Accepted version
doi: 10.1016/j.jhydrol.2022.129002

Deep learning for earthquake hydrology? Insights from the karst Gran Sasso aquifer in central Italy

Anna Rita Scorzini [a,*], Mario Di Bacco [a], Gaetano De Luca [b], Marco Tallini [a]

[a] Dipartimento di Ingegneria Civile, Edile-Architettura e Ambientale, Università degli Studi dell'Aquila, Via G. Gronchi, 18, 67100 L'Aquila, Italy

A.R.S.: annarita.scorzini@univaq.it ; M.D.B.: mario.dibacco@univaq.it ; M.T.: marco.tallini@univaq.it

[b] Istituto Nazionale di Geofisica e Vulcanologia – Osservatorio Nazionale Terremoti, Via F. Crispi, 43-45, 67100 L'Aquila, Italy

G.D.L.: gaetano.deluca@ingv.it

* Corresponding author: annarita.scorzini@univaq.it

Abstract

By leveraging monitoring data for the Gran Sasso carbonate aquifer during two significant seismic sequences that hit central Italy in recent years, this study investigates the possibility of using memory-enabled deep learning algorithms as meaningful tools for an enhanced modelling of the hydrological response of karst aquifers subject to earthquake phenomena. Meteorological, hydrological and seismic data are used to train and validate long short-term memory networks (LSTM) in one- and multiple-day ahead flow forecasting exercises, aimed at assessing model sensitivities to input variables and modelling choices (training data and parameters of the models). Results indicate that the models fairly reproduce the flow patterns for the considered spring in the Gran Sasso aquifer, thus supporting the potential use of these models for hydrological applications in similar areas, provided that sufficient data are available for the training of the network.

Keywords: earthquake hydrology; seismic sequences; karst aquifer; deep learning; LSTM; central Italy

1. Introduction

The groundwater flow in fractured karst aquifers is the result of the complex interplay between many components, involving hydrological processes (e.g., rainfall and snow accumulation, surface runoff, infiltration), geological and morphological features of the aquifer, as well as anthropogenic factors (Bonacci, 2004; Ding et al., 2020). All these elements affect recharge-discharge mechanisms with overlapping effects, thus making the identification of individual contribution of each component to the final outflow very difficult (Duran et al., 2020). The quantitative modelling of karst hydrology typically relies on physically-based or conceptual models, which are however inherently uncertain (Birk et al., 2006; Hartmann et al., 2014; Li et al., 2016; Zoghbi and Basha, 2019; Duran et al., 2020).

In seismic areas the situation is even more complex, given that earthquakes are known to be responsible for significant alterations in the hydrological systems, as a consequence of dynamic strain modifications inducing formation of microcracks, fracture cleaning phenomena and temporary changes in permeability (Brodsky et al., 2003; Casini et al., 2006; Wang and Manga, 2010; Adinolfi Falcone et al., 2012; Manga and Wang, 2015). These alterations cause short- and/or long-term impacts on spring and river flows (Rojstaczer et al., 1995; Manga et al., 2003; Montgomery and Manga, 2003; Manga and Wang, 2015; Mohr et al., 2017; Petitta et al., 2018; Di Matteo et al., 2020), water table (Roeloffs, 1998; Wang et al., 2004; Wang and Chia, 2008; Shi et al., 2015; Petitta et al., 2018) as well as chemical composition of water (Claesson et al., 2004; Onda et al., 2008; Adinolfi Falcone et al., 2012; Barberio et al., 2017; Nakagawa et al., 2020).

In recent years, the aquifers in the Italian Apennines have been described in many studies to be prone to hydrological alterations following major seismic events, with a common observed increase in spring and streamflow discharge. For instance, Esposito et al. (2001) reported significant hydrogeological changes within 200 km from the epicentral zone for four earthquakes that struck central-southern Apennines in the last two centuries. Amoruso et al. (2011) and Adinolfi Falcone et al. (2012) analyzed the changes in the hydrogeology of the Gran Sasso carbonate aquifer after the 2009 L'Aquila earthquake. Increases in river and spring flows, as well as in water table, have been registered also for the most recent 2016-2017 seismic sequence in Central Italy, with effects observed up to a distance of 100 km from the epicentral area (De Luca et al., 2018; Petitta et al., 2018; Mastrorillo et al., 2020; Valigi et al., 2019; Di Matteo et al., 2020).

Despite the relatively large amount of literature on the topic, none of these previous studies have tried to quantitatively interpret the relationship between fluctuations in groundwater flow, precipitation, snowmelt, etc. and the features of the seismic events (e.g., magnitude and distance between the observation site and the epicenter). In such contexts, when the underlying relationships cannot be fully explained in physical terms, machine learning, data-driven approaches may be considered useful tools for modelling the hydrological response of the aquifer (Dikshit et al., 2020). These methods, starting with simpler artificial neural networks (ANNs), have been proved to be effectively employed in many hydrological applications, from streamflow forecasting to groundwater and water management modelling (Sivakumar et al., 2002; Hu et al., 2008; Wu et al., 2009; He et al., 2014; Zhang et al., 2015). In recent years, to overcome ANNs' limitation in not properly accounting for the sequential order information of

input data, recurrent neural networks (RNNs) have been developed to selectively convey information across sequence steps (Kratzert et al., 2018; Dikshit et al., 2020). Among RNNs, long short-term memory (LSTM) neural networks (Hochreiter and Schmidhuber, 1997) have gained increasing attention among hydrologists, due to their powerful capability in learning long-term dependencies in time series data (Kratzert et al., 2018; Le et al., 2019; An et al., 2020; Cheng et al., 2020).

The objective of the present paper is then to test the implementation of deep learning algorithms (LSTMs) to analyze the possible contribution of these approaches in supporting a quantitative modelling of the hydrological alterations in karst aquifers subject to seismic activity and, then, to assess their potential as meaningful tools for reliably predicting spring flows in such areas. The study region is represented by the northern part of the Gran Sasso carbonate aquifer, impacted by the 2009 and 2016-2017 seismic sequences in central Italy, which induced abrupt changes in the flow regime of monitored springs. Details on the available hydrometeorological and seismic variables, as well as on the methodological approach followed in this study, are provided in Section 2, while the results, presented and discussed in Section 3, form the basis for the conclusions reported in Section 4.

2. Material and methods

2.1. Overview of the study area and available data

The study area (Figure 1) is the Gran Sasso carbonate aquifer, located in the northern part of the Abruzzo region (Italy), which, as most of the central-southern Apennines, is a seismic area characterized by normal faulting earthquakes reaching maximum magnitudes close to 7.0 (De Luca et al., 2000, 2009; Bagh et al., 2007; Frepoli et al., 2017). This is a large fractured aquifer with well-defined boundaries, extending over an area of about 700 km^2 (Petitta and Tallini, 2002) and perennial groundwater reserves estimated in the order of 10^{10} m^3 (Amoruso et al., 2011). Due to its karst features, with joints, faults and intense fracturing of rock masses, recharge quantities are huge and characterized by a net infiltration of over 800 mm per year (Boni et al., 1986; Scozzafava and Tallini, 2001), on an annual average rainfall of about 1200 mm (Curci et al., 2021). The Gran Sasso supplies a total discharge of more than 18 m^3/s from its springs, with the major ones located at the discharge zones at its boundaries. Since the 1980s, after the construction of underground infrastructures, including highway tunnels and laboratories, a portion of the groundwater has been partially drained and exploited for drinking purposes (Adinolfi Falcone et al., 2008). This drainage system consists of two 5 km long channels conveying water to the south-western (Assergi site (AS), Figure 1) and north-eastern exists of the highway tunnels (Petitta and Tallini, 2002). As described in the Introduction, previous studies described significant short and mid-term changes in the hydrological characteristics of groundwater flow at the Assergi site after the 2009 earthquake in L'Aquila (with an epicentral distance of about 14 km from AS), as a possible consequence of changes in the pore pressure and increase in the equivalent hydraulic conductivity of the aquifer, induced by mechanisms of fracture unclogging (Amoruso et al., 2011; Adinolfi Falcone et al., 2012). Indeed, the available hydrological data measured on April 2009 indicated a sudden increase in the spring discharge, which remained higher than the historical flow registered before the mainshock for several months after the event, as typical in fractured aquifers, where recovery

to pre-earthquake conditions is governed by the time to reblock the fractures (Manga et al., 2012; Mastrorillo et al., 2020; Vittecoq et al., 2020). Moreover, quick flow alterations, favored by the fractured carbonate nature of the aquifer, were also observed for the seismic sequence occurred in 2016-2017 in central Italy, with epicentral area located about 30 km away from the spring located at the end of the drainage tunnel in Assergi (Figure 1). This sequence started on August 24th 2016, near Amatrice (M_L (local magnitude) 6.0), Lazio Region, and it continued in the following months with several events of magnitude greater than 5.0, culminating in the main shock on October 30th 2016, when a M_L 6.5 earthquake occurred near Norcia in the Umbria Region, at a distance of about 50 km from the Assergi site.

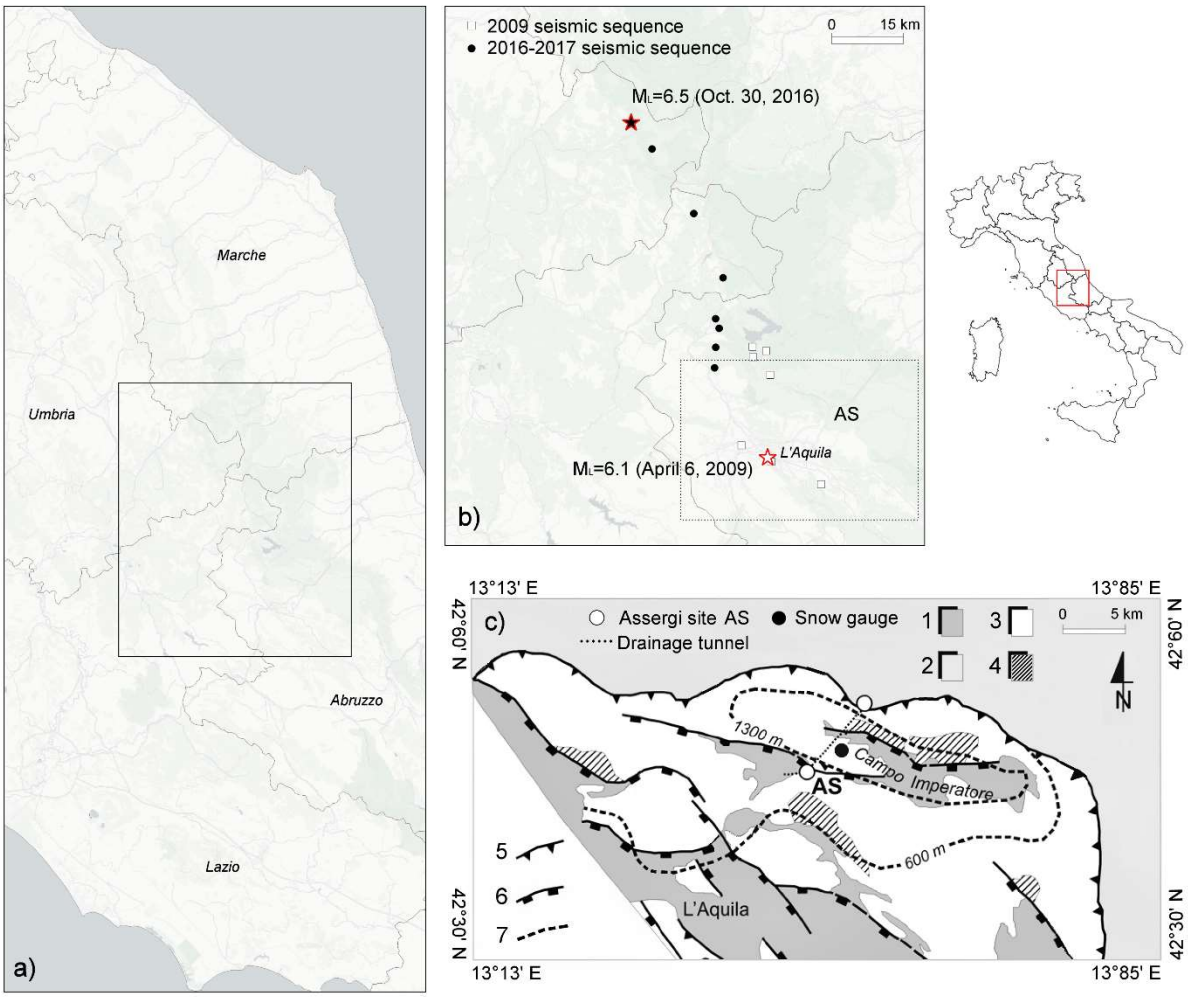


Figure 1. a) Overview of the study area (northern part of the Gran Sasso aquifer in central Italy); b) location of the main seismic events during the 2009 and 2016-2017 sequences (indicated with the red bars in Figures 2 and 3); c) hydrogeological setting, with indication of the drainage tunnel and Assergi (AS) and Campo Imperatore sites: 1. aquitard (continental detrital units of intramontane basins, Quaternary); 2. aquiclude (terrigenous turbidites, Mio-Pliocene); 3. aquifer (calcareous sequences of platform – reef included – and slope to basin lithofacies, Meso-Cenozoic); 4. low permeability substratum (dolomite, upper Triassic); 5. thrust; 6. extensional fault; 7. presumed water table in m a.s.l.

Within the mentioned seismic sequence, sub-hourly observed flow data were provided by the water supply company which manages the tapped spring. These measurements refer to monitoring sensors applied to derived discharge channels for the south-western drainage tunnel of the Gran Sasso aquifer (AS). The available discharge (Q) time series spans over September

2015 – January 2018, which was the maximum time window characterized by continuous records, without missing data and/or system malfunctioning. The fourth panel of Figure 2 clearly highlights the abrupt flow increases observed in the spring discharge after the main earthquake events in central Italy during 2016-2017, as also described by Petitta et al. (2018) and Mastrorillo et al. (2020) in other neighboring springs.

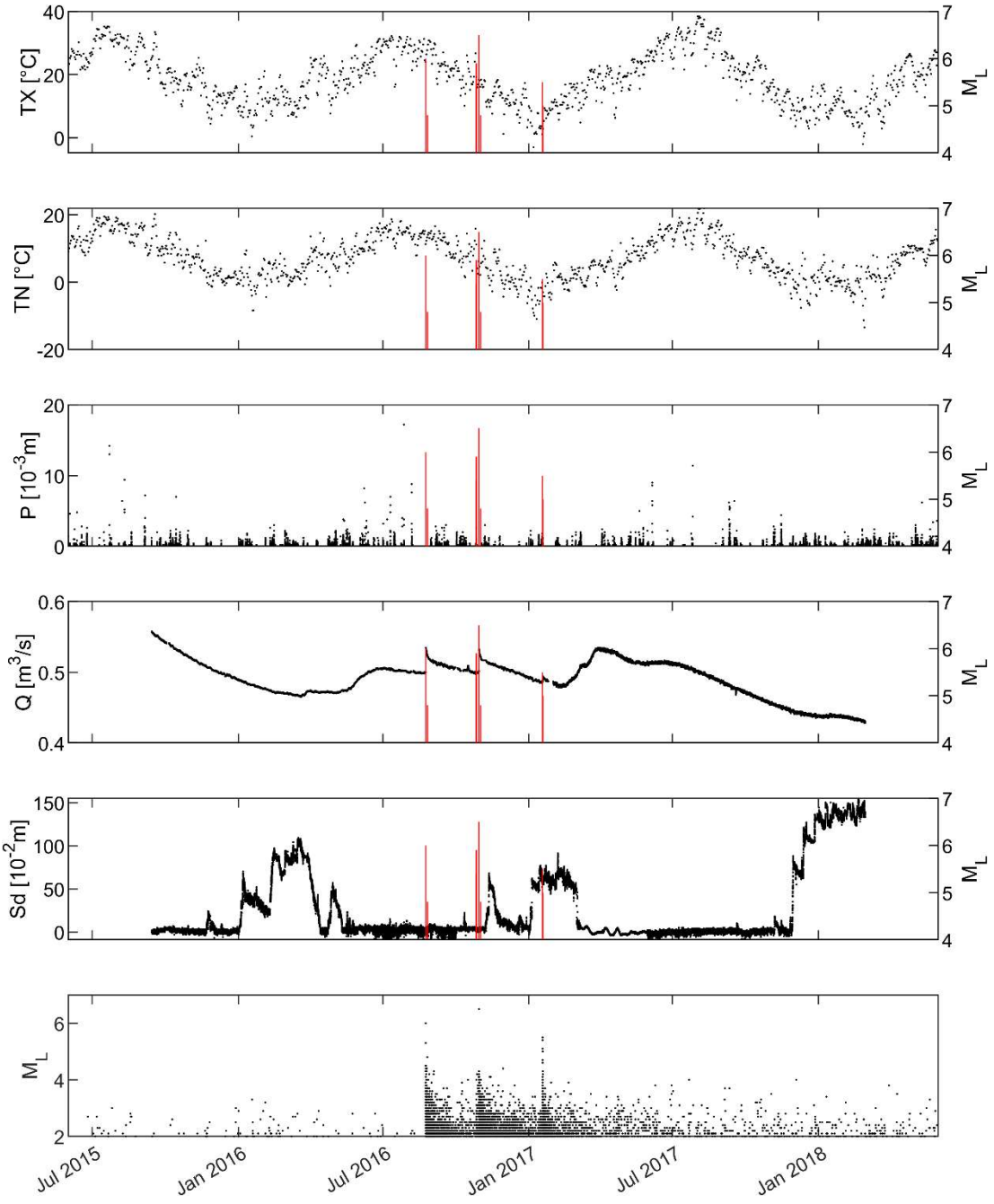


Figure 2. Hydrometeorological and seismic time series available for the study area in the 2015-2018 period: maximum and minimum temperature (*TX* and *TN*), rainfall (*P*), spring discharge (*Q*), snow depth (*Sd*) and earthquake magnitude (M_L). The red bars highlight the main events of the seismic sequence.

Complete seismic data, consisting of magnitude ($M_L \geq 2.0$) and epicentral distance (d_e) from the examined spring site, were retrieved from Istituto Nazionale di Geofisica e Vulcanologia (INGV, http://cnt.rm.ingv.it/); the events outside a radius of 60 km from Assergi were excluded from the analysis, in agreement with the observations of Petitta et al. (2018), who qualitatively described alterations in the flow regimes at different locations in central Italy.
In order to account for the effect of other meteorological factors on the hydrological response of the aquifer, rainfall (*P*), snow depth (*Sd*), minimum and maximum temperature (*TN* and *TX*) data over the considered period were obtained from the Hydrographic Office of the Abruzzo Region. Assergi (AS) was selected as the most representative site for rainfall and temperature measurements, while Campo Imperatore for snow depth data (Figure 1 and 2).
The same kind of information were also retrieved for the period encompassing the 2009 seismic sequence, where reliable flow data ranged from January 2009 to October 2010 (Figure 3), while for other periods, the flow time series were discontinuous and not useful for the aim of this study. Because of the different data sources, the time series were characterized by sampling rates (Table 1) ranging from 15 minutes to 24 hours; to ensure uniformity, data were then aggregated and analyzed at daily scale. Regarding earthquake data, in case of multiple daily events, the one with the highest magnitude was retained for the analysis.
Figure 4 shows the time-lagged cross-correlation between the spring flow and the other considered variables, obtained by incrementally shifting the different time series vectors and repeatedly calculating the Pearson correlation coefficient between these shifted signals and *Q*. It can be recognized the delayed recharge effect caused by snowmelt and driven by temperatures, while weaker appears the contribution of rainfall due to the storage capacity of the aquifer. As also visible from the time series (Figures 2 and 3), M_L seems to be more related to daily variations of *Q* and then the peak correlations reported in Figure 4 at certain lags may be considered a spurious consequence of synchronous (lagged) flow seasonality and occurrence of intense seismic activity. Moreover, the different correlation patterns shown in Figure 4 for the two considered periods are mostly due to the occurrence of the main earthquakes in different seasons of the year, which are late summer-autumn (recession phase) in 2015-2018 and spring (rising limb) in 2009-2010.

Table 1. Location and original sampling rates for the hydrometeorological variables considered in the study.

Location	Variable	Original sampling rate
Southwestern drainage of Gran Sasso aquifer (Assergi)	Spring discharge	15 min (2015-2018) daily (2009-2010)
Assergi	Rainfall	15 min (2015-2018) daily (2009-2010)
Assergi	Temperature (max, min)	daily
Campo Imperatore	Snow depth	15 min

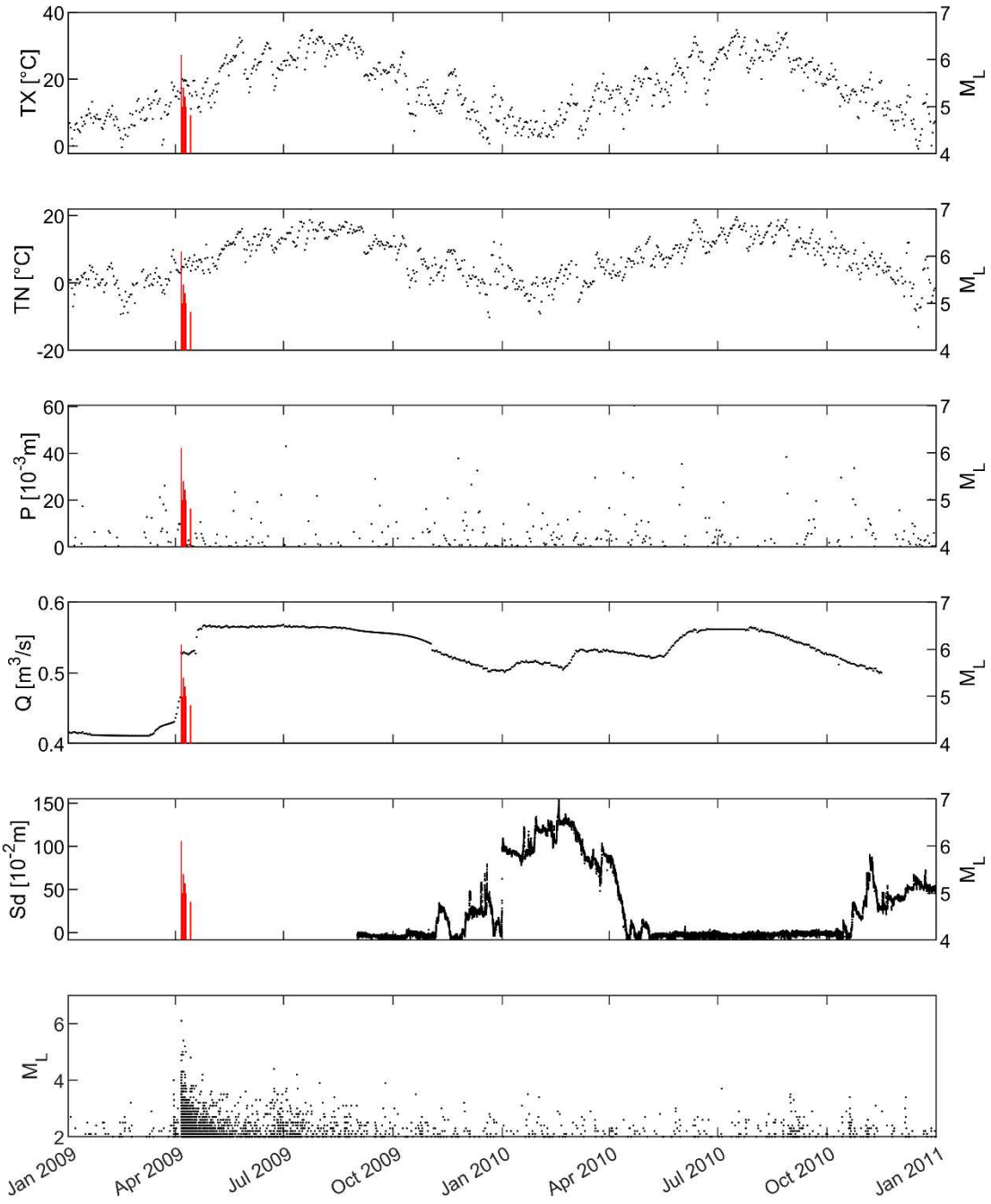


Figure 3. Hydrometeorological and seismic time series available for the study area in the 2009-2010 period: maximum and minimum temperature (*TX* and *TN*), rainfall (*P*), spring discharge (*Q*), snow depth (*Sd*) and earthquake magnitude (M_L). The red bars highlight the main events of the seismic sequence.

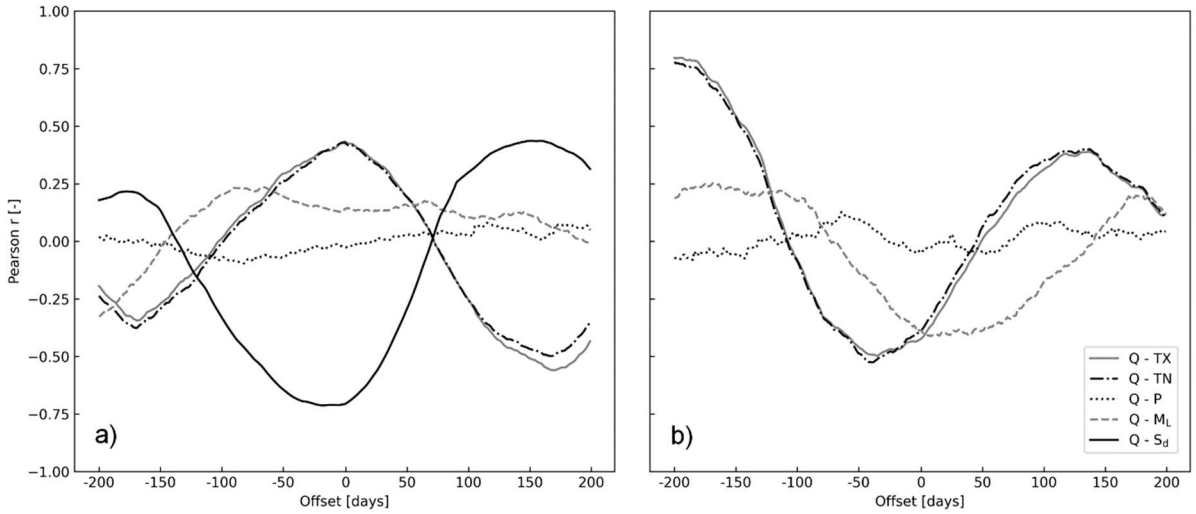


Figure 4. Time-lagged cross-correlation between the spring flow and the hydrometeorological and seismic variables considered in this study: a) 2015-2018 period; b) 2009-2010 period (in this case, the correlation between *Q* and *Sd* is not calculated due to the limited length of the *Sd* series).

2.2. Implemented machine-learning algorithms

In this study, the flow forecasting at daily scale in the considered karst aquifer subject to earthquake phenomena was carried out by using memory-enabled machine learning algorithms.

RNNs are a type of neural networks capable of learning order dependence in sequence data (Rumelhart et al., 1985), although to a limited extent (Bengio et al., 1994). Indeed, while their hidden state should be able to preserve the memory of past input data, they encounter vanishing and exploding gradient problems for network training that can hinder their capacity of learning long-term dependencies (Bengio et al., 1994). LSTMs have been then developed to overcome this issue, by including, in their recurrent structure, cell state and gating mechanisms which control the flow of information through the network (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018).

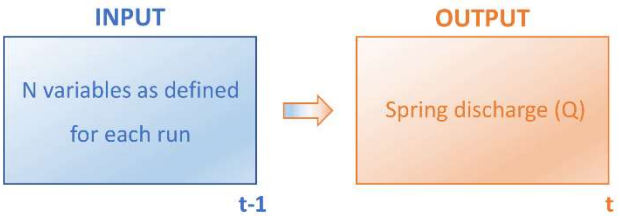
While for the specific mathematical expressions of LSTMs the readers can refer to the original published literature (Rumelhart et al., 1985; Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018), hereinafter we focus on the methodological approach followed in this study to investigate the possibility of using a deep learning approach for simulating the flow patterns in karst aquifers subject to earthquake phenomena. In detail, as a consequence of the relatively short time series available for the study area, two scenarios were analyzed in order to obtain reliable insights on models' predictive ability, trying to minimize the effect of the temporal limitation in the data. The first group of runs leveraged only on the most recent time series (2015-2018) for both (appropriate) training and validation as well as test of the neural networks for one-step ahead forecasting, while in a second multistep-ahead prediction exercise, the 2015-2018 period was exclusively considered for training and validation and the 2009-2010 as a test set. As regards the first scenario, several tests were carried out to evaluate the sensitivity of the models to: (i) models' hyperparameters; (ii) start date of the learning period (ID_train), necessary due to the relatively shorth length of the available time series; (iii) number and combination of the input features considered for model training (i.e., feature importance on the

predictive accuracy). Regarding the last point, for the sake of conciseness, only ten different combinations of model input parameters were selected in model training for the one-step-ahead flow forecasting (Figure 5). Starting from the most complete combination including all the parameters at stake (I), the others originate by excluding one or two related variables (as M_L and d_e or *TX* and *TN*) at time (II-V), except for *Q* and *P*, which are instead always taken into account in the models. Runs VI and VII derive from a further simplification of the previous combinations, by considering only hydrometeorological variables, while runs VIII and IX exclude the possible influence of the epicentral distance d_e, in combination with temperature or snow depth data. The last run (X), similar to the first but without *Q*, is instead considered as a comparison mean for evaluating the impact of the information on the antecedent values of the spring flow on model accuracy.

For each feature combination, training was repeated 200 times by randomly changing ID_train and models' hyperparameters. Indeed, although it is widely acknowledged that hyperparameters are crucial for the predictive accuracy of black-box models, no established general instructions exist for setting them (Hutter et al., 2014; Reimers and Gurevych, 2017; Kratzert et al., 2018). In this study, the MATLAB's deep learning toolbox was used to develop the LSTMs and therefore, after a preliminary assessment, most of the hyperparameters were kept at their default values, while the ones considered in the sensitivity analysis are described hereinafter.

Adam optimizer (Kingma and Ba, 2014), based on stochastic gradient descent, was employed in all runs; in the optimization process, one iteration for the complete training dataset is referred to as an epoch and Nepochs defines the maximum number of iterations, which was randomly changed between 100 and 300. Other tested hyperparameters included: the number of hidden units (NHU, ranging from 40 to 100) in the LSTM layer and the batch size (BS, ranging from 40 to 100), which defines the number of samples to be processed before updating internal parameters of the LSTM model.

In the first scenario, the daily time series were split into training, validation and test datasets in the proportion of 55-20-25, with changing start date of the learning period (ID_train, up to the 400th day in the time series, as shown in the scheme provided in Figure 6).



	Variables						
Run	Q	P	Sd	TX	TN	M_L	d_e
I	■	■	■	■	■	■	■
II	■	■	■	■	■	■	
III	■	■		■	■	■	■
IV	■	■	■			■	■
V	■	■	■	■	■		
VI	■	■	■				
VII	■	■		■	■		
VIII	■	■	■			■	
IX	■	■		■	■	■	
X		■	■	■	■	■	■

Figure 5. Scheme for the developed one-day-ahead models: inputs are the available observed hydro-meteorological and earthquake data at the previous time levels *(t-1)* and the target output is the discharge at the time level *t*. The different combinations of considered input data are shown in the table on the right.

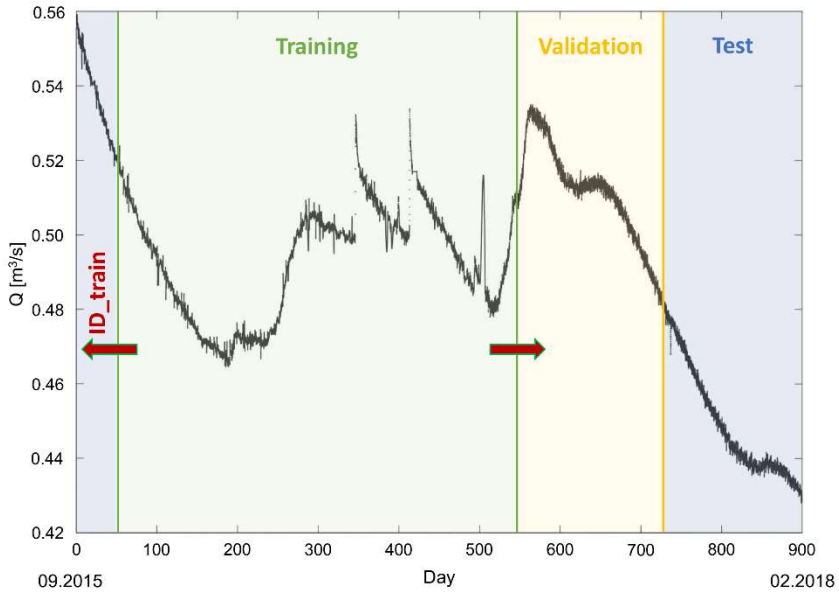


Figure 6. Scheme for the splitting of the time series into training, validation and test sets, exemplified for the flow record over the 2015-2018 period. In the experiments, the training time window is variable with changing ID_train.

The model inputs are the available observed hydrometeorological and earthquake data at the previous time levels *(t-1)*, while the target output is the discharge at the time level *t* (Figure 5). The different data series were standardized by rescaling the observations to have a null mean and a standard deviation of 1.

During training, the data input to the layers of the networks are converted by weights, which are the parameters of the layer; the simulated and observed discharges are compared using an objective function, here represented by the root mean square error (RMSE, Eq. (1)) and the weights are adjusted through the optimizer until the RMSE is minimized:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_i^0 - Q_i^p\right)^2} \tag{1}$$

with $Q_i^0$and Q_i^p respectively representing the standardized observed and predicted flow at the day *i*, while *n* is the length of the dataset.

Eq. (1) was also used to measure the accuracy of the trained networks on the test datasets, in addition to the mean absolute error (MAE, Eq. (2)) and the coefficient of determination (r^2), all calculated by always excluding the first 40 values of the time series, due to initialization issues of the implemented models (Zimmermann et al., 2012; Mehdipour Ghazi et al., 2019).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Q_i^0 - Q_i^p\right| \tag{2}$$

Furthermore, to better understand the importance on the predictive accuracy of the different variables, and specifically of seismic events, for the best schemes of input combinations, a recursive strategy in the trained models was adopted to achieve a multi-step-ahead forecast (i.e., flow forecasting at several lead-times, up to 14 days, with respect to the known flow

values). For a given lead time, the procedure consisted in predicting Q_{t+1} at each time step t, using the time series of the input variables from 1 to t, with the last t_{lag-1} elements of the input Q series ($Q_{t+1-t_{lag}},\ldots,Q_{t-1},Q_t$) replaced by the last t_{lag-1} values of the predicted series.

3. Results and discussion

3.1. One-day ahead flow forecasting exercise: model sensitivities

Figure 7 summarizes the results for the one-day ahead forecast in terms of RMSE (similar patterns have been observed in terms of MAE and r^2) between the standardized observed and predicted flows, under the different combinations of input variables (runs I to X, as defined in Figure 5), model's hyperparameters (Nepochs, BS and NHU) and start date of the training period (ID_train) over the 2015-2018 time series.

The first evident and quite obvious result regards the influence of the antecedent daily flow values on the predictive performances of the developed LSTMs, with the RMSE registering an average threefold increase when Q is excluded from the input parameters (run X). In addition, while indicating an overall modest influence of the hyperparameters on model accuracy, especially for BS and NHU, Figure 7 highlights some apparently controversial results, which, however, can be explained and used to support general warnings about potential pitfalls that may be encountered when implementing black-box models for time series predictions (Kratzert et al., 2018; Cho and Kim, 2022). Indeed, the first panel of the figure shows that the most noticeable differences in the RMSE are a function of ID_train, with minimum errors (median values of about 0.15) identified when the training period starts later than June-July 2016 (i.e., ID_train>300) for most of the runs, thus suggesting the existence of a best training dataset, basically characterized by a significant consistency and similarity with the target dataset. This is a direct consequence of the short records available for implementing the learning process, which explains, besides the observed limited sensitivity of the LSTMs to the hyperparameters, some unexpected results among the different runs. In fact, Figure 7 clearly shows that not only the predictive accuracy does not improve with the number of considered predictors, but even worsens when potentially important factors for the Gran Sasso aquifer, as earthquake features or snow accumulation (e.g., Amoruso et al., 2011, 2013; Adinolfi Falcone et al., 2012), are included into the model. For instance, the most complete runs (I or II) provide slightly lower performances than those reported by run III, which excludes *Sd* from the set of predictors. Such pattern is even more evident when analyzing the results of run V, VI and VII, characterized by using only hydrometeorological variables as input features, but providing the minimum RMSEs. To a certain extent, by emphasizing the role of the learning period, these results then suggest that the length of the available records could be a limiting factor for deriving robust conclusions on the use of LSTMs for predicting spring flows in karst aquifers subject to seismic activity.

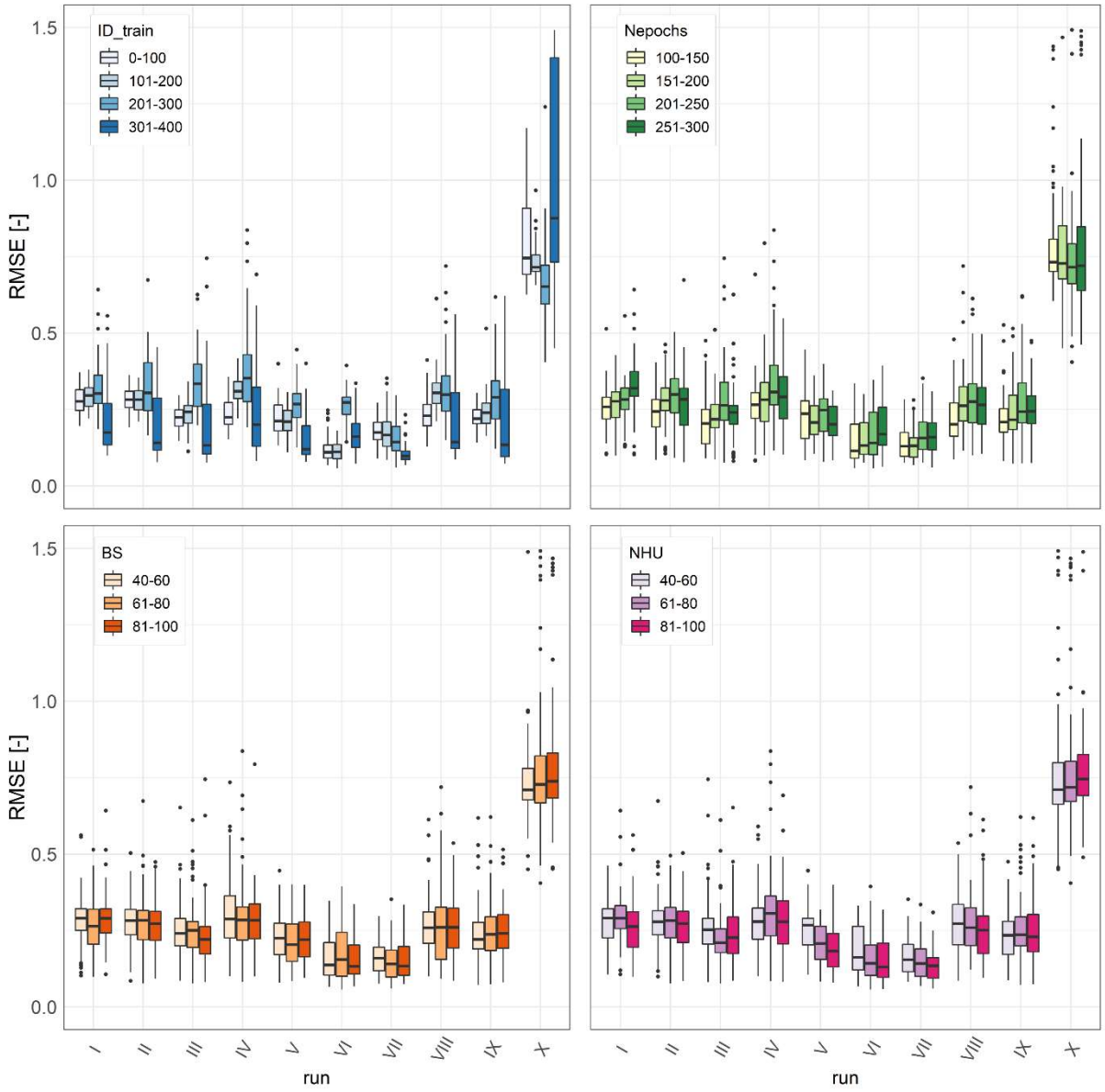


Figure 7. Results of the sensitivity test performed for the one-day ahead flow forecasting exercise: RMSE between the standardized observed and predicted flows as a function of the start date of the learning period (ID_train) and LSTM hyperparameters (number of epochs (Nepochs), batch size (BS) and number of hidden units (NHU)). The run numbers identify the input feature combinations (Figure 4) considered for model training.

3.2. Multiple-step ahead flow forecasting exercise

In consideration of the discussed problems for deriving solid conclusions when relying only on the most recent 2015-2018 time series for both training and test, this section analyzes the results on the test for the 2009-2010 period in order to identify possible reliable insights on the ability of LSTMs in reproducing the flow features in seismic karst aquifers as the one under investigation

To this aim, the LSTM models providing the best performances in the previous exercise for the different runs (Figure 7) were selected to be further applied for one- and multi-day ahead flow predictions over the 2009-2010 period. Since snow depth records started only on late 2009,

after L'Aquila earthquake (Figure 3), here we focus on the runs which do not include *Sd* among the input variables, namely the most complete III and IX, and the simpler VII, relying only on hydrometeorological parameters (Figure 5).

This lag-exercise can be seen as a "stress test" for the predicting ability of the models, in consideration of the importance of antecedent flow discharge values: indeed, when increasing the lead times, such influence is partly reduced, thus allowing to recognize the possible influence of other factors. Furthermore, tests at longer lead times can be even more interesting for practical applications, such as predictions for what-if analysis or missing data filling.

Table 2 summarizes the error metrics for the selected runs at different lead times (t_{lag}), while Figure 8 shows, for exemplificatory purposes, the comparative plots of the 2009-2010 forecasting results obtained from the application of the LSTMs for the feature combinations III and VII. If excluding the short initialization period, the figure depicts a relatively good agreement between the patterns of observed and predicted spring discharges, especially at smaller lead times, thus indicating a fair ability of the LSTM models in resembling the flow features for a test set completely taken out from the training process.

Interestingly, while the differences between forecasts and observation become more evident with increasing t_{lag}, more markedly for run VII (Figure 8b), the plots highlight the skill of the models in well predicting the sharp flow increase induced by the earthquake event occurred on April 2009 (~ day 100 of the time series). Such result suggests that the networks seem to have effectively learned the physical mechanisms leading to sudden flow changes after main seismic events in the karst aquifer, attributed in the literature, from a qualitative perspective, to fracture cleaning phenomena and consequent increase in the bulk hydraulic conductivity (Wang and Manga, 2010; Adinolfi Falcone et al., 2012; Manga and Wang, 2015; Petitta et al. 2018; Valigi et al., 2019). This is quantitatively confirmed by the results shown in Table 2, which indicates similar models' performances at short lead times, with r^2 ranging from 0.98 to 0.83 for t_{lag} between 1 to 4 days for the three configurations.

Table 2. Results for the multiday-ahead forecast exercise on the 2009-2010 test set: error metrics for the three considered runs (run III and IX include seismic input parameters, while run VII relies only on hydrometeorological factors).

Run	**Error metric**	**t_{lag}**						
		1 day	2 days	3 days	4 days	7 days	10 days	14 days
III	RMSE	0.39	0.49	0.60	0.73	1.10	1.39	1.48
	r^2	0.97	0.95	0.93	0.89	0.74	0.51	0.53
	MAE	0.15	0.24	0.36	0.55	1.22	1.94	2.18
VII	RMSE	0.37	0.59	0.78	0.96	2.35	3.22	3.09
	r^2	0.98	0.94	0.90	0.83	0.04	0.001	0.001
	MAE	0.14	0.34	0.61	0.93	5.54	10.34	9.55
IX	RMSE	0.44	0.56	0.66	0.77	1.06	1.32	1.46
	r^2	0.96	0.93	0.90	0.86	0.75	0.63	0.56
	MAE	0.20	0.31	0.44	0.60	1.12	1.73	2.13

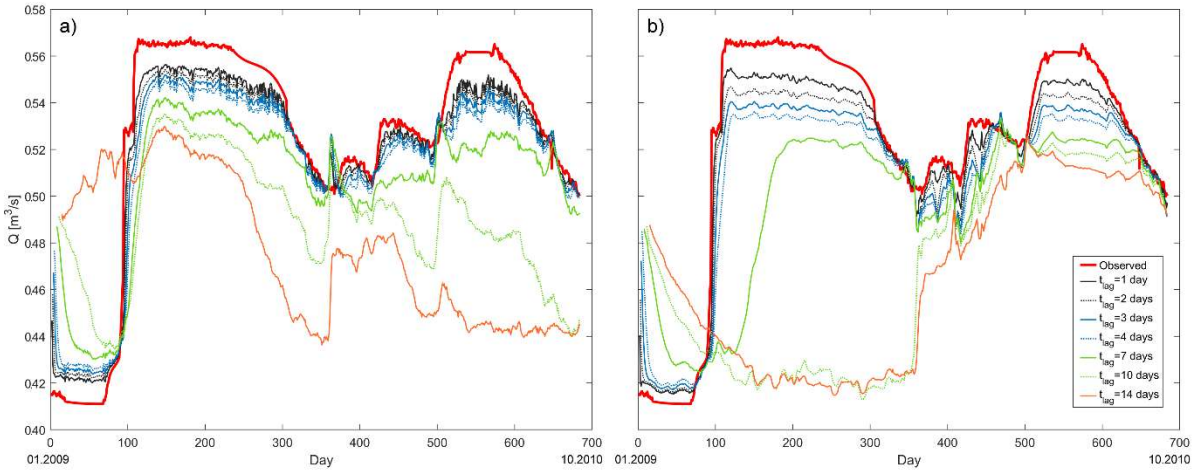


Figure 8. Results for the multiple-step ahead flow forecasting exercise: predicted versus observed flow time series for the 2009-2010 test period at several lead times for: a) model scheme of run III; b) model scheme of run VII.

However, when t_{lag} is further increased, the performances of run VII (the one trained with only hydrometeorological data for the 2015-2018 period) drop significantly ($r^2<0.05$), while a more gradual and limited worsening (with r^2 values of about 0.75-0.55 for $t_{lag} \geq 7$ days) can be observed for the other two combinations, thus proving the positive effect of the seismic parameters on models' accuracy.

It is worth noting that the good predictions provided by run VII at short lead times are mainly a consequence of the tendency of the model to "copy" the antecedent discharge values in the time series: indeed, as shown in the results reported in the previous section, the network heavily relies on antecedent flow rates and thus predictions do not deviate much from observations for small t_{lag}, while, on the other hand, to capture the sudden increase in the flow rate for larger t_{lag}, information on earthquake events certainly improves the quality of the forecasting (runs III and IX).

4. Conclusions

The characterization and investigation of karst aquifers is a complex task, which makes quantitative studies on their governing physical processes quite difficult. In recent years, because of the excellent performance of data-driven approaches in many scientific fields, hydrologists have been starting to implement such algorithms in place of (or combined with) physically based models in their applications.

Given this trend, for the first time, to the best of our knowledge, the present study attempted to investigate the usability and reliability of memory-enhanced neural networks (LSTMs) for modelling the hydrological behavior of karst aquifers exposed to seismic hazard, which exacerbates the system complexity. Indeed, while plenty of literature has been published on the use of machine learning algorithms for purely hydrological applications (Ardabili et al., 2019), no previous studies on the topic have dealt with the inclusion of earthquake impacts into the modelling.

The results presented here, especially in the multiday-ahead forecasting exercise, demonstrated that, albeit machine learning approaches may suffer from limitations arising from insufficient

data for model training (which may lead to incorrect conclusions, as pointed out in the sensitivity analysis discussed in Section 3.1), they have potential for a successful use for flow forecasting in seismic karst areas by virtue of their ability in implicitly learning the complex geophysical mechanisms causing sudden alterations in the hydrological properties of the aquifers after main earthquake events.

Clearly, the key point will be the possibility of relying on representative long hydrometeorological and seismic time series for a proper training of the networks. Therefore, we promote future studies aimed at corroborating our findings in regions experiencing similar geohydrological conditions.

References

Adinolfi Falcone, R., Falgiani, A., Parisse, B., Petitta, M., Spizzico, M., and Tallini, M. (2008). Chemical and isotopic (δ18O‰, δ2H‰, δ13C‰, 222Rn) multi-tracing for groundwater conceptual model of carbonate aquifer (Gran Sasso INFN underground laboratory–central Italy). *Journal of Hydrology*, *357*(3-4), 368-388. doi: 10.1016/j.jhydrol.2008.05.016.

Adinolfi Falcone, R., Carucci, V., Falgiani, A., Manetta, M., Parisse, B., Petitta, M., Rusi, S., Spizzico, M., and Tallini, M. (2012). Changes on groundwater flow and hydrochemistry of the Gran Sasso carbonate aquifer after 2009 L'Aquila earthquake. *Italian Journal of Geosciences*, *131*(3), 459-474. doi: 10.3301/IJG.2011.34.

Amoruso, A., Crescentini, L., Petitta, M., Rusi, S., and Tallini, M. (2011). Impact of the 6 April 2009 L'Aquila earthquake on groundwater flow in the Gran Sasso carbonate aquifer, Central Italy. *Hydrological Processes*, *25*(11), 1754-1764. doi: 10.1002/hyp.7933.

Amoruso, A., Crescentini, L., Petitta, M., and Tallini, M. (2013). Parsimonious recharge/discharge modeling in carbonate fractured aquifers: The groundwater flow in the Gran Sasso aquifer (Central Italy). *Journal of Hydrology*, *476*, 136-146. doi: 10.1016/j.jhydrol.2012.10.026.

An, L., Hao, Y., Yeh, T.C.J., Liu, Y., Liu, W., and Zhang, B. (2020). Simulation of karst spring discharge using a combination of time–frequency analysis methods and long short-term memory neural networks. *Journal of Hydrology*, *589*, 125320. doi: 10.1016/j.jhydrol.2020.125320.

Ardabili, S., Mosavi, A., Dehghani, M., and Várkonyi-Kóczy, A.R. (2019). Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In: Várkonyi-Kóczy, A. (Eds.) Engineering for Sustainable Future. INTER-ACADEMIA 2019. Lecture Notes in Networks and Systems, Vol 101. Springer, Cham, Switzerland. doi: 10.1007/978-3-030-36841-8_5.

Bagh, S., Chiaraluce, L., De Gori, P., Moretti, M., Govoni, A., Chiarabba, C., Di Bartolomeo, P., and Romanelli, M. (2007). Background seismicity in the Central Apennines of Italy: The Abruzzo region case study. *Tectonophysics*, *444*(1-4), 80-92. doi: 10.1016/j.tecto.2007.08.009.

Barberio, M. D., Barbieri, M., Billi, A., Doglioni, C., and Petitta, M. (2017). Hydrogeochemical changes before and during the 2016 Amatrice-Norcia seismic sequence (central Italy). *Scientific Reports*, *7*(1), 1-12. doi: 10.1038/s41598-017-11990-8.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on neural networks, 5*(2), 157–166, 1994.

Birk, S., Liedl, R., and Sauter, M. (2006). Karst spring responses examined by process-based modeling. *Groundwater*, *44*(6), 832-836. doi: 10.1111/j.1745-6584.2006.00175.x.

Bonacci, O. (2004). Hazards caused by natural and anthropogenic changes of catchment area in karst. *Natural Hazards and Earth System Sciences*, *4*(5/6), 655-661. SRef-ID: 1684-9981/nhess/2004-4-655.

Boni C., Bono P., Capelli G. (1986). Schema idrogeologico dell'Italia centrale. *Memorie della Società Geologica Italiana*, *35*, 991-1012, 2 Tav., Roma.

Brodsky, E.E., Roeloffs, E., Woodcock, D., Gall, I., and Manga, M. (2003). A mechanism for sustained groundwater pressure changes induced by distant earthquakes. *Journal of Geophysical Research: Solid Earth*, *108*(B8). doi: 10.1029/2002JB002321.

Casini, S., Martino, S., Petitta, M., and Prestininzi, A. (2006). A physical analogue model to analyse interactions between tensile stresses and dissolution in carbonate slopes. *Hydrogeology Journal*, *14*(8), 1387-1402. doi: 10.1029/2002JB002321.

Cheng, M., Fang, F., Kinouchi, T., Navon, I.M., and Pain, C.C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, *590*, 125376. doi: 10.1016/j.jhydrol.2020.125376.

Cho, K., and Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *Journal of Hydrology*, *605*, 127297. doi: 10.1016/j.jhydrol.2021.127297.

Claesson, L., Skelton, A., Graham, C., Dietl, C., Mörth, M., Torssander, P., and Kockum, I. (2004). Hydrogeochemical changes before and after a major earthquake. *Geology*, *32*(8), 641-644. doi: 10.1130/G20542.1.

Curci, G., Guijarro, J.A., Di Antonio, L., Di Bacco, M., Di Lena, B. and Scorzini, A.R. (2021). Building a local climate reference dataset: application to the Abruzzo region (Central Italy), 1930-2019. *International Journal of Climatology, 41*(8), 4414-4436. doi: 10.1002/joc.7081.

De Luca, G., Scarpa, R., Filippi, L., Gorini, A., Marcucci, S., Marsan, P., Milana, G., and Zambonelli, E. (2000). A detailed analysis of two seismic sequences in Abruzzo, Central Apennines, Italy. *Journal of Seismology*, *4(1)*, 1-21. doi: 10.1023/A:1009870327083.

De Luca, G., Cattaneo, M., Monachesi, G., and Amato, A. (2009). Seismicity in central and northern Apennines integrating the Italian national and regional networks. *Tectonophysics*, *476(1-2)*, 121-135. doi: 10.1016/j.tecto.2008.11.032.

De Luca, G., Di Carlo, G., and Tallini, M. (2018). A record of changes in the Gran Sasso groundwater before, during and after the 2016 Amatrice earthquake, central Italy. *Scientific Reports*, *8*(1), 1-16. doi: 10.1038/s41598-018-34444-1.

Dikshit, A., Pradhan, B., and Alamri, A.M. (2020). Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Research, 100*, 290-301. doi: 10.1016/j.gr.2020.08.007.

Di Matteo, L., Dragoni, W., Azzaro, S., Pauselli, C., Porreca, M., Bellina, G., and Cardaci, W. (2020). Effects of earthquakes on the discharge of groundwater systems: The case of the 2016 seismic sequence in the Central Apennines, Italy. *Journal of Hydrology*, *583*, 124509. doi: 10.1016/j.jhydrol.2019.124509.

Ding, H., Zhang, X., Chu, X., and Wu, Q. (2020). Simulation of groundwater dynamic response to hydrological factors in karst aquifer system. *Journal of Hydrology*, *587*, 124995. doi: 10.1016/j.jhydrol.2020.124995.

Duran, L., Massei, N., Lecoq, N., Fournier, M., and Labat, D. (2020). Analyzing multi-scale hydrodynamic processes in karst with a coupled conceptual modeling and signal decomposition approach. *Journal of Hydrology*, *583*, 124625. doi: 10.1016/j.jhydrol.2020.124625.

Esposito, E., Pece, R., Porfido, S., and Tranfaglia, G. (2001). Hydrological anomalies connected to earthquakes in southern Apennines (Italy). *Natural Hazards and Earth System Sciences*, *1*(3), 137-144. doi: 10.5194/nhess-1-137-2001.

Frepoli, A., Cimini, G. B., De Gori, P., De Luca, G., Marchetti, A., Monna, S., Montuori, C. and Pagliuca, N.M. (2017). Seismic sequences and swarms in the Latium-Abruzzo-Molise Apennines (central Italy): New observations and analysis from a dense monitoring of the recent activity. *Tectonophysics*, *712*, 312-329. doi: 10.1016/j.tecto.2017.05.026.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, *52*(3), 218-242. doi: 10.1002/2013RG000443.

He, Z., Wen, X., Liu, H., and Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, *509*, 379-386. doi: 10.1016/j.jhydrol.2013.11.054.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735.

Hu, C., Hao, Y., Yeh, T.C.J., Pang, B., and Wu, Z. (2008). Simulation of spring flows from a karst aquifer with an artificial neural network. *Hydrological Processes, 22*(5), 596-604. doi: 10.1002/hyp.6625.

Hutter, F. Hoos, H., and Leyton-Brown, K. (2014). An Efficient Approach for Assessing Hyperparameter Importance. In: Proceedings of the 31st International Conference on International Conference on Machine Learning – Vol. 32, ICML'14, pag. I-754-I-762.

Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences, 22*(11), 6005-6022. doi: 10.5194/hess-22-6005-2018.

Le, X.H., Ho, H.V., Lee, G., and Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*, *11*(7), 1387. doi: 10.3390/w11071387.

Li, G., Goldscheider, N., and Field, M.S. (2016). Modeling karst spring hydrograph recession based on head drop at sinkholes. *Journal of Hydrology*, *542*, 820-827. doi: 10.1016/j.jhydrol.2016.09.052.

Manga, M., Brodsky, E.E., and Boone, M. (2003). Response of streamflow to multiple earthquakes. *Geophysical Research Letters*, *30*(5). doi: 10.1016/j.jhydrol.2016.09.052.

Manga, M., Beresnev, I., Brodsky, E.E., Elkhoury, J.E., Elsworth, D., Ingebritsen, S.E., .Mays, D.C. and Wang, C.Y. (2012). Changes in permeability caused by transient stresses: Field observations, experiments, and mechanisms. *Reviews of Geophysics*, *50(2)*, 1-24. doi: 10.1029/2011RG000382.

Manga, M. and Wang, C.Y. (2015). Earthquake Hydrology. In: Schubert, G. (Ed.), *Treatise on Geophysics*, 2nd Edition, Vol 4. Oxford: Elsevier, 305-328. doi: 10.1016/B978-0-444-53802-4.00082-8

Mastrorillo, L., Saroli, M., Viaroli, S., Banzato, F., Valigi, D., and Petitta, M. (2020). Sustained post-seismic effects on groundwater flow in fractured carbonate aquifers in Central Italy. *Hydrological Processes*, *34*(5), 1167-1181. doi: 10.1002/hyp.13662.

Mehdipour Ghazi, M., Nielsen, M., Pai, A., Modat, M., Cardoso, M. J., Ourselin, S., and Sørensen, L. (2019). On the Initialization of Long Short-Term Memory Networks. In: Gedeon, T., Wong, K., Lee, M. (Eds.) *Neural Information Processing. ICONIP 2019. Lecture Notes in Computer Science*, 11953. Springer, Cham, Switzerland. doi: 10.1007/978-3-030-36708-4_23.

Mohr, C.H., Manga, M., Wang, C.Y., and Korup, O. (2017). Regional changes in streamflow after a megathrust earthquake. *Earth and Planetary Science Letters*, *458*, 418-428. doi: 10.1016/j.epsl.2016.11.013.

Montgomery, D.R., and Manga, M. (*2003).* Streamflow and water well responses to earthquakes. *Science*, *300,* 2047–2049. doi: 10.1126/science.1082980.

Nakagawa, K., Yu, Z.Q., Berndtsson, R., and Hosono, T. (2020). Temporal characteristics of groundwater chemistry affected by the 2016 Kumamoto earthquake using self-organizing maps. *Journal of Hydrology*, *582*, 124519. doi: 10.1016/j.jhydrol.2019.124519.

Onda, S., Sano, Y., Takahata, N., Kagoshima, T., Miyajima, T., Shibata, T., Pinti, D.L., Lan, T., Kim, N.K., Kusakabe, M., and Nishio, Y. (2018). Groundwater oxygen isotope anomaly before the M6.

6 Tottori earthquake in Southwest Japan. *Scientific Reports*, *8*(1), 1-7. doi: 10.1038/s41598-018-23303-8.

Petitta, M., and Tallini, M. (2002). Idrodinamica sotterranea del massiccio del Gran Sasso (Abruzzo); muove indagini idrologiche, idrogeologiche e idrochimiche (1994-2001). *Bollettino della Società geologica italiana*, *121*(3), 343-363.

Petitta, M., Mastrorillo, L., Preziosi, E., Banzato, F., Barberio, M.D., Billi, A., Cambi, C., De Luca, G. … and Doglioni, C. (2018). Water-table and discharge changes associated with the 2016–2017 seismic sequence in central Italy: hydrogeological data and a conceptual model for fractured carbonate aquifers. *Hydrogeology Journal*, *26*(4), 1009-1026. doi: 10.1007/s10040-017-1717-7.

Reimers, N., and Gurevych, I. (2017). Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. arXiv preprint, arXiv:1707.06799.

Roeloffs, E.A. (1998). Persistent water level changes in a well near Parkfield, California, due to local and distant earthquakes. *Journal of Geophysical Research: Solid Earth, 103*(B1), 869-889. doi: 10.1029/97JB02335.

Rojstaczer, S., Wolf, S., and Michel, R. (1995). Permeability enhancement in the shallow crust as a cause of earthquake-induced hydrological changes. *Nature*, *373*(6511), 237-239. doi: 10.1038/373237a0.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1985). Learning internal representations by error propagation. No. ICS-8506, California Univ. San Diego, La Jolla Institute for Cognitive Science.

Scozzafava, M., and Tallini, M. (2001). Net infiltration in the Gran Sasso Massif of central Italy using the Thornthwaite water budget and curve-number method. *Hydrogeology Journal, 9*(5), 461-475. doi: 10.1007/s100400100151.

Shi, Z., Wang, G., Manga, M., and Wang, C.Y. (2015). Mechanism of co-seismic water level change following four great earthquakes–insights from co-seismic responses throughout the Chinese mainland. *Earth and Planetary Science Letters*, *430*, 66-74. doi: 10.1016/j.epsl.2015.08.012.

Sivakumar, B., Jayawardena, A.W., and Fernando, T.M.K.G. (2002). River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *Journal of Hydrology*, *265*(1-4), 225-245. doi: 10.1016/S0022-1694(02)00112-9.

Valigi, D., Mastrorillo, L., Cardellini, C., Checcucci, R., Di Matteo, L., Frondini, F., Mirabella, F., Viaroli, S., and Vispi, I. (2019). Springs discharge variations induced by strong earthquakes: The Mw 6.5 Norcia event (Italy, October 30th 2016). *Rendiconti Online Della Società Geologica Italiana*, *47*, 141-146. doi: 10.3301/ROL.2019.25.

Vittecoq, B., Fortin, J., Maury, J., and Violette, S. (2020). Earthquakes and extreme rainfall induce long term permeability enhancement of volcanic island hydrogeological systems. *Scientific Reports*, *10(1)*, 1-13. doi: 10.1038/s41598-020-76954-x.

Wang, C.Y., C.H. Wang, and Kuo, C.H. (2004). Temporal change in groundwater level following the 1999 (Mw=7.5) Chi-Chi earthquake, Taiwan. *Geofluids, 4,* 210–220. doi: 10.1111/j.1468-8123.2004.00082.x.

Wang, C.Y., and Chia, Y. (2008). Mechanism of water level changes during earthquakes: Near field versus intermediate field. *Geophysical Research Letters*, *35*(12). doi: 10.1111/j.1468-8123.2004.00082.x.

Wang, C.Y., and Manga, M. (2010). Hydrologic responses to earthquakes and a general metric. *Geofluids*, *10*(1-2), 206-216. doi: 10.1111/j.1468-8123.2009.00270.x.

Wu, C.L., Chau, K.W., and Li, Y.S. (2009). Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resources Research*, *45*(8). doi: 10.1029/2007WR006737.

Zhang, X., Peng, Y., Zhang, C., and Wang, B. (2015). Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *Journal of Hydrology*, *530*, 137-152. doi: 10.1016/j.jhydrol.2015.09.047.

Zimmermann, H.G., Tietz, C., and Grothmann, R. (2012). Forecasting with recurrent neural networks: 12 tricks. In: Montavon, G., Orr, G.B., Müller, K.R. (Eds.) *Neural Networks: Tricks of the Trade*, 687-707. Springer, Berlin, Heidelberg, Germany. ISBN: 978-3-642-35289-8.

Zoghbi, C., and Basha, H. (2019). Simplified physically based models for free-surface flow in karst systems. *Journal of Hydrology*, *578*, 124040. doi: 10.1016/j.jhydrol.2019.124040.