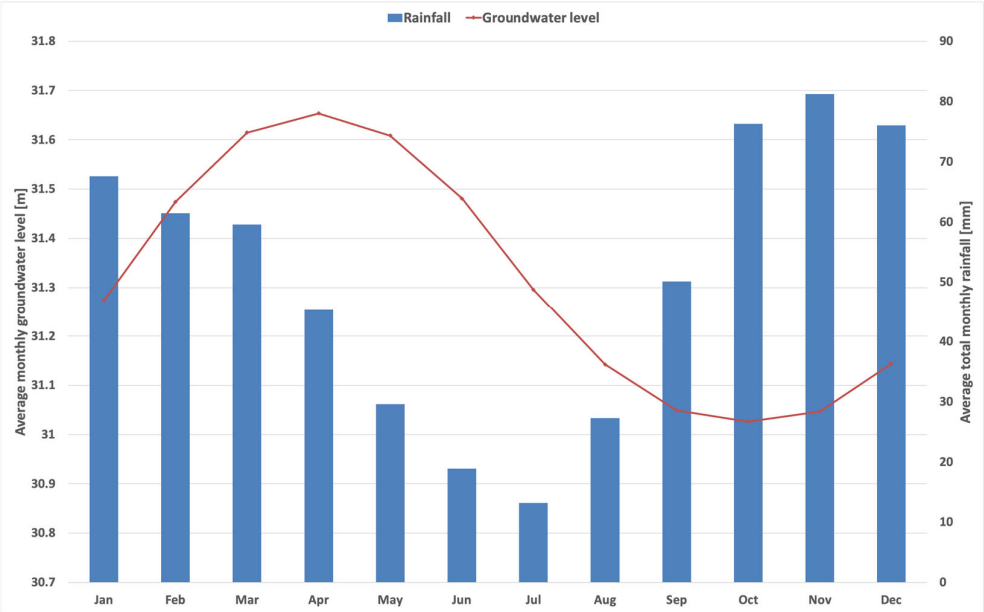

Figure 2. Average monthly water table levels of Brindisi shallow aquifer and average precipitation of Brindisi rain-gauge station, estimate for the time.

Recharge becomes more evident in groundwater level fluctuations when transitioning from autumn to spring, provided there are no sharp or anomalous weather events, such as periods of atypical precipitation. Even in such cases, clear water table fluctuations are evident, linked to long-term climatic variations. A notable example of this is the dry periods between 1987 and 1995.

Six historical series, from six distinct wells, are available for analysis (Table 1 and Figure 3). Each dataset is characterized by a unique temporal length of historical observations. The maximum data length extends to 45 years, while the minimum duration is 27 years. Table 1 shows the main characteristics of these six wells. An aggregated representation of all historical time series is provided in Figure 3.

Table 1. Details of the Investigated Wells.

Well Name	Date	Elevation (m a.m.s.l.)	Depth (m)
S. Pacrazio Salentino	1951–1996 (45 years)	58	2
Casa Cantoniera	1953–1996 (43 years)	35	4
Salice Salentino	1951–1992 (41 years)	46	1
Cellino San Marco	1951–1987 (36 years)	57	6
Novoli	1962–1996 (34 years)	35	11
Francavilla Fontana	1955–1982 (27 years)	136	6

Figure 3. Historical Groundwater Level and Monthly Precipitation Time Series for the Six Study Wells. Each panel (**a**–**f**) shows the long-term variations in groundwater level (red line, left *y*-axis, measured in meters above sea level (m a.s.l.)) and corresponding monthly precipitation (light blue bars, right *y*-axis, measured in millimeters [mm]). The toponyms of wells are the following: (**a**) S. Pancrazio Salentino, (**b**) Casa Cantoniera, (**c**) Salice Salentino, (**d**) Cellino San Marco, (**e**) Novoli, and (**f**) Francavilla Fontana.

3.2. Data Processing

All data from the six different wells were normalized using Min-Max Scaling, a widely employed data preprocessing technique in machine learning. This method rescales numerical features into a predefined range, typically [0, 1]. This linear process is performed by subtracting the minimum value of the feature and dividing by the range (the difference between the maximum and minimum), as shown in Equation (1).

Normalization was used to eliminate scale differences between variables. Such differences can negatively affect the performance of magnitude-sensitive algorithms, such as neural networks. For each well, the parameters X_{min} and X_{max} were calculated exclusively on the training dataset and then reused to transform the validation and test sets. This approach prevents data leakage and ensures the model does not gain information about the future data distribution [9].

$$X_{norm} = \frac{X - X_{min}}{X_{\max} - X_{\min}} \quad (1)$$

3.3. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are an advanced type of recurrent neural network (RNN) specifically designed to overcome an intrinsic limitation of traditional RNNs: their difficulty in retaining information over long sequences of data within their internal structure [4]. This challenge is known as the "vanishing gradient problem," which refers to the tendency, during learning, to lose information acquired from previously read data series [10]. To address this challenge, LSTMs are designed with a different internal structure compared to RNNs. Figure 4 illustrates the internal configuration of a single Long Short-Term Memory (LSTM) computational unit, also referred to as an LSTM cell. The operation of each LSTM unit is governed by the interaction of three primary inputs at a given time step t [4]:

- The current input vector (Xt): This represents the set of observed data at time t. It is organized as a vector that contains all previously normalized input features, such as groundwater level H and precipitation value P.
- The previous hidden state vector (Ht − 1): This represents an output from the LSTM cell at the previous time step. It serves as a compressed representation of the unit's short-term memory, incorporating information derived from past time steps.
- The previous cell state vector (Ct − 1): This component represents the unit's long-term memory.

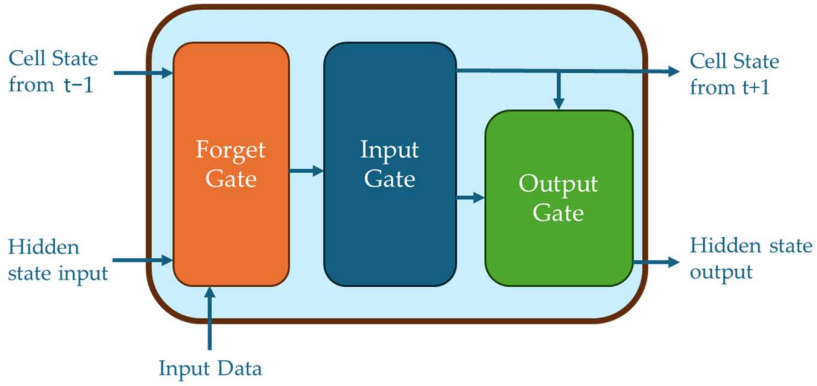

Figure 4. Schematic representation of an LSTM cell.

The control over which information is added to or removed from the memory cell is managed by a specific part called Gates. These are neural components that decide which information to allow through and what to block. There are three different types of gates in an LSTM cell [4]:

- Forget Gate: determines which information from the previous memory LSTM cell should be discarded or "forgotten."
- Input Gate: determines which new information, from the current input vector, should be included in the current state of the memory cell.
- Output Gate: determines which part of the memory cell will be produced as the output for the current time step.

This makes LSTMs exceptionally powerful tools in contexts where sequence and temporal context are crucial. For this reason, they are effective in analyzing time series, such as GWL or meteorological forecasts, where they capture patterns and dependencies that extend over prolonged periods. The pattern and number of gates and memory states allow LSTM cells to control the flow of information during the training phase, allowing them to store and retrieve relevant data over very long time intervals. While Long Short-Term Memory (LSTM) models, like most Machine Learning models, are often perceived as

"black boxes" due to their nonlinear architecture and numerous parameters, the ability to vary the number of forecasting horizons and the input window size allows for a functional probing during evaluation. This enables us to understand how the model effectively grasps short-term and long-term relationships within the data [11].

3.4. Embedding Dimensions

When modeling multiple time series, such as piezometric levels from different wells, it is crucial to provide the model with a way to distinguish and characterize each data source. Since the well identifier (e.g., "Cellino San Marco", "Novoli") is a categorical variable, it cannot be used directly as a numerical input in a deep learning model.

To overcome this limitation, our model incorporates an Embedding layer. This layer transforms each well identifier, represented by an integer, into a vector within an *n*-dimensional vector space, known as the embedding space. This means that each well in the model is uniquely represented by a vector of *n* real numbers. These vectors are not predefined; instead, their components are trainable parameters of the model. Figure 5 illustrates an example representation of an embedding dimension.

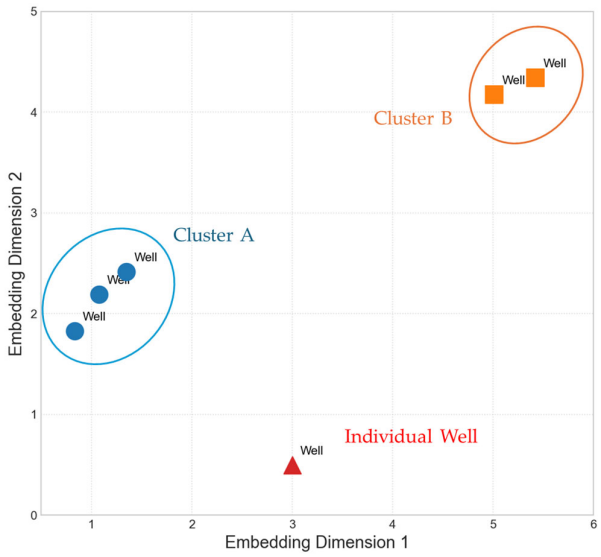

Figure 5. Example of scatter plot representation of well identifiers in a 2D embedding space. Cluster A (orange circle) wells are grouped due to their learned similarities, while Cluster B wells form another distinct similar group but are distant from Cluster A (blue circle). An Individual Well (red triangle).

During the training process, the model learns how to position the well embedding vectors in the n-dimensional space. Their placement captures the latent characteristics relevant for predicting the piezometric level (H). The validity of this learned representation can be visually inspected after training, providing useful insights. Wells that the model groups closely together in the latent space are those identified as having similar numerical behavior, offering further understanding of the model's ability to comprehend the relationships that may exist among the various wells [12]. The strength of this approach lies in its ability, during the training phase, to incorporate a growing number of wells by adding a new embedding dimension vector to their historical time series, without altering the core model structure. This allows for amending results, without using complex physically based models, requiring for non-easily available data.

3.5. Loss Function

For the training of the neural network, a probabilistic loss function was selected. Instead of employing traditional metrics such as Mean Squared Error (MSE), which produces

a single expected value, our model does not directly predict the future piezometric level H. Rather, it predicts the parameters of a Gaussian (Normal) probability distribution from which the true value is assumed to be sampled.

This approach enables the model to quantify its uncertainty for each prediction. The implemented cost function is the Negative Log-Likelihood (NLL) for a Gaussian distribution. The Probability Density Function (PDF) (Equation (2)) of a single observation y for a Gaussian distribution with mean μ and standard deviation σ is given by the following:

$$P(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \quad (2)$$

The Negative Log-Likelihood (NLL), when constants that do not affect minimization are ignored, simplifies to the following:

$$L(y|\mu,\sigma) == \frac{(y-\mu)^2}{2\sigma^2} + \frac{1}{2}\log(\sigma^2) \quad (3)$$

Equation (3) represents the formula that our implemented cost function aims to calculate and minimize. This approach allows the model to not only provide a point estimate of the future groundwater level H (μ) but also a measure of confidence in that estimate (σ) [13].

3.6. Modelling Approach

The model architecture, implemented in Python 3.13.5 using the Keras and TensorFlow libraries, is designed to simultaneously process two distinct types of input: time-series data, consisting of groundwater levels and precipitation values, and a fixed-dimension embedding vector representing each individual well. The weights of the embedding vector for each well are not fixed; instead, they are learned during the training phase. This allows the model to autonomously show potential inter-well relationships. This embedding vector is then replicated and concatenated to the main input sequence at each time step.

The experimental framework was designed to systematically evaluate all possible combinations of wells, including both models trained on single wells and a global model trained on data from all wells. This includes models for all pairs, triplets, quartets, and so on, culminating in a single "global model" trained on the data from all six wells. The framework model hyperparameters are shown in Table 2, and the experimental setup can be outlined in the following scenarios:

- Single-Well Models Scenario: A distinct model is created and trained exclusively on the data from a single well (e.g., Well A). This model becomes highly specialized for Well A but has no information about the others. This process is repeated for all six wells, resulting in six independent models.
- Combinatorial Scenario: A new model instance is trained using data from a combination of wells (e.g., Well A and Well B). In this case, the LSTM and its embedding layer must learn to manage and distinguish between the two wells and infer the relationships between them. This procedure is iterated for all possible pairs (A, C), (A, D), etc., and subsequently for all possible triplets, quartets, and so on.
- Global Model Scenario: A single LSTM model is trained on data from all six wells simultaneously, with each well identified by its unique embedding vector.

In the proposed model, the embedding dimension was set to 4. This choice is supported by the findings of Gu et al. [14], who observed that for more than four inputs (specifically, for six inputs), the model tends to achieve optimal performance, with marginal or no improvements for larger dimensions. For the case studied here, with a limited number

of wells, an embedding dimension of 4 offers an optimal balance between representational capacity and model complexity.

Table 2. Summary of framework hyperparameters and configuration settings.

Parameter	Value	Description
SEED	32	Seed for random number generators
WINDOW_SIZE	5	Number of past time steps (e.g., months) used as input
HORIZONS	(1, 2, 3)	Future prediction horizons (in time steps) for the model's output during the open-loop phase
LSTM_UNITS	46	Number of units in the LSTM layer
LSTM_LAYERS	1	Number of LSTM layers in the core model
EMBEDDING_DIM	4	Size of the embedding vector for each well
EPOCHS	150	Maximum number of training epochs for each model
TEST_START_DATE	1 January 1978	Start date for the test set for all wells
VALIDATION_DURATION_YEARS	4	Duration of the validation set in years
ROLLING_FORECAST_HORIZONS	(1, 3, 7, 12)	List of horizons (L) to test in 'Rolling Multi-Step Forecast' mode during the closed-loop simulation

The SEED parameter was set to 32 to ensure experimental reproducibility. The WINDOW_SIZE, representing the number of past time steps (e.g., months) used as input for the model, was set to 5. This value was chosen based on the aquifer's hydraulic response to precipitation, as discussed in previous sections.

The number of units in an LSTM layer (LSTM_UNITS) was set to 46, and the number of LSTM layers (LSTM_LAYERS) to 1. This specific result was achieved through an iterative, trial-and-error optimization approach, aiming for a favorable balance between model accuracy and computational efficiency.

Three distinct forecasting methodologies were employed to evaluate the model's performance. The first, open-loop prediction, consists of "one-shot" predictions where the model uses actual historical data as input for each time step. This represents the standard evaluation scenario and measures the model's single-step accuracy. A second approach, closed-loop simulation, has the model generate predictions in a fully autoregressive manner over the entire validation and test periods. In this mode, it uses its own predicted groundwater level outputs as input for subsequent steps, while still being fed the true precipitation values, enabling a complete, long-term simulation of groundwater dynamics. The final methodology is the rolling multi-step forecast. For this, at each time step in the test set, the model initiates a forecast over a horizon of length L by starting with the true historical data and then advancing autoregressively. It uses its own predicted groundwater levels as input while retaining the true precipitation values. After completing the forecast, the origin is advanced by one step, and the process is repeated using the newly available ground truth data, a procedure performed across the entire test series for various forecast horizons L.

The complete dataset is chronologically partitioned into three distinct subsets, with the same time splits applied to all six wells: a training set extending up to 1973, a validation set from 1974 to 1977, and a test set from 1978 onwards, which is reserved exclusively for the final, unbiased evaluation of the model's performance. Each input sample consists of a fixed-length sequence (a look-back window of 5 steps) of past groundwater level and precipitation observations. The target is represented by future piezometric levels at specific forecast horizons, which in the open-loop phase are set to 1, 2, and 3 months ahead.

For each of the scenarios, the model outputs the two parameters of a Gaussian distribution: the mean (μ) and the logarithm of the standard deviation ($\log\sigma$). During the training

phase, the Gaussian Negative Log-Likelihood (NLL) is employed as the loss function. Finally, the resulting model is evaluated on the test set.

3.7. Avoidance of Overfitting in Predictive Models

During neural network training, particularly when a high number of epochs is set, a prevalent phenomenon observed is "learning speed slow-down." This shows as the model's tendency to exhibit no significant increase in training set accuracy after reaching a certain threshold. Concurrently, if training continues, a reversal of accuracy on the validation set may be observed, indicating overfitting to the training data. The model essentially begins to lose its generalization capability [15].

For this reason, the proposed model employs an early stopping technique. This methodology allows the model to monitor the change in validation set accuracy during the training phase. Training is halted upon identifying the optimal epoch, which is the point where both underfitting (insufficient training, typical of epochs preceding the validation accuracy reversal point) and overfitting (excessive training occurring after this point) are avoided [15,16].

Upon completion of training, the model's weights were automatically reverted to the configuration that achieved the best validation accuracy. This ensures the adoption of the most robust and performant model on unseen data. This technique allowed for the use of a fixed number of epochs across all analyses, enabling the model to determine the appropriate maximum epoch threshold [15].

3.8. Model Performance Evaluation

This research uses the following metrics to evaluate the efficiency of various models in terms of accuracy: Mean Square Error (MSE), Coefficient of Determination (R^2) and Mean Absolute Error (MAE).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(H_i - \overline{H}_i\right)^2 \tag{4}$$

$$\mathrm{R}^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(H_i - \overline{H}_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(H_i - \overline{H}\right)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|H_i - \overline{H}_i\right|^2 \tag{6}$$

The first metric is the Mean Squared Error (MSE), as shown in Equation (4), which calculates the average of the squared differences between the predicted values (H_i) and the actual values ($\overline{H}_i$), over a total of n observations. The second is the Mean Absolute Error (MAE), described in Equation (6), which calculates the average of the same differences, but in absolute value. These two error metrics are always greater than or equal to zero; a value of 0 indicates a perfect match, and they do not have an upper bound.

The third metric is the coefficient of determination (R^2), defined in Equation (5), where $\overline{H}$ is the mean of the actual values. Unlike MSE and MAE, R^2 has a range of values from 0 to 1, and higher values signal a better predictive capability of the model [9].

4. Results

The model yields results for 63 distinct combinations. Six of these represent individual models, each trained on the data of a single well, while one is the global model, trained on all six wells simultaneously using a dimensional embedding. Given the substantial volume of data generated, detailed time series results are graphically presented solely for the most sensitive well in the system, namely 'Casa Cantoniera'. The scatter plots and performance metrics illustrate the comparison between the single-well model configurations and the global model, which was trained with data from all six wells.

4.1. Open Loop

The results from the open-loop phase indicate a clear enhancement in performance as the number of wells included in the model increases. The analysis of "Casa Cantoniera" well shows how the global model achieves greater predictive accuracy for future groundwater levels (Figures 6–8). Second, it achieves a significant reduction in prediction uncertainty, as quantified by the standard deviation (σ) (Figures 6–8, Table 3). This suggests that during training, the model not only learns the underlying long-term relationships more effectively but also becomes more confident in its predictions, substantially reducing their associated uncertainty (Figures 6 and 7). This trend culminates in the global model, which processes data from all six wells simultaneously and demonstrates improvement in two key domains (Figure 9, Table 4).

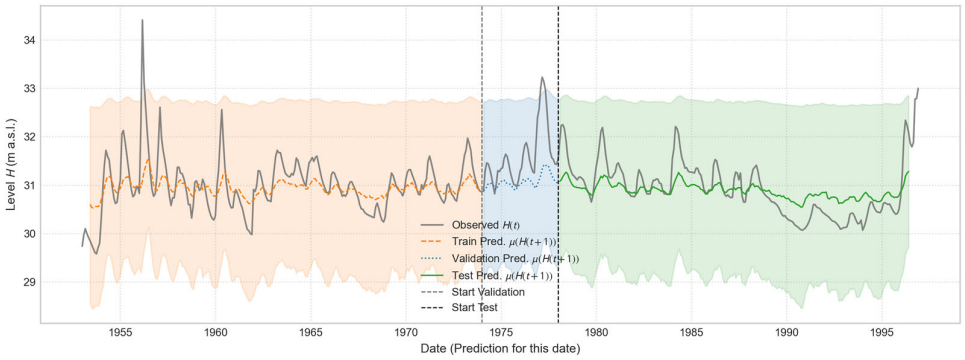

Figure 6. One month ahead Open-Loop prediction for "Casa Cantoniera" well, generated by the individually trained model.

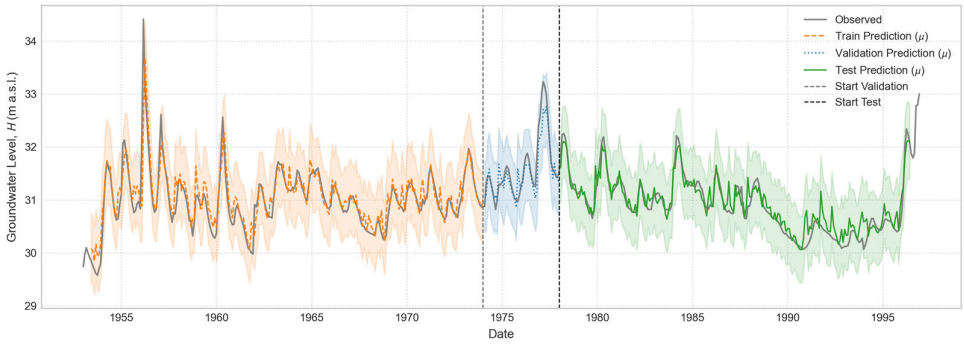

Figure 7. One month ahead Open-Loop prediction for "Casa Cantoniera" well, generated by the globally trained model.

Table 3. Model σ mean for the entire test series. The table presents the forecasted standard deviation (σ) values one, two, and three months ahead, comparing Open-Loop simulation results for individual wells ("1 Well alone") against models where the target well is supported by five additional wells ("1 Well supported by 5 Wells").

Well	σ 1 Month	σ 2 Months	σ 3 Months	σ 1 Months	σ 2 Months	σ 3 Months
		1 Well Alone		1 Well Supported by 5 Wells		
Casa Cantoniera	0.921731633	1.121058262	0.829408957	0.337532867	0.488939191	0.583101378
Cellino San Marco	2.861131741	3.147446002	3.115651124	0.685021842	1.022347641	1.240546083
Francavilla Fontana	1.165261519	1.194734437	0.816706758	0.191408214	0.271964254	0.347186325
Novoli	0.757368482	1.071530992	1.243149763	0.552066195	0.703314831	0.887721345
S. Pancrazio	0.224041011	0.348641935	0.422292953	0.291580448	0.398767864	0.506903108
Salice Salentino	0.229141452	0.364612261	0.430799924	0.259029697	0.36862028	0.487317109

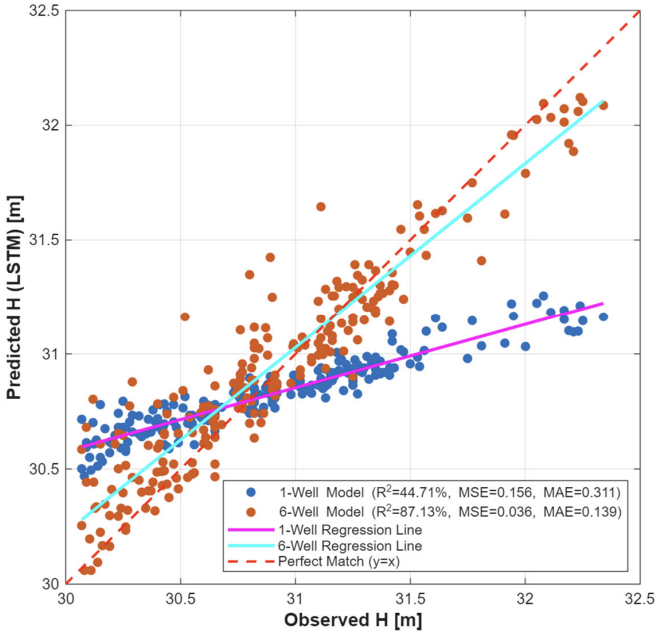

Figure 8. Scatter plot: In blue, the one month ahead Open-Loop prediction for "Casa Cantoniera" well, generated by the individually trained model; in red, one month ahead Open-Loop prediction for "Casa Cantoniera" well, generated by the globally trained model.

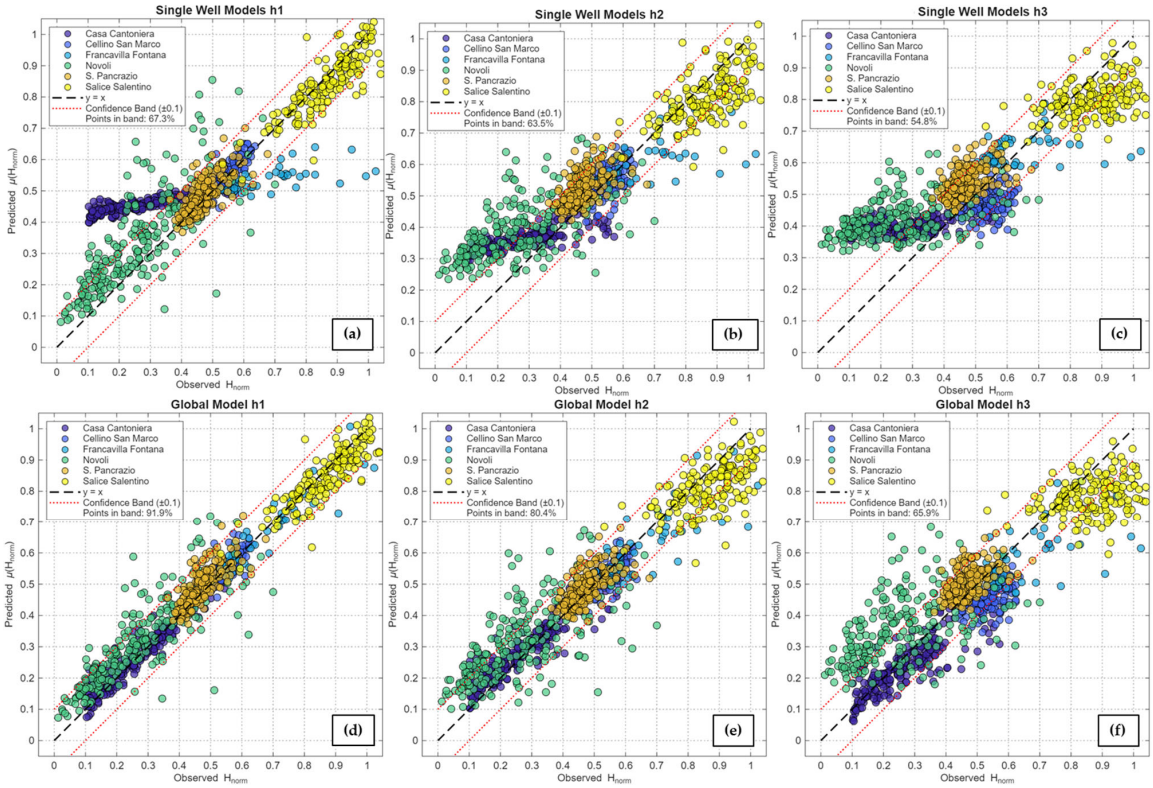

Figure 9. Scatter plots of the Open-Loop predictions at different forecast horizons (h1, h2, h3). The plots compare the observed normalized groundwater level values (*x*-axis) with the predicted values (*y*-axis). The top row (**a**–**c**) illustrates the performance of the models trained individually on each well ('Single Well Models'), while the bottom row (**d**–**f**) refers to the global model ('Global Model') trained on all wells simultaneously. The columns represent forecast horizons of one (h1), two (h2), and three (h3) months. The black dashed line indicates perfect agreement (y = x), and the red lines define a confidence band of ±0.1.

Table 4. Model performance evaluation, in Open-Loop simulation one month ahead, for individual wells ("1 Well alone") versus models supported by five additional wells ("1 Well supported by 5 Wells").

Well	MSE	MAE	R2	MSE	MAE	R2
	1 Well Alone			1 Well Supported by 5 Wells		
Casa Cantoniera	0.042076	0.187538	−2.47802	0.000986	0.023351	0.918526
Cellino San Marco	0.0008	0.022094	0.608493	0.000701	0.018652	0.657276
Francavilla Fontana	0.041174	0.152356	−0.60462	0.003043	0.038853	0.881425
Novoli	0.016188	0.100466	0.209543	0.009164	0.075448	0.552507
S. Pancrazio	0.001652	0.030518	0.344462	0.001866	0.031845	0.259462
Salice Salentino	0.003361	0.04218	0.637009	0.003012	0.041365	0.674698

4.2. Rolling Multi-Step Forecast

To assess the model's predictive robustness under simulation conditions, an in-depth analysis was conducted using the rolling multi-step forecast methodology. This approach enabled the simulation of groundwater level evolution over various forecast horizons (L), offering a more rigorous assessment of the model's capacity to predict future sequences.

The results from this analysis confirm that increasing the number of wells in the training set leads to an overall enhancement of the model's performance (Figure 10, Table 5). However, a decline in the model's predictive accuracy was observed for "Casa Cantoniera" well, specifically during the 1990–1995 period (Figures 11 and 12). This performance degradation is likely attributable to the intense drought that characterized this period. This condition is a known significant challenge for modeling this hydrogeological system, a finding consistent with previous studies using the EPR-MOGA approach [2,3].

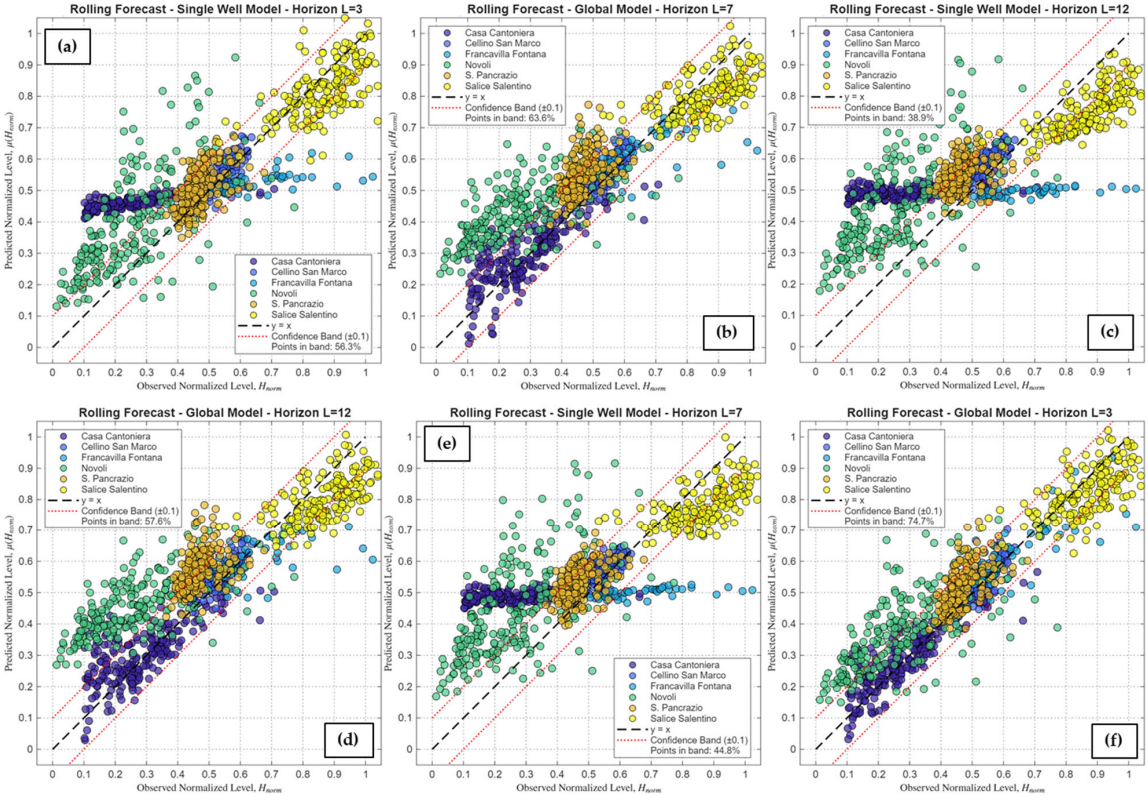

Figure 10. Scatter plots of the Rolling Forecast predictions at different forecast horizons. Plots (**a**–**c**) illustrate the performance of the models trained individually on each well ('Single Well Model') for forecast horizons of L = 3, L = 7, and L = 12. Plots (**d**–**f**) refer to the global model ('Global Model'), trained on all wells simultaneously, for the same forecast horizons. The black dashed line indicates perfect agreement (y = x), while the red lines define a confidence band of ±0.1.

Table 5. Model performance evaluation, in forecast horizons L = 12 simulation, for individual wells ("1 Well alone") versus models supported by five additional wells ("1 Well supported by 5 Wells").

Well	MSE_L12	MAE_L12	R2_L12	MSE_L12	MAE_L12	R2_L12
	Model: 1 Well Alone			Model: 1 Well Supported by 5 Wells		
Casa Cantoniera	0.057432	0.217541	−2.98651	0.007049	0.062549	0.510699
Cellino San Marco	0.002909	0.049205	−0.35837	0.001124	0.025329	0.475057
Francavilla Fontana	0.054971	0.181728	−1.25017	0.021883	0.090293	0.104233
Novoli	0.051709	0.199839	−1.39868	0.049527	0.203103	−1.29746
S. Pancrazio	0.009598	0.088073	−2.80072	0.014196	0.105775	−4.62163
Salice Salentino	0.017778	0.120302	−0.8018	0.010472	0.089603	−0.06136

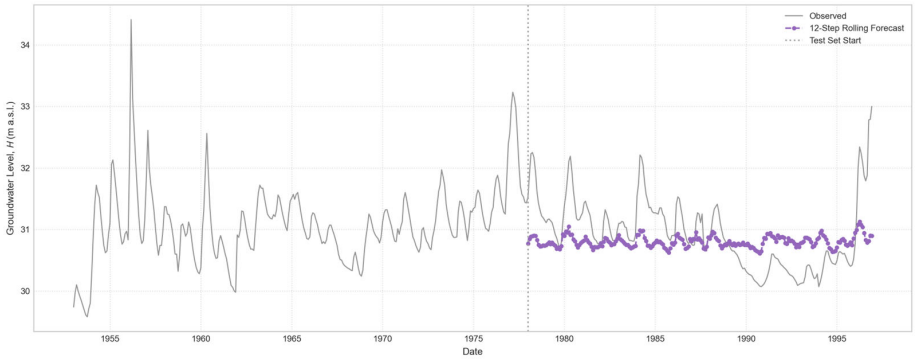

Figure 11. 12-step horizon rolling multi-step forecast for "Casa Cantoniera" well, generated by the individually trained model.

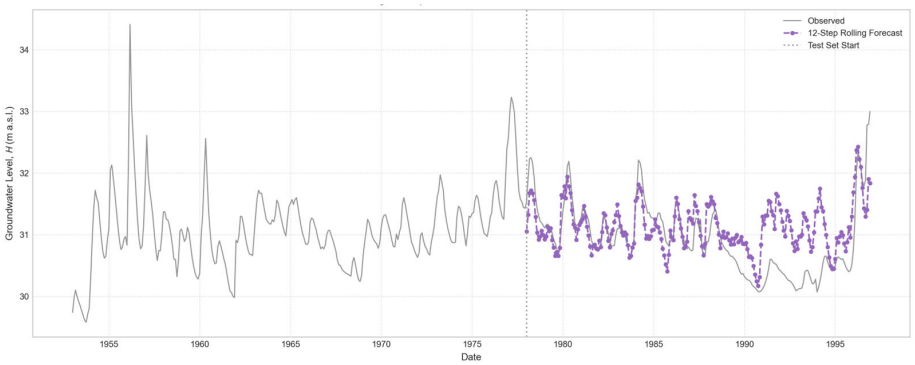

Figure 12. 12-step horizon rolling multi-step forecast for "Casa Cantoniera" well, generated by the globally trained model.

4.3. Closed Loop—Full Simulation

The trend of improved performance with an increasing number of wells was also confirmed in the closed-loop simulations (Figure 13, Table 6). This result was consistent with the performance observed in the open-loop and 12-step rolling forecast analyses.

Table 6. Model performance evaluation, full Closed-Loop simulation, for individual wells ("1 Well alone") versus models supported by five additional wells ("1 Well supported by 5 Wells").

Well	MSE	MAE	R2	MSE	MAE	R2
	1 Well Alone			1 Well Supported by 5 Wells		
Casa Cantoniera	0.051167	0.20294	−1.99272	0.013089	0.086262	0.234419
Cellino San Marco	0.014317	0.107462	−3.60932	0.001108	0.025084	0.643265
Francavilla Fontana	0.067972	0.211615	−1.46748	0.021038	0.100068	0.236299
Novoli	0.051417	0.193869	−0.2211	0.05477	0.208617	−0.30072
S. Pancrazio	0.010134	0.08682	−1.40912	0.015588	0.110489	−2.7058
Salice Salentino	0.02366	0.140252	−1.57054	0.010815	0.090826	−0.175

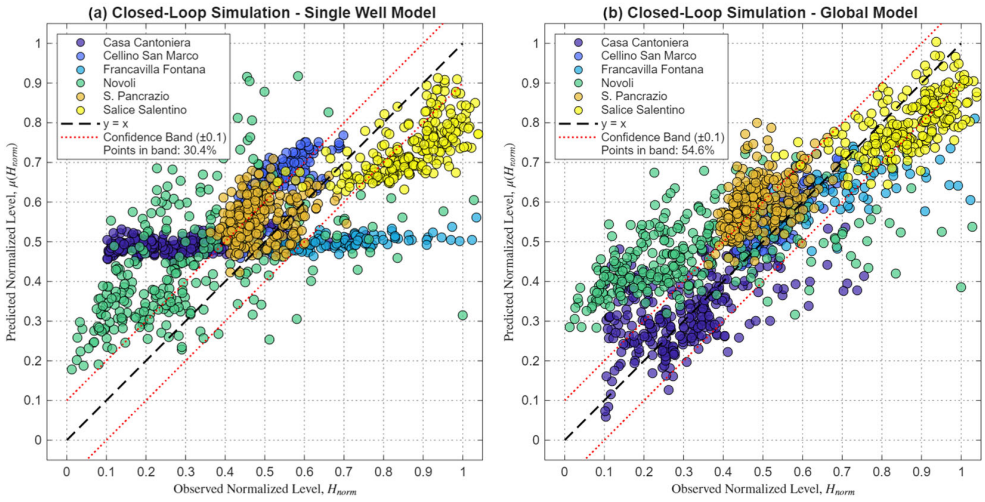

Figure 13. Scatter plots comparing the performance of the models in Closed-Loop Simulation. The plots show the observed normalized values (Hnorm) on the *x*-axis versus the predicted values on the *y*-axis. (**a**) illustrates the results for the Single Well Model (trained on each well individually), while (**b**) shows the results for the Global Model (trained on all wells simultaneously). The black dashed line indicates perfect agreement (y = x), and the red lines define a confidence band of ±0.1.

It should be noted, however, that the error propagation inherent in the fully autoregressive, closed-loop approach leads to a general decline in performance, even for the global LSTM configuration. For the "Casa Cantoniera" well, similarly to what was observed in other modes, the intense drought of 1990–1995 represented a significant challenge, adversely affecting the model's predictive accuracy during that specific interval (Figures 14 and 15).

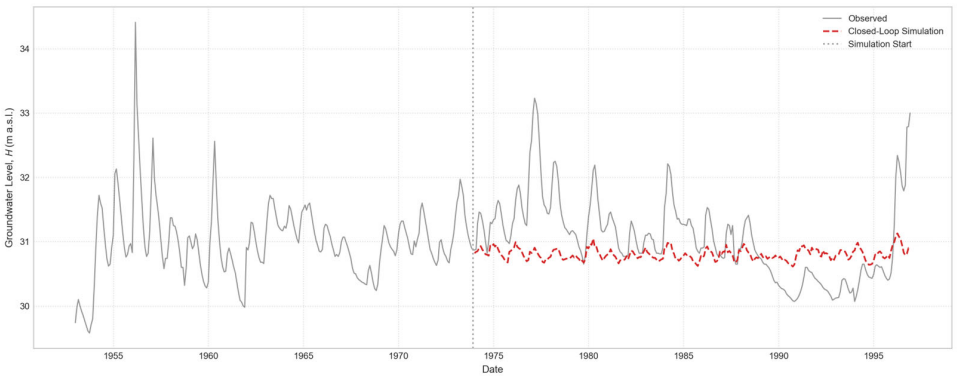

Figure 14. Closed-Loop simulation forecast for "Casa Cantoniera" well, generated by the individually trained model.

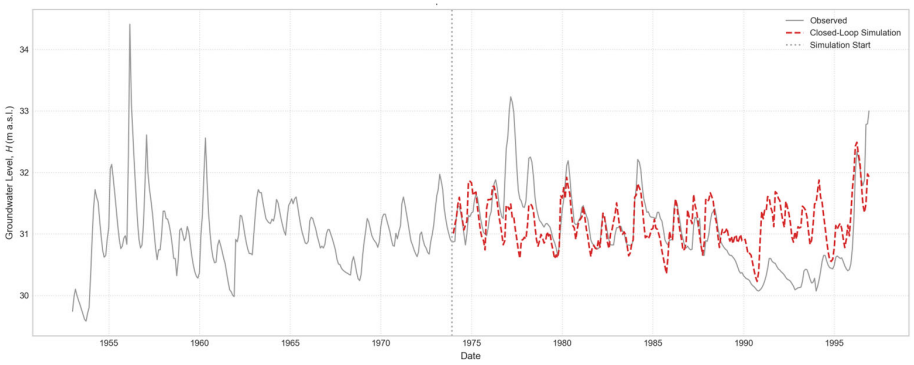

Figure 15. Closed-Loop simulation forecast for "Casa Cantoniera" well, generated by the globally trained model.

4.4. Sensitivity Analysis

The vast number of results generated from the 63 model configurations necessitated a systematic approach for their presentation and interpretation. To clearly illustrate the correlation between the number of wells included in the training set and the resulting predictive performance, the output data were aggregated for each of the six monitored wells. Specifically, while the outputs for the single-well models and the global model (trained on the entire six-well dataset) are presented individually, the performance for intermediate combinations (e.g., models trained on 2, 3, 4, or 5 wells) is reported differently. For these cases, the performance for any given well is expressed as the average of the results from all model configurations in which that well was included.

For example, to evaluate the performance of the "Casa Cantoniera" well within the two-well combination scenarios, its reported metric is the average calculated from all five possible pairs that included "Casa Cantoniera".

This aggregation method provides an effective way to summarize the model's overall performance trend as the number of wells used for training increases. Figures 16–21 show the one month ahead open-loop prediction results.

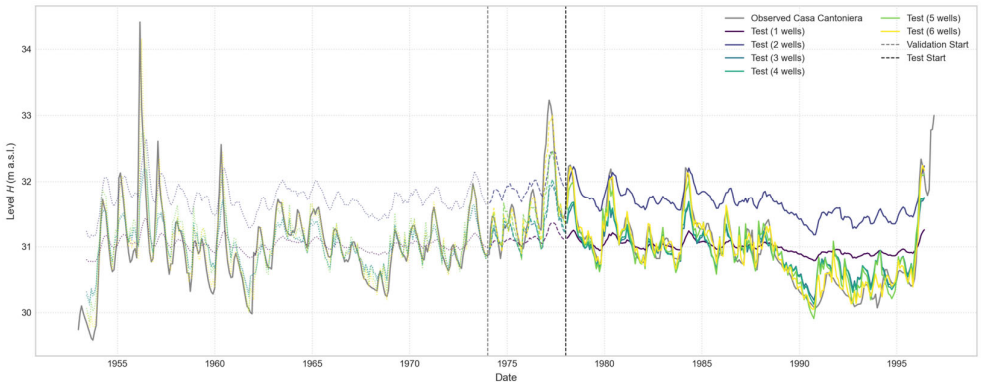

Figure 16. Comparison of Open-Loop predictions for the "Casa Cantoniera" well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

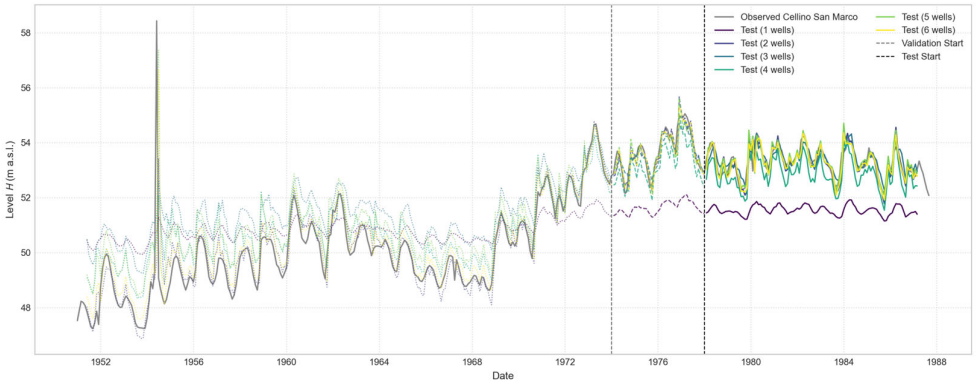

Figure 17. Comparison of Open-Loop predictions for the "Cellino San Marco" well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

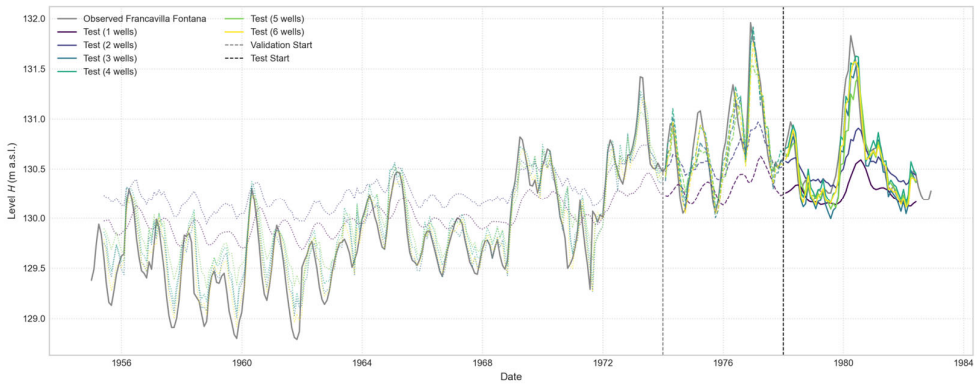

Figure 18. Comparison of Open-Loop predictions for the "Francavilla Fontana" well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

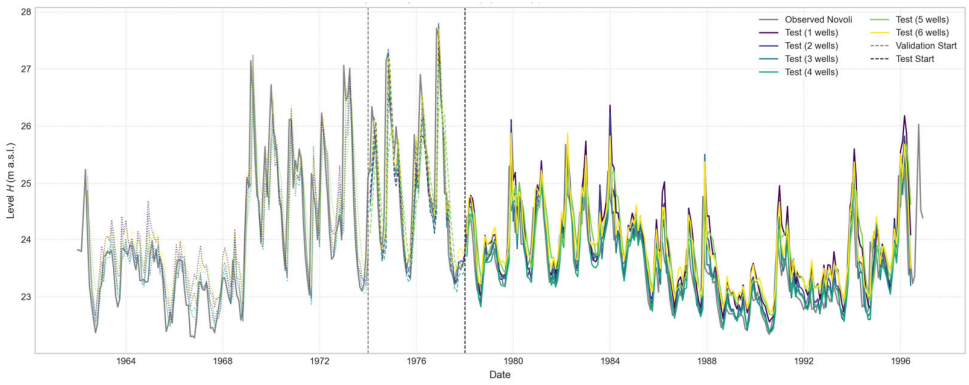

Figure 19. Comparison of Open-Loop predictions for the "Novoli" well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

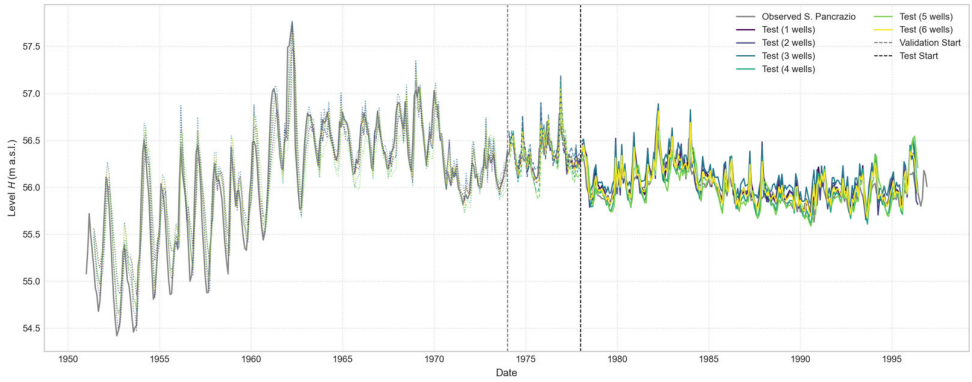

Figure 20. Comparison of Open-Loop predictions for the "S. Pancrazio" well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

The data show that all wells exhibited an increase in predictive accuracy as the number of training wells grew. In particular, the "Casa Cantoniera", "Cellino San Marco", and "Francavilla Fontana" wells demonstrate a significant performance enhancement, especially when scaling from a single-well model to a four-well model configuration. The peak performance for all wells was achieved with the global model, which was trained on the data from all six wells simultaneously.

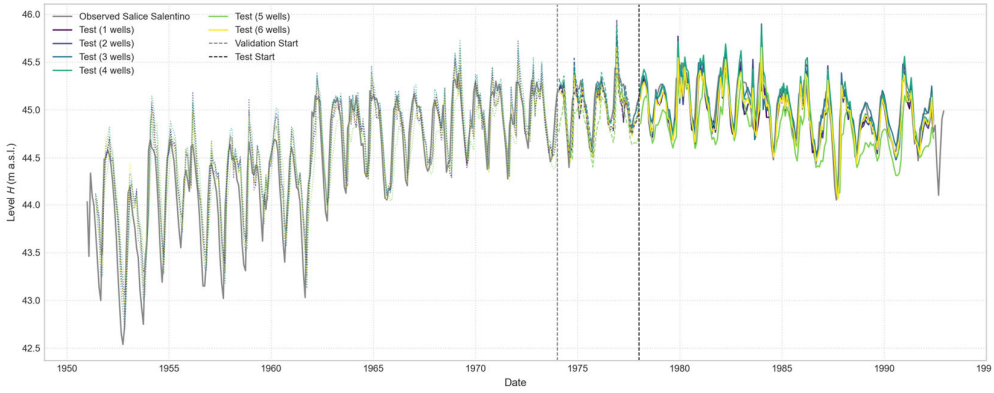

Figure 21. Comparison of Open-Loop predictions for the Salice Salentino well, with a 1-month forecast horizon. The plot illustrates the performance of a series of models, each trained using an increasing number of input wells, from one to six.

4.5. Embedding Dimensions Results

To gain insight into the internal representations learned by the model, the embedding vectors from the global model (trained simultaneously on all six wells) were extracted and analyzed. Although these vectors exist in a high-dimensional space, dimensionality reduction techniques were applied to project them onto a two-dimensional plane for visualization, as shown in Figure 22 (left).

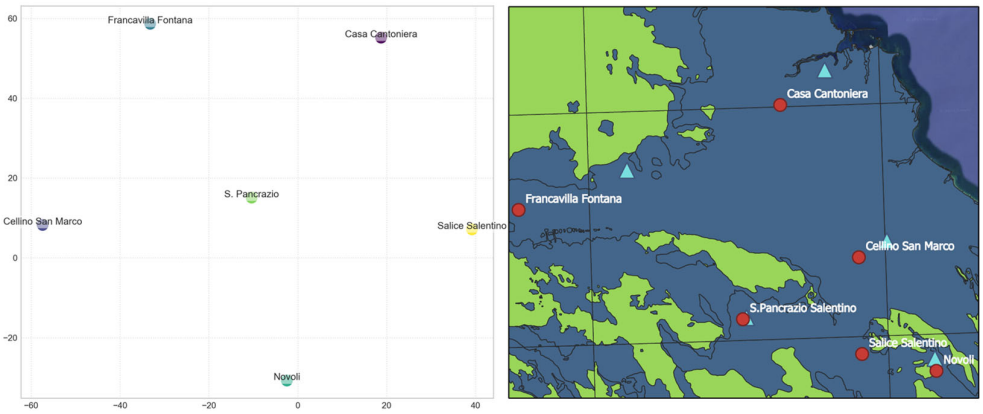

Figure 22. Correlation between the positioning of the wells in the model's learned embedding space (**left**) and their actual geographical layout (**right**), where in the right figure the red circles correspond to sampling wells, while cyan triangles represent rain gauge stations.

The resulting plot reveals a significant characteristic: the wells are distributed as distinct points rather than coalescing into separate, tight clusters. This spatial distribution suggests that the model did not simply group the wells into independent subsets. Instead, it learned a complex set of shared dependencies and relationships that link all six locations, treating them as components of a single, interconnected system. The most significant result emerges when comparing the learned embedding space (Figure 22, left) with the actual geographical map of the wells (Figure 22, right). A clear visual correlation is apparent: the relative arrangement of the wells in the abstract embedding space closely mirrors their physical proximity in the real world.

This outcome is of critical importance. It demonstrates that the learned embeddings are not merely abstract parameters but are grounded in the physical reality of the aquifer system. Even if not directly implemented, they agree with observed hydrological responses reported in previous studies [1,3,8]. This provides a powerful and data-driven method for understanding and quantifying the hydrogeological dependencies between the time series

of the various wells, offering insights that might not be immediately obvious from the raw data alone.

4.6. Simulation with Different Rain Scenarios

In modeling hydrogeological systems, running simulation scenarios is a powerful tool for the in-depth understanding of relationships and responses between precipitation and groundwater levels (GWL). These simulations are particularly valuable for assessing the model's ability to generalize extreme scenarios not encountered during the training phase. Observing the GWL response to specific rainfall inputs in a controlled setting helps to identify the most influential factors and to determine the time lag of their effects.

The best-performing model from the previous analyses, the global model, was selected for these simulations. Several scenarios (Figure 23) were designed based on normalized precipitation values to evaluate their response:

- Sustained Drought: Precipitation is set to the historical minimum value (or 0 if the minimum is undefined), replicating prolonged drought conditions.
- Sustained Average P: Precipitation is set to the historical average value.
- Sustained High P (P75): Precipitation is set to the 75th percentile of historical values, simulating a period of abundant but not extreme rainfall.
- Rainfall Pulse: This is a dynamic scenario that simulates an initial period of drought followed by a period of intense rainfall (90th percentile or the historical maximum, if the 90th percentile is zero or undefined), and then a return to average conditions.

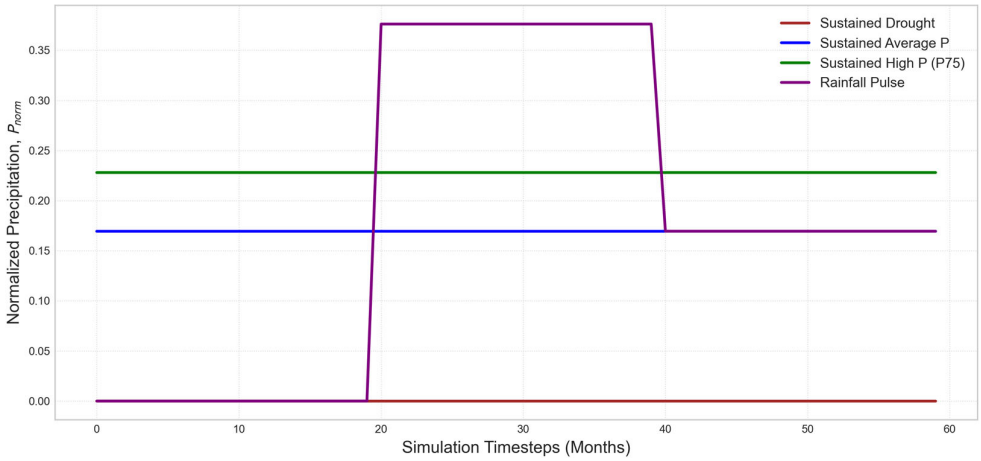

Figure 23. Graphical representation of the four precipitation scenarios used for model simulations over a 60-month period. Each scenario is defined by a time series of normalized precipitation.

As expected, the groundwater level (GWL) simulations reveal distinct responses corresponding to the different rainfall scenarios (Figure 24). Under the Sustained Drought scenario (red line), the wells at Cellino San Marco and Casa Cantoniera exhibit the most pronounced GWL declines, indicating high vulnerability to prolonged dry conditions. In contrast, wells like S. Pancrazio show a more delayed and muted response.

With Sustained Average Precipitation (blue line), the GWL in most wells, such as Cellino San Marco, Casa Cantoniera, and S. Pancrazio, stabilizes or slightly increases. Francavilla Fontana and Salice Salentino show a minor declining trend, whereas Novoli is the only well to display a significant, steady GWL rise under average conditions. The Sustained High P (P75) scenario (green line) is the only condition sufficient to induce a consistent GWL increase across all monitored wells.

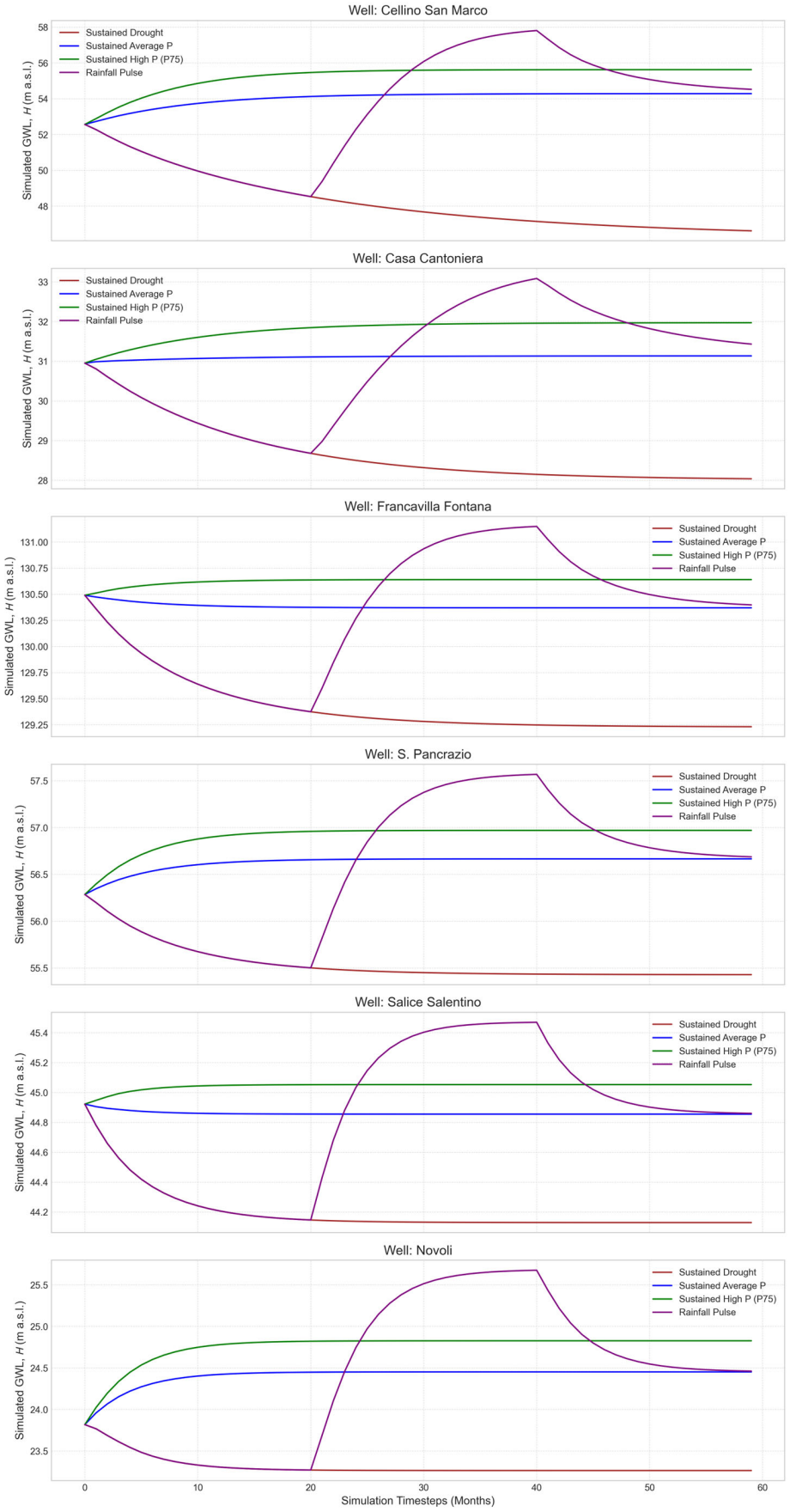

Figure 24. Groundwater level projections for all six wells, comparing all scenarios. The plots show the predicted mean (solid lines) and uncertainty ranges (shaded areas) for each scenario.

The Rainfall Pulse scenario (purple line) highlights differences in aquifer reactivity. While all wells show an initial decline followed by a sharp recovery, the response dynamics vary. Wells like Cellino San Marco, Casa Cantoniera, and Francavilla Fontana demonstrate a highly dynamic recovery with a significant and rapid rise in water levels. The response at Novoli is also very pronounced, while the recovery at S. Pancrazio and Salice Salentino appears comparatively more dampened.

5. Discussion and Conclusions

This study aimed to improve the available methodological approaches for understanding the complex relationships among rainfall supply and water levels in wells using data from the shallow aquifer of Brindisi, Puglia. To achieve this, we trained a single global Long Short-Term Memory (LSTM) model for modeling and forecasting groundwater levels, specifically considering wells whose data are not directly correlated but which are located within the same aquifer and share similar hydrogeological responses.

Unlike traditional machine learning approaches, which often treat time series well independently or with localized models, our methodology introduces a 'global' LSTM model capable of simultaneously processing multiple time series of groundwater levels and precipitation. This is enhanced by an embedding layer that allows the model to learn the latent relationships between the different wells.

The results demonstrate that the global model, trained with data from the six wells simultaneously, not only allows obtaining superior performance metrics but also highlights its ability to generalize. Thus, it is possible to improve the prediction performance without the use of other parameters, which are not always available and with the same reliability of measures.

Also, the results demonstrate that including a greater number of wells during the LSTM model's training phase translates into a significant improvement in predictive accuracy. This improvement, observed in all simulation modes (open-loop, closed-loop, and rolling forecast), suggests that the model leverages joint information from multiple locations to capture the complex and general hydrogeological dynamics of the aquifer. In particular, the reduction in predictive uncertainty, quantified by the standard deviation (σ), indicates that learning from larger and more diverse datasets leads to forecasts that are not only more accurate but also more reliable and robust. The clear correlation between the arrangement of the wells in the two-dimensional embedding space and their actual geographical proximity (Figure 22) provides empirical evidence that the model has autonomously identified and quantified the hydrogeological dependencies between the different time series.

This study aligns with the latest trends in the application of LSTMs in hydrology [5–7], confirming their suitability for forecasting groundwater levels. However, it represents an advancement over previous studies that had not systematically explored the impact of increasing the number of wells on the predictive performance of an LSTM model.

Applying the model to other aquifers would allow for additional analyses on its generalizability and to explore its predictive capabilities in the scenario of climate change.

Future work could explore the integration of genetic algorithms to perform dual optimization: simultaneously identifying the optimal combination of wells that maximize performance while also finding the ideal set of model hyperparameters for each specific LSTM configuration. Additionally, the implementation of a larger number of wells across different aquifers would allow for more comprehensive analyses. Finally, incorporating further variables such as evapotranspiration, land use, and pumping may be particularly effective at modeling other systems where, unlike the investigated aquifers, these factors heavily influence groundwater behavior.

Author Contributions: L.D.T. worked on data processing and wrote the text of the manuscript; A.F. worked on data processing; A.D. contributed to the methodology conceptualization and the analysis of results; V.S. contributed with a critical review of the manuscript and on the analysis of results. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

LSTM	Long Short-Term Memory
MSE	Mean Absolute Error
MAE	Mean Absolute Error
R2	Coefficient of Determination
EPR	Evolutionary Polynomial Regression
RNN	Recurrent Neural Networks
NARX	Nonlinear Autoregressive Networks with Exogenous Inputs
CNN	Convolutional Neural Network
NLL	Negative Log-Likelihood
PDF	Probability Density Function
GWL	Groundwater Level

References

1. 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]
2. Giustolisi, O.; Savic, D. A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinform.* **2006**, *8*, 207–222. [CrossRef]
3. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
4. Vizi, Z.; Batki, B.; Rátki, L.; Szalánczi, S.; Fehérváry, I.; Kozák, P.; Kiss, T. Water level prediction using long short-term memory neural network model for a lowland river: A case study on the Tisza River, Central Europe. *Environ. Sci. Eur.* **2023**, *35*, 92. [CrossRef]
5. Feng, Z.K.; Zhang, J.S.; Niu, W.J. A state-of-the-art review of long short-term memory models with applications in hydrology and water resources. *Appl. Soft Comput.* **2024**, *167*, 112352. [CrossRef]
6. Wunsch, A.; Liesch, T.; Broda, S. Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1671–1687. [CrossRef]
7. Doglioni, A.; Simeone, V. Evolutionary Modeling of Response of Water Table to Precipitations. *J. Hydrol. Eng.* **2017**, *22*, 04016055. [CrossRef]
8. Mancarella, D.; Simeone, V. Modellazione e previsione nei sistemi idrogeologici mediante la tecnica E.P.R. (Evolutionary Polynomial Regression). *G. Geol. Appl.* **2008**, *8*, 8–16.
9. Geron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019.
10. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef] [PubMed]
11. Sahoo, B.B.; Jha, R.; Singh, A.; Kumar, D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys.* **2019**, *67*, 1471–1481. [CrossRef]
12. Camplani, M.; Cannas, B. The Role of the Embedding Dimension and Time Delay in Time Series Forecasting. *IFAC Proc. Vol.* **2009**, *42*, 316–320. [CrossRef]
13. Hilbe, J.M. *Negative Binomial Regression*; Cambridge University Press: Cambridge, UK, 2011.
14. Gu, W.; Tandon, A.; Ahn, Y.-Y.; Radicchi, F. Principled approach to the selection of the embedding dimension of networks. *Nat. Commun.* **2021**, *12*, 3772. [CrossRef] [PubMed]

15. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [CrossRef]
16. Caruana, R.; Lawrence, S.; Giles, C. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. In *Advances in Neural Information Processing Systems*; Leen, T., Dietterich, T., Tresp, V., Eds.; MIT Press: Sebastopol, CA, USA, 2000.

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