



# Physics-informed machine learning for understanding rock moisture dynamics in a sandstone cave

5 Kai-Gao Ouyang<sup>1</sup>, Xiao-Wei Jiang<sup>1,2\*</sup>, Gang Mei<sup>3</sup>, Hong-Bin Yan<sup>4</sup>,

Ran Niu<sup>1</sup>, Li Wan<sup>1</sup>, Yijian Zeng<sup>5</sup>

1. MWR Key Laboratory of Groundwater Conservation, China University of Geosciences, Beijing, China

2. MNR Key Laboratory of Shallow Geothermal Energy, China University of Geosciences, Beijing, China

3. School of Engineering and Technology, China University of Geosciences, Beijing, China

10 4. Yungang Research Institute, Datong, Shanxi, China

5. Department of Water Resources, ITC Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, the Netherlands

\*Correspondence: Xiao-Wei Jiang (jxw@cugb.edu.cn)

## 15 Abstract

Rock moisture, which is considered as a hidden component of the terrestrial hydrological cycle, has received little attention. In this study, the frequency-domain reflectometry (FDR) is used to obtain fluctuating rock water content in a sandstone cave of the Yungang Grottoes, China. We identified two major cycles of rock moisture addition and depletion, one in the summer and the other in the winter. By using the LSTM (Long Short-Term Memory) network and the SHAP (SHapley Additive exPlanations) method, relative humidity, air temperature and wall temperature are found to have contributions to rock moisture in the summer. By using vapor concentration and the difference between dew point temperature and wall temperature as two input variables of the LSTM network, the predicted rock water content has a very good agreement with the measured rock water content, with the Nash–Sutcliffe efficiency coefficient (NSE) being as high as 0.978. Because the two new input variables are factors directly controlling vapor condensation, they provide informative priors to the deep learning model and improved prediction performance. After identifying the causal factors of rock water content fluctuations, we also identified the mechanism controlling the multi diurnal vapor condensation. The increased vapor concentration

20

25



30 accompanying a precipitation event leads to transport of water vapor into rock pores, which is subsequently adsorbed onto the surface of rock pores and then condensed into liquid water. With the aid of the deep learning model, this study increases understanding of sources of water in caves, which would contribute to future strategies of alleviating weathering in caves.

## 1 Introduction

35 Water movement in the unsaturated zone is a fundamental component of the hydrologic cycle regulating the atmosphere, the hydrosphere and the lithosphere (Arora et al., 2019; Brubaker and Entekhabi, 1996; Lu and Likos, 2004; Tindall et al., 1998). Although there are abundant studies on water movement in various scales of unsaturated soils (Larson et al., 2022; Schoups et al., 2005; Vereecken et al., 2014; Vinnikov et al., 1996; Yu et al., 2016), much less attention has been paid to water in unsaturated rocks. In a recent study, Rempe and Dietrich (2018) defined water stored in unsaturated rocks as rock moisture and pointed out that rock moisture is a hidden component of the 40 terrestrial hydrologic cycle critical to ecosystems and weathering processes. Due to the ubiquitous occurrence of precipitation infiltration through unsaturated rocks, infiltrating precipitation was found to be the main source of rock moisture (Rempe et al., 2018; Sass, 2005). In fact, as early as in the fourth century BC, Aristotle (384-322 BC) hypothesized that atmospheric water vapor could penetrate into rocks in caves with low temperature and condense into liquid water (after Meinzer, 1934). Due to occurrence of hidden water in the form of rock moisture, many stone 45 heritages inside caves have suffered from weathering (Auler and Smart, 2004; Camuffo, 1998; de Freitas et al., 2006; Guerriera et al., 2019; Linan et al., 2021). However, up to now, there is no observations of rock moisture in caves, which hampers a comprehensive understanding of the source and control factors of rock moisture.

By using such techniques as downhole neutron probe (Rempe and Dietrich, 2018), time domain reflectometry (TDR) (Salve et al., 2012) and nuclear magnetic resonance (NMR) (Schmidt and Rempe, 2020), the responses of 50 rock moisture to precipitation were identified in some small catchments. However, these devices, which are usually long in length (the lengths of TDR and neutron probe are larger than 30 mm) or large in diameter (the diameter of NMR is around 70 mm), are not suitable to be used in stone heritages. The frequency domain reflectometry (FDR), which has the advantage of small in volume (the length is less than 60 mm), has been widely used to characterize the temporal variability of soil moisture (Irmak and Irmak, 2005; Li et al., 2020; Xie et al., 2021; Zhang et al., 55 2019). Note that both TDR and FDR measure the dielectric constant of porous media, and FDR is less influenced by temperature (Zhu et al., 2019). In this study, we attempt to use the FDR for monitoring rock moisture in the field



for the first time and identify the atmospheric conditions controlling rock moisture in caves not exposed to rain.

Establishing the cause-and-effect relationship between rock moisture and various atmospheric conditions is a feasible approach to identify the source of rock moisture in caves and reveal mechanisms controlling rock moisture fluctuations. Machine learning has the ability to acquire knowledge and establish the complicated nonlinear relationship between variables in a vast domain (Chen et al., 2019a; Jumin et al., 2020). Although machine learning models have the ability of high accuracy prediction, they are notorious for being a black-box model. Lundberg and Lee (2017) proposed the SHAP (SHapley Additive exPlanations) values as a unified measure of feature importance, which led to a combination of accuracy and interpretability of predictions by machine learning models. In almost all applications of machine learning in the field of hydrology, the directly measured meteorological factors like precipitation, temperature, radiation, humidity and wind speed are used as input variables (e.g., Barzegar et al., 2017; Fang et al., 2017; Gao et al., 2020; Lees et al., 2021; Liu et al., 2022; Xiang et al., 2019; Zhao et al., 2022). In fact, using prior knowledge stemming from physical or mathematical understanding as model inputs could improve the performance of a machine learning algorithm, which is called physics-informed learning (Karniadakis et al., 2021).

In this study, the classic Long Short-Term Memory (LSTM) network, which is deep learning model, is combined with the SHAP values to predict rock moisture and evaluate the relative importance of four directly measured variables (precipitation, relative humidity, air temperature and wall temperature). After excluding the possible control by precipitation, based on the physics controlling vapor condensation, two new variables derived from relative humidity, air temperature and wall temperature are used as inputs of the LSTM network, which not only leads to improved prediction performance, but also improve understanding of source of water in caves.

## 2 Study site and field monitoring

### 2.1. Study site

The Yungang Grottoes (40°07' N, 113°08' E), which are located in Datong, Shanxi Province, China, were declared World Heritage Site by the UNESCO in 2001 (<http://whc.unesco.org/en/list/1039>). According to meteorological data in recent 20 years in the Datong city (data from The China Meteorological Data Service Center, <http://data.cma.cn/en>), the site has a semi-arid climate, with an annual average precipitation of 393 mm and an annual average pan evaporation of 1243 mm. The precipitation in the rainy season from June to September accounts for 73% of the annual precipitation. The annual average temperature is 8 °C, the average temperature in summer is 20.3 °C, and the average temperature in winter is -8.2 °C.



Most statues in the Yungang Grottoes were carved in sandstone caves (Fig. 1a). In the summer, water droplets with planar distribution can be occasionally observed on the walls of some caves (Fig. 1b). Although no water droplets occur in other sandstone walls, by absorbing water, the high rock moisture leads to slight changes in the color of some walls. These different forms of water are responsible for weathering of the statues, however, the sources of these forms of water remains controversial. Previous studies suggested that the possible sources of water in caves include infiltrating precipitation through the overlying thick unsaturated zone (Wang et al., 2012) and condensation of water vapor onto walls (Cao et al., 2005; Huang, 2008; Huang, 2010).

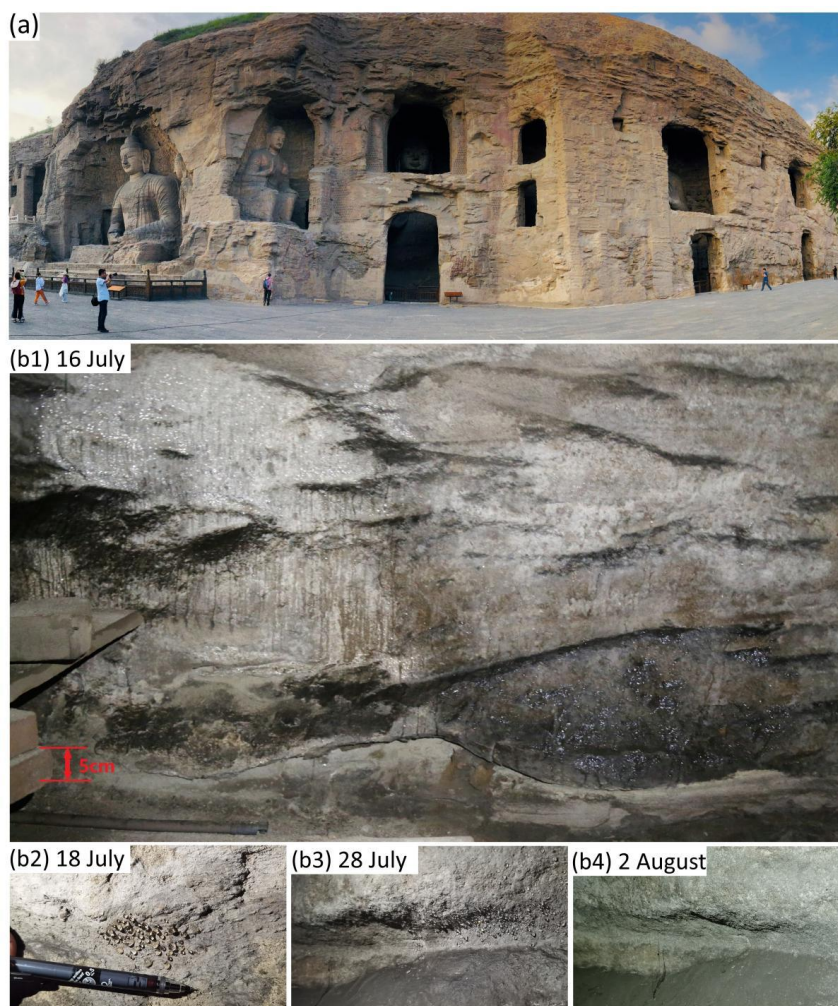


Figure 1. (a) Some caves and statues in the Yungang Grottoes; (b) The occurrence (b1-b3) and disappearance (b4) of water droplets in Cave #5 of the Yungang Grottoes in the summer of 2021.

95



## 2.2. Monitoring of rock moisture and atmospheric conditions

To monitor variations of rock moisture in the shallow part of a cave wall, a FDR-based ECH2O EC-5 sensor (produced by DECAGON, USA) was installed at 3-8 cm inside the north wall of Cave #9 (Fig. 2). Due to the difficulty of calibrating the actual water content in the field, the apparent rock water content measured by the sensor is directly used in the current study. As reported in previous experimental studies (Mollo and Greco, 2011; Sakaki and Rajaram, 2005), there is a good linear relationship between actual rock water content and rock moisture transformed from dielectric constant. Because the purpose of the current study to identify the source of rock moisture addition by establishing the relationship between rock moisture and atmospheric conditions, the relative changes of rock water content is more important than the actual value of rock water content. Therefore, direct use of apparent rock water content does not influence the results of our study, and the apparent rock water content is directly called rock water content.

Air temperature ( $T_a$ ) and air relative humidity ( $RH$ ) are simultaneously monitored near the monitoring site of rock moisture (Fig. 2). The wall temperature is also monitored to analyze whether the wall meets the condition for condensation. Moreover, hourly precipitation is available from a meteorological station outside the cave.

110

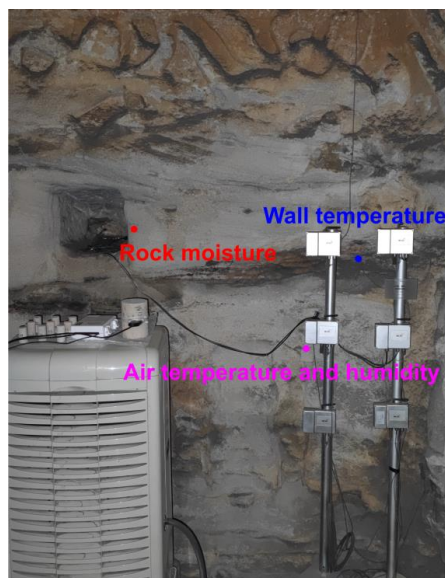


Figure 2. A photo showing the arrangements of sensors for rock moisture, wall temperature, air temperature and relative humidity in the north wall of Cave #9



### 3 Methods

#### 3.1 Long short-term memory (LSTM) network

115 The LSTM network is an improved variant of the conventional Recurrent Neural Network (RNN) and has  
 the same fundamental framework as the conventional RNN. Therefore, we first give a brief introduction to the  
 conventional RNN. As shown in Fig. 3, in the RNN, there is an input layer, a hidden layer and an output layer.  $x_t$  is  
 the input vector at time step  $t$ ,  $h_t$  is the hidden state at time step  $t$  determined by both the input vector  $x_t$  at time step  
 $t$  and the hidden state ( $h_{t-1}$ ) at time step  $t-1$  (Zhao et al., 2017), and  $op_t$  is the output of the RNN at time step  $t$ .  
 120 Mathematically, the relationship between the three layers can be written as:

$$h_t = \tanh(Ux_t + Wh_{t-1}) \quad (1)$$

$$op_t = Vh_t \quad (2)$$

where  $\tanh$  is the activation function which means the hyperbolic tangent performs nonlinear transformations of  
 the inputs,  $U$ ,  $W$  and  $V$  are the network weight matrices for input-to-hidden, hidden-to-hidden and hidden-to-output  
 125 connections, respectively,  $b_o$  and  $b_h$  are bias vectors. In this study, the open-source framework TensorFlow (version  
 1.14.0) written in Python 3.7.6 is used to build and train the LSTM model.

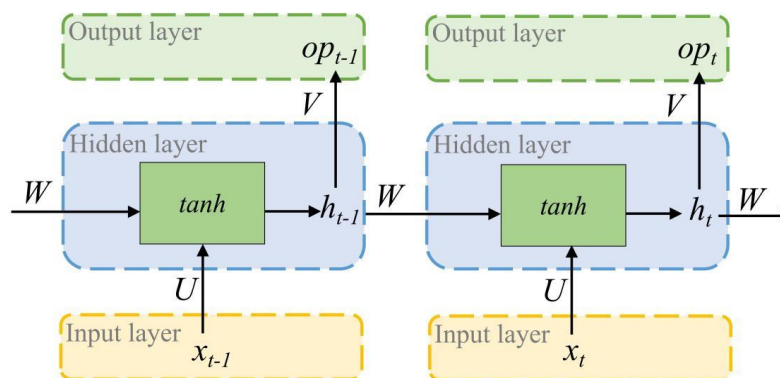


Figure 3. The structure of the Recurrent Neural Network (RNN) (Modified from Hopfield, 1982).

130 When the time sequence is too long, the RNN is not sufficiently efficient to learn because of vanishing  
 gradient and exploding gradient problems (Bengio, 1994; Hochreiter and Schmidhuber, 1997). To solve such  
 problems, Hochreiter and Schmidhuber (1997) proposed an improved variant of RNN whose hidden layer can  
 capture the correlation within time series in both short and long term. This improved variant was named the LSTM



network, and its hidden layer was called the LSTM cell. As shown in Fig. 4, a LSTM cell has three gates, including  
 135 a forget gate, an input gate and an output gate, to maintain and adjust its cell state  $c_t$  (also called memory) and  
 hidden state  $h_t$ . The forget gate determines what information should be discarded from the cell state, the input gate  
 specifies what new information need be stored in the cell state, and the output gate determines what information  
 from the cell state  $c_t$  should be passed to the next step (Gao et al., 2020; Fischer and Krauss, 2018; Lipton et al.,  
 2015). Note that in the input gate, the cell state is determined simultaneously by the the sigmoid activation function  
 140 which determines the weights of the information and the  $\tanh$  activation function which eliminates the bias of the  
 network, the latter of which is termed abstract cell state.

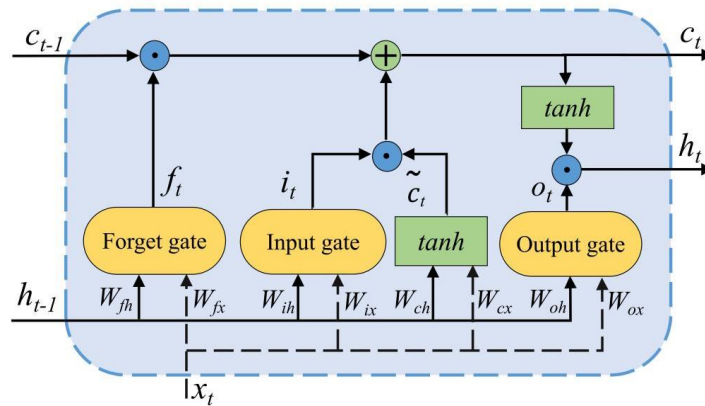


Figure 4. The structure of the LSTM cell (Modified from Hochreiter and Schmidhuber, 1997).

145 The formulas of the three gates, abstract cell state, cell state and hidden state in the LSTM cell are shown as  
 follows:

$$\text{(forget gate)} \quad f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (3)$$

$$\text{(input gate)} \quad i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (4)$$

$$\text{(output gate)} \quad o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (5)$$

150 (abstract cell state)  $\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (6)$

$$\text{(cell state)} \quad c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (7)$$

$$\text{(hidden state)} \quad h_t = \tanh(c_t) \odot o_t \quad (8)$$

where  $i_t, f_t$  and  $o_t$  are the vectors of input, forget and output gates at time step  $t$ , respectively, all of which have the



155 same sizes as  $c_i$  and  $h_i$ ,  $\sigma$  is the logistic sigmoidal activation function,  $\tilde{c}_i$  is the vector of abstract cell state at time step  $t$ ,  $\odot$  is element wise multiplication of two vectors. Similar to RNN,  $W_{ix}$ ,  $W_{ih}$ ,  $W_{ox}$ ,  $W_{oh}$ ,  $W_{cx}$  and  $W_{ch}$  are the matrices for different connections in the network,  $b_i$ ,  $b_o$  and  $b_c$  are bias vectors. In this study, the open-source framework TensorFlow (version 1.14.0) written in Python 3.7.6 is used to build and train the LSTM model.

160 Although purely data-driven models may have a high accuracy of prediction, most machine learning models cannot extract interpretable information and knowledge from this data deluge (Karniadakis et al., 2021). The performance of machine learning models could be improved through providing physical prior knowledge. In this study, we simply integrate physics controlling vapor condensation into the input variables to improve the performance of the LSTM model.

### 3.2 Model interpretation and evaluation

165 To interpret the performance of a machine model, Lundberg and Lee (2017) proposed the SHAP (SHapley Additive exPlanations) explanation method, which is based on game theory (Štrumbelj and Kononenko, 2014). The Shapely value of every input variable represents the contribution of it on the prediction, and the importance of each input variable is clarified by comparing model performances with and without it. The formula for calculation of the Shapely value is

$$\phi_i = \sum_{S \subseteq F, i \in S} \frac{|S|! (F - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (9)$$

170 where  $\phi_i$  is the contribution of variable  $i$ ;  $F$  is the set of all input variables;  $v(S \cup \{i\})$  is the result of a model trained with the variable  $i$ , and  $v(S)$  is the result without the variable  $i$ , so the difference between them represents the effect of feature  $i$  on the model prediction. This method requires retraining the model on all feature subsets  $S \subseteq F$  (Shapley, 1953).

175 To assess the accuracy of prediction by the LSTM network, we use the statistical metrics of Nash–Sutcliffe efficiency coefficient (NSE), mean absolute error (MAE) and root mean squared error (RMSE), all of which are widely used in the literature. NSE is the ratio of the sum of the squares of the regression to the total sum of the squares, which reflects the linear fit between the predictions and observations. The closer the value is to 1, the better the linear fit. The expression of NSE is

$$NSE = 1 - \frac{\sum_{i=1}^N (y_{Pred} - \bar{y})^2}{\sum_{i=1}^N (\bar{y} - y)^2} \quad (10)$$

180 where  $N$  is the number of data, and  $y_{Pred}$ ,  $y$ , and  $\bar{y}$  are the predicted, observed, and mean observed value, respectively.





MAE is the mean of the distance between the predicted and the observations, whereas RMSE is the square root of the mean of the square of the deviation between the predicted and the observations. The expressions of MAE and RMSE are

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{Pred} - y| \quad (11)$$

185

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (|y_{Pred} - y|)^2} \quad (12)$$

## 4 Results and discussion

### 4.1 The seasonal variation of rock moisture and atmospheric conditions

In the north wall of Cave #9, although there is no obvious occurrence of liquid water throughout the year, there is a clear trend of seasonal variation in rock water content (Fig. 5). From April to May, the rock water content is relatively stable and maintains at around  $0.013 \text{ cm}^3/\text{cm}^3$ . From June to September, which correspond to the rainy season with high relative humidity and high air temperature, there is a cycle of significant addition and depletion of rock moisture. From October to December, there is a trend of gradual decrease in rock water content. The cycles of precipitation, relative humidity, air temperature and wall temperature from spring to early winter have quite similar trends as the cycle of rock water content, indicating that they are possible environmental conditions leading to the fluctuating rock water content.

195

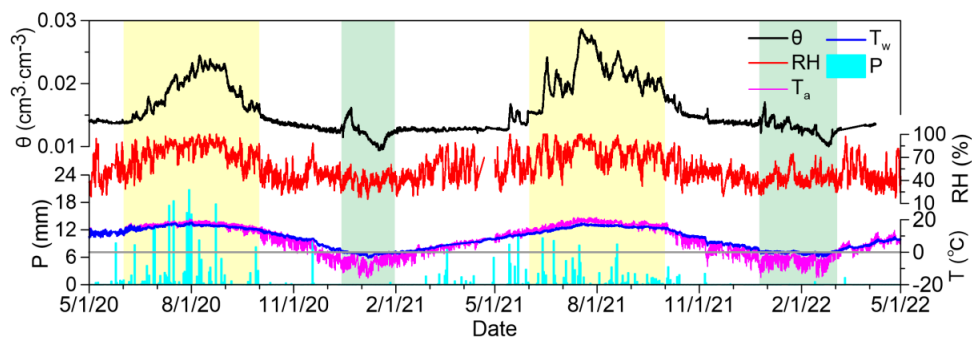


Figure 5. The temporal variations of rock water content ( $\theta$ ), air relative humidity ( $RH$ ), air temperature ( $T_a$ ), precipitation ( $P$ ) and wall temperature ( $T_w$ ). The two periods in yellow correspond to the summer period with high temperature and high humidity, whereas the two periods in green correspond to the fluctuation of rock moisture caused by freezing-thawing.

200



In the summer of 2021, the rock water content has a sharp increase since 9 July, reaches a maximum value equaling  $0.029 \text{ cm}^3/\text{cm}^3$  on 17 July, and is maintained at relatively high values until 28 July. The high rock water content indicates that there are atmospheric conditions responsible for water infiltration or condensation of water vapor. Note that this period corresponds to the period with occurrence of water droplets in Cave #5 as shown in Fig. 1b1-b3. Although there is no water droplet in Cave #9, the color of the sandstone changes slightly, indicating that a slight change in the color of sandstone is a result of the fluctuating rock water content. Therefore, we believe that the rock water content measured by FDR reflects the actual change of water content in the rock.

From mid December in 2020 to the end of January in 2021, and from late December in 2021 to the end of February in 2022, there are also significant fluctuations of rock water content (Fig. 5). This pattern of fluctuating rock water content is a direct of freezing-thawing, which can be confirmed by the negative wall temperature. At the beginning of the freezing-thawing cycle, there is a trend of increasing rock water content around the sensor due to freezing-induced liquid water migration towards the wall surface with the lowest temperature. By the end of the freezing period, the rock water content reaches a minimum value of the year because most liquid water has been transformed into ice. In the two years, the minimum liquid water content is  $0.009 \text{ cm}^3/\text{cm}^3$  (on 16 January 2021) and  $0.010 \text{ cm}^3/\text{cm}^3$  (on 25 February 2022), respectively. In the thawing stage, there is a trend of increasing liquid water content.

The pattern of freezing-thawing-induced rock water content fluctuations is similar to that of freezing-thawing-induced soil water content fluctuations (Deprez et al., 2020; Matsuoka and Murton, 2008; Sun and Scherer, 2010; Xie et al., 2021; Yu et al., 2018), demonstrating that the FDR technique is very sensitive to liquid water content in sandstone and is suitable to measure rock moisture. The fluctuating rock water content during the freezing-thawing cycle also has implications for understanding weathering processes. The decreasing low liquid water content induced by freezing indicates that rock moisture in spring and autumn belongs to movable bound water that could be responsible for chemical weathering. Moreover, the freezing of accumulated liquid water near the wall surface might cause physical weathering.

#### 4.2 The performance of LSTM model with two different schemes

In the rainy season, as we pointed out in 4.1, precipitation, relative humidity, air temperature and wall temperature have quite similar trends of seasonal variation as apparent rock moisture. Apparently, they are all possible factors determining the fluctuating apparent rock moisture. Therefore, in 4.2.1, we first use all of them as input variables (scheme #1) of the LSTM model to predict rock water content, and use the SHAP values to evaluate



the contribution of each input variable. After excluding precipitation whose mean |SHAP value| equals 0, in 4.2.2, we use two new parameters (vapor concentration, dew point temperature minus wall temperature) calculated from relative humidity, air temperature and wall temperature as input variables (scheme #2) of the LSTM model to predict apparent rock moisture.

235 Deep learning models require a large amount of data for training, as well as data sets with a longtime span to ensure the mastery of complete data features. Because the period from 1 June to 1 October has the most significant trends of rock moisture addition and depletion, the hourly data during this period in the year 2020 are used to construct the training set, whereas the hourly data in the year 2021 are used to construct the test set.

#### 4.2.1 The predicted results based on directly monitored variables

240 By using relative humidity, air temperature, precipitation and wall temperature as model input variables, there is a fairly good match between the predicted and measured rock water content, with similar patterns of fluctuations (Fig. 6a). Although there is obvious underestimation of rock water content in mid and late July, and slight underestimation or overestimation in other months, the NSE is as high as 0.958, indicating that the fluctuating relative humidity, air temperature, precipitation and wall temperature can capture the major patterns of fluctuating  
245 rock water content.

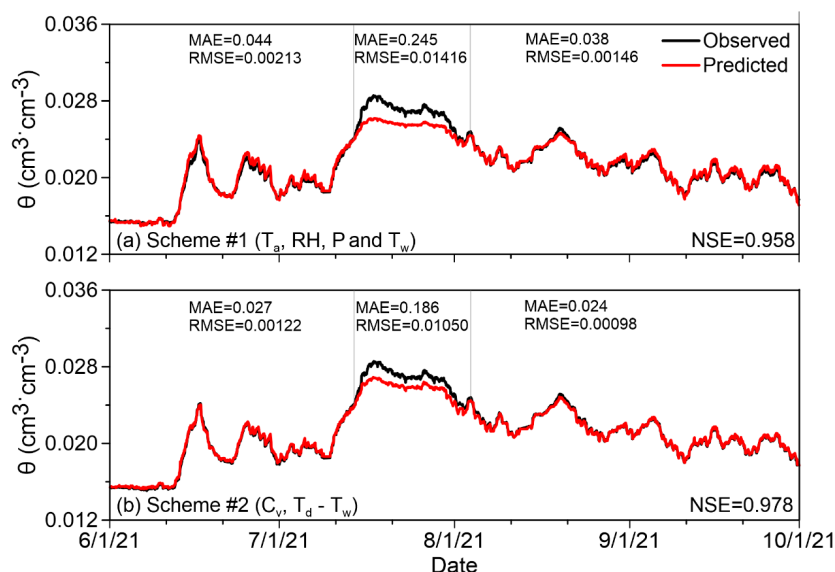
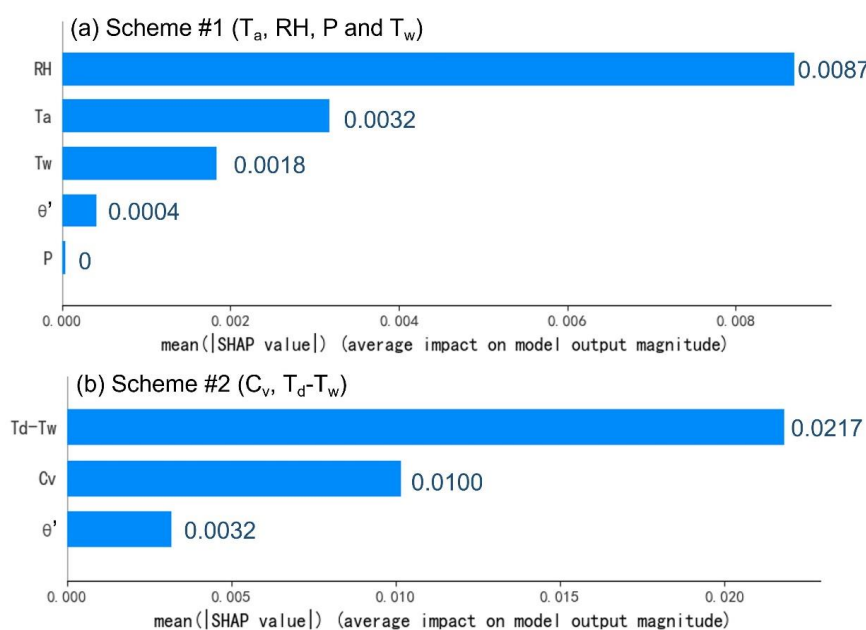


Figure 6. The predicted and measured rock water content obtained by two schemes. (a) Scheme #1 using four directly measured variables; (b) Scheme #2 using two calculated variables controlling vapor condensation. Also shown are NSE of the whole time series, MAE and RMSE of three different stages.



250 Fig. 7a shows the mean absolute SHAP value of each input variable, which represents the relative importance of each variable for the prediction. The mean absolute SHAP values of air relative humidity, air temperature, wall temperature, rock water content at previous time step and precipitation are 0.0087, 0.0032, 0.0018, 0.0004 and 0, respectively. Therefore, precipitation has no direct contribution to rock moisture in caves, and we infer that vapor condensation should be the source of rock moisture in caves.



255

Figure 7. The relative importance of each input variable in the two schemes with different input variables. (a) Scheme #1 using four directly measured variables; (b) Scheme #2 using two calculated variables controlling vapor condensation. Note that rock water content at previous step ( $\theta'$ ) also has contribution to prediction.

#### 260 4.2.2 The prediction results based on variables controlling vapor condensation

Based on the SHAP values of scheme #1, precipitation can be excluded as an input variable for the LSTM network. Among the three directly monitored variables that have contributions to rock water content, air relative humidity and air temperature determines the vapor concentration and the dew point temperature (Nguyen et al., 2014), and whether the wall temperature is below the dew point temperature determines whether vapor condensation could occur (Fernández-Cortés et al., 2006; Gabrovšek et al., 2010; Li et al., 2021). Because water  
265 vapor is the direct source of condensation water and whether the wall temperature is below the dew point



temperature is the precondition of condensation, we use vapor concentration and dew point temperature minus wall temperature as two input variables, both of which are direct variables controlling vapor condensation.

Vapor concentration and dew point temperature are both functions of actual vapor pressure, which is determined by air temperature, saturated vapor pressure and relative humidity. For air with a temperature of  $T$  (K), the formula for calculating saturated vapor pressure and actual vapor pressure are (Lawrence, 2002; Lu, 2004):

$$u_{v,\text{sat}} = 0.611 \exp\left(17.27 \frac{T - 273.2}{T - 36}\right) \quad (13)$$

$$u_v = u_{v,\text{sat}} \cdot RH \quad (14)$$

where  $u_{v,\text{sat}}$  is the saturation vapor pressure (kPa),  $RH$  is the relative humidity of air (%), and  $u_v$  is the actual vapor pressure (kPa). After obtaining  $u_v$ , the vapor concentration,  $C_v$  ( $\text{g}/\text{m}^3$ ), and the dew point temperature,  $T_d$  (K), can be calculated as (Lu, 2004)

$$C_v = 217 \cdot \frac{u_v}{T - 273.15} \quad (15)$$

$$T_d = \frac{36 \ln(u_v) - 4700}{\ln(u_v) - 16.78} \quad (16)$$

As indicated in Equation 16, a higher water vapor content in the air,  $u_v$ , corresponds to a higher dew point temperature, thus a higher possibility of condensation at the wall. Fig. 8 shows that the patterns of fluctuating rock water content, vapor concentration and difference between dew point temperature and wall temperature (denoted as  $T_d - T_w$  hereafter) in the whole non-freezing period are quite similar. Moreover, we find the period with a positive  $T_d - T_w$  has a good correspondence with the period with a high level of rock moisture.

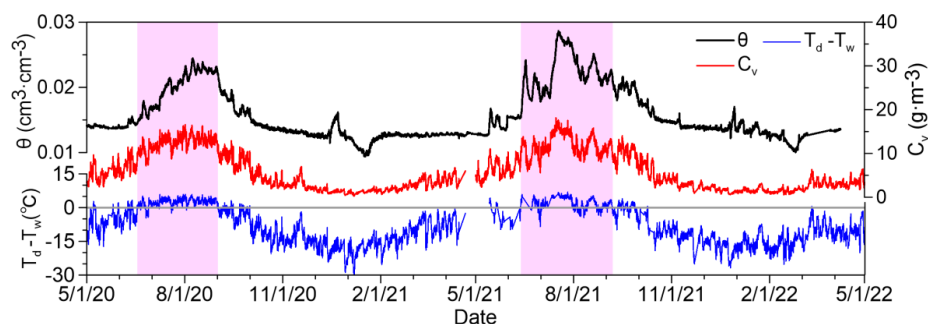


Figure 8. The fluctuating apparent rock moisture ( $\theta$ ), vapor concentration ( $C_v$ ), and the difference between dew point temperature and wall temperature ( $T_d - T_w$ ). The zones in purple represents the periods with positive  $T_d - T_w$  and high rock moisture.



In scheme #2, by using the two new variables as input of the LSTM model, the mean absolute SHAP values of  $T_d-T_w$  and vapor concentration are 0.0217 and 0.0100, respectively (Fig. 7b), indicating that both variables have significant contributions to rock moisture. Moreover, the NSE of predicted rock moisture is further increased to 0.978 (Fig. 6b). Although the prediction still underestimates rock water content from mid July to the end of July, the MAE reduced from 0.245 in scheme #1 to 0.186 in scheme #2, and RMSE reduced from 0.01416 in scheme #1 to 0.01050 in scheme #2. In the other two time durations, the MAE and RMSE of scheme #2 also decrease obviously. Therefore, scheme #2 has much better performance of prediction, showing that using physics-informed variables would improve accuracy of prediction.

#### 4.3 The mechanism of water vapor condensation

As we illustrated in 4.2, precipitation is not directly responsible for rock moisture fluctuations, but other atmospheric conditions controlling vapor concentration and the condition of vapor condensation are directly responsible for rock moisture fluctuations in the cave. In fact, vapor concentration fluctuations are more or less related to precipitation events. As shown in Fig. 9, the vapor concentration usually begins to rise prior to the occurrence of a precipitation event, and declines under the control of solar radiation after a precipitation event.

Under the control of convection and diffusion, the increased water vapor in the air could be transported into porous media. When the sandstone is dry, water vapor can be adsorbed onto the surface of the rock pores, forming an adsorbed layer as thin water film; as curved menisci begin to form under increasing relative humidity, capillary condensation occurs in the rock pores (Broekhoff, 1969; Lu and Likos, 2004; Xu et al., 1998). Both adsorption and capillary condensation would lead to rock moisture addition. As shown in Fig. 9, in the summer of 2021, there are 10 stages with obvious rock moisture additions. In the majority of the 10 stages, there are lagged responses of rock moisture additions to rising vapor concentration in the air, probably due to time required for vapor convection and diffusion.

Among the 10 stages, the magnitude of rock moisture addition is controlled by  $T_d-T_w$ . In stages IV, V, VII, VIII, IX and X, because the dew point temperature seldom exceeds the wall temperature, the magnitudes of rock moisture additions are relatively small. In stages I, II, III and VI, there are long durations with dew point temperature being higher than the wall temperature, causing large magnitudes of rock moisture addition. However, at the beginning of these four stages, even if dew point temperature is still lower than the wall temperature, increasing vapor concentration has resulted in rock moisture addition. Therefore, although a negative  $T_d-T_w$  does not exclude the possible occurrence of capillary condensation, a positive  $T_d-T_w$  does promote capillary



condensation.

Following the 10 stages of rock moisture additions, we find rock moisture depletion is very sensitive to decreasing vapor concentration. Moreover, in stage III with very high rock water content, a slight decrease in vapor concentration results in a slight decrease in rock water content. Therefore, we believe that rock water content measured by the FDR technique is sensitive enough to fluctuating vapor concentrations and can be applied in future rock moisture monitoring in other settings.

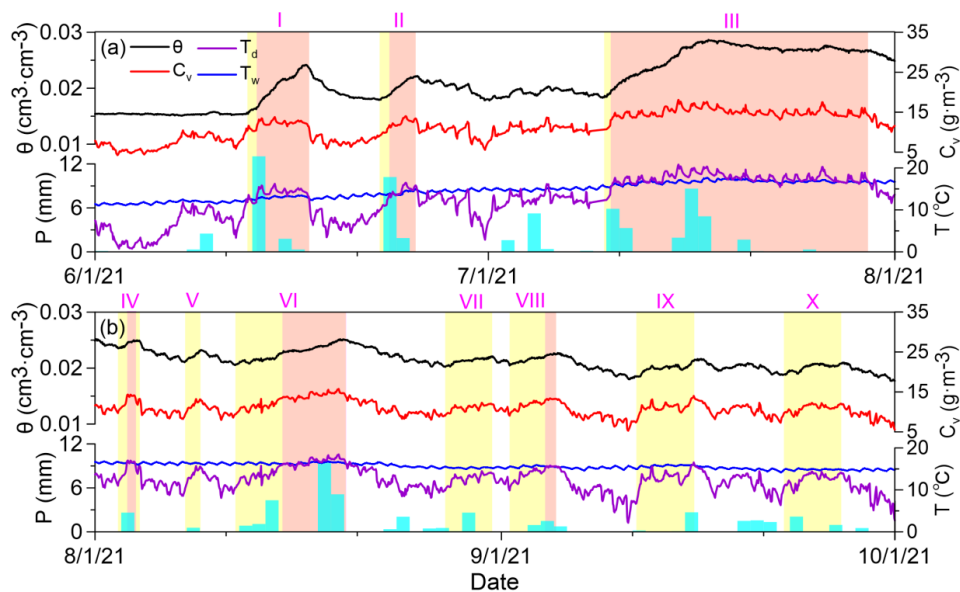


Figure. 9 Plots showing the responses of rock water content ( $\theta$ ) to precipitation ( $P$ ), vapor concentration ( $C_v$ ), dew point temperature ( $T_d$ ) and wall temperature ( $T_w$ ) in the summer of 2021. The zones in yellow have increasing rock water content, and zones in red have increasing rock water content as well as positive  $T_d - T_w$ .

## 5 Conclusion

The source of water in the sandstones caves in the Yungang Grottoes responsible for weathering was a long-standing unresolved scientific question. In this study, we use the FDR to monitor the rock moisture in a cave, which shows clear rock moisture addition-depletions cycles due to various controlling mechanisms. By using relative humidity, air temperature, precipitation and wall temperature as the input variables of the LSTM network, the predicted rock moisture well reproduced the pattern of monitored rock moisture fluctuations. Moreover, we find that precipitation has no contribution, but all other three variables have contribution to the fluctuating rock moisture.



335 Because relative humidity, air temperature and wall temperature belong to factors controlling vapor condensation, this scheme of deep learning reveals that vapor condensation instead of precipitation is the source of rock moisture in the cave.

By calculating vapor concentration and dew point temperature from air temperature and relative humidity, we proposed two new variables, vapor concentration and the difference between dew point temperature and wall  
340 temperature as input variables of the LSTM network. Because the two variables are direct control factors of vapor condensation, this scheme leads to a much better accuracy of prediction, confirming that rock moisture in the cave is derived from vapor condensation. We also analyzed how precipitation events control vapor concentration, thus indirectly control vapor condensation inside the rock.

By examining the sensitiveness of rock moisture to vapor concentration and the difference between dew  
345 point temperature and wall temperature, this study demonstrates for the first time that the FDR technique is an effective means for monitoring rock moisture. By using variables directly controlling vapor condensation as the input variables of the LSTM model, this study shows that “physics-informed” deep learning can improve prediction performance. Moreover, by identifying how vapor condensation controls rock moisture and occasional occurrence of water droplets in the study area, this study contributes to the understanding of sources of water in caves, which  
350 is important in providing scientific-based proofs to propose future strategies for alleviating weathering of stone heritages.

#### **Code availability.**

All code used to generate, train, and test the models are available in a dedicated Zenodo repository:  
355 <https://doi.org/10.5281/zenodo.7382827>.

#### **Data availability.**

The data of rock water content and atmospheric conditions at the field site (from 1 May 2020 to 1 May 2022) are available in a dedicated Zenodo repository: <https://doi.org/10.5281/zenodo.7382895>.

#### **Author contributions.**

360 All authors were involved in interpreting data provided by HY. XJ developed the initial idea of the current study, and KO run the models with the help of GM. KO and XJ wrote the manuscript with contributions from GM and YZ.





### Competing interests.

The authors declare that they have no conflict of interest.

### 365 Financial support.

This study was funded by the National Key R&D Program of China through grant number 2019YFC1520500 and Shanxi Cultural Relics Bureau through contract number 208141400237.

### References

- Arora, B., Dwivedi, D., Faybishenko, B., Jana, R. B. and Wainwright, H. M.: Understanding and predicting vadose  
370 zone processes, *Rev. Mineral Geochem.*, 85, 303–328, <https://doi.org/10.2138/rmg.2019.85.10>, 2019.
- Auler, A. S. and Smart, P. L.: Rates of condensation corrosion in speleothems of semiarid northeastern Brazil, *Speleogenesis Evol. Karst Aquifers*, 2, 2004.
- Barzegar, R., Fijani, E., Moghaddam, A. A. and Tziritis E.: Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models, *Sci. Total. Environ.*, 599–600, 20–31,  
375 <https://doi.org/10.1016/j.scitotenv.2017.04.189>, 2017.
- Bengio, Y., Simard, P. and Frasconi, P.: Learning long-term dependencies with gradient descent is difficult, *IEEE Trans. Neural Netw. Learn Syst.*, 5, 157–166, <https://doi.org/10.1109/72.279181>, 1994.
- Broekhoff, J., C., P.: *Sorption and Capillarity*, Waltman, Delft, 1969.
- Brubaker, K. L. and Entekhabi, D.: Analysis of feedback mechanisms in land-atmosphere interaction, *Water Resour. Res.*, 32, 1343–1357, <https://doi.org/10.1029/96WR00005>, 1996.
- 380 Camuffo, D.: *Microclimate for cultural heritage*. Elsevier, Amsterdam, 1998.
- Cao, W. B., Wan, L., Zeng, Y. J., Gan, X. M., Jiang, X. W., Huang, J. Z., Yan, H. B. and Zhang, L.: Study on formation mechanism and prevention of condensation water in Yungang Grottoes, in: *Collection of Papers of the 2005 Yungang International Academic Seminar (Protection Volume)*. Cultural Relics Publishing House, Beijing, 191–198, 2005. (in Chinese)
- 385 Chen, J., Gao, X., Zheng, X., Miao, C., Zhang, Y., Du, Q., and Xu, Y.: Simulation of Soil Freezing and Thawing for Different Groundwater Table Depths, *Vadose Zone J.*, 18, 18057, <https://doi.org/10.2136/vzj2018.08.0157>, 2019.
- De Freitas, C. R., Schmekal, A.: Studies of condensation/evaporation processes in the Glowworm Cave, New Zealand. *Int. J. Speleol.*, 35, 75–81, <http://dx.doi.org/10.5038/1827-806X.35.2.3>, 2006.
- 390 Deprez M., De Kock, T., De Schutterc, G. and Cnudde, V.: A review on freeze–thaw action and weathering of rocks,



- Earth–Sci. Rev. 203, 103143, 2020.
- Fernández-Cortés, Á., Calaforra, J.M. and García-Guinea, J.: The Pulpí gigantic geode (Almería, Spain): geology, metal pollution, microclimatology, and conservation, *Environ. Geol.*, 50, 707–716, 395  
<https://doi.org/10.1007/s00254-006-0243-z>, 2006.
- Fang, K., Pan, M. and Shen, C. P.: The value of SMAP for long-term soil moisture estimation with the help of deep learning, *IEEE T. Geosci. Remote*, 57, 2221–2233, <https://doi.org/10.1109/TGRS.2018.2872131>, 2019.
- Fischer, T. and Krauss, C.: Deep learning with long short-term memory networks for financial market predictions, *Eur. J. Oper. Res.*, 270, 654–669, <https://doi.org/10.1016/j.ejor.2017.11.054>, 2018.
- 400 Gabrovšek, F., Dreybrodt, W. and Perne, M., Physics of condensation corrosion in caves. In: *Advances in Research in Karst Media. Environmental Earth Sciences*, edited by: Andreo, B., Carrasco, F., Durán, J. J. and LaMoreaux, J. W., Springer, Berlin, Heidelberg, 491–496, [https://doi.org/10.1007/978-3-642-12486-0\\_75](https://doi.org/10.1007/978-3-642-12486-0_75), 2010.
- Gao, S., Huang Y. F. Zhang, S., Han, J. C., Wang, G. Q., Zhang, M. X. and Lin, Q. S.: Short-term runoff prediction 405  
with GRU and LSTM networks without requiring time step optimization during sample generation, *J. Hydrol.*, 589, 125188, <https://doi.org/10.1016/j.jhydrol.2020.125188>, 2020.
- Guerrier, B., Doumenc, F., Roux, A., Mergui, S. and Jeannin, P.Y.: Climatology in shallow caves with negligible ventilation: Heat and mass transfer. *Int. J. Therm. Sci.*, 146, 106066, <https://doi.org/10.1016/j.ijthermalsci.2019.106066>, 2019.
- 410 Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Comput.*, 9, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>, 1997.
- Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities, *P. Natl. Acad. Sci. USA*, 79, 2554–2558, <https://doi.org/10.1073/pnas.79.8.2554>, 1982.
- Huang, K. Z.: Harm and Prevention of Water in Yungang Grottoes, in: *Cultural relics protection technology (1981- 415  
1991)*, China Science Publishing & Media Ltd, Beijing, 284–290, 2010. (in Chinese)
- Huang, J. Z.: Waterproof protection of Yungang Grottoes, a world cultural heritage. *Sciences of Conservation and Archaeology*, 20, 114–121, 2008. (in Chinese)
- Huang, J. Z., Wan, L., Peng, T., Cao, W. B. and Wang, X. S.: Survey on water sources in Yungang Grottoes and some results. *Geotechnical Investigation & Surveying*, 40, 1–5, 2012. (in Chinese)
- 420 Irmak, S. and Irmak, A.: Performance of frequency-domain reflectometer, capacitance, and psuedo-transit time-based soil water content probes in four coarse-textured soils, *Appl. Eng. Agric.*, 21, 999–1008,



- <https://doi.org/10.13031/2013.20035>, 2005.
- Jumin E., Zaini, N., Ahmed A. N., Abdullah, S., Ismail, M., Sherif, M., Sefelnasr, A. and Ahmed El-Shafie, A.:  
Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction,  
425 Eng. Appl. Comp. Fluid, 14, 713–725, <https://doi.org/10.1080/19942060.2020.1758792>, 2020.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S. F. and Yang, L.: Physics-informed machine  
learning. Nat. Rev. Phys., 3, 422–440, <https://doi.org/10.1038/s42254-021-00314-5>, 2021.
- Larson, J., Lidberg, W., Ågren, A. M. and Laudon, H.: Predicting soil moisture conditions across a heterogeneous  
boreal catchment using terrain indices, Hydrol. Earth Syst. Sci., 26, 4837–4851, <https://doi.org/10.5194/hess->  
430 26-4837-2022, 2022
- Lawrence, D. S.: Physical hydrology, Prentice Hall, Upper Saddle River, 2002.
- Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G. and Dadson, S. J.: Benchmarking data-driven  
rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with  
four lumped conceptual models, Hydrol. Earth Syst. Sci., 25, 5517–5534, <https://doi.org/10.5194/hess-25->  
435 5517-2021, 2021.
- Li, P. J., Zha, Y. Y., Tso, C. M., Shi, L. S., Yu, D. Y., Zhang, Y. G. and Zeng, W. Z.: Data assimilation of uncalibrated  
soil moisture measurements from frequency-domain reflectometry, Geoderma, 374, 114432,  
<https://doi.org/10.1016/j.geoderma.2020.114432>, 2020.
- Li, Y. F., Aemisegger, F., Riedl, A., Buchmann, N. and Eugster, W.: The role of dew and radiation fog inputs in the  
440 local water cycling of a temperate grassland during dry spells in central Europe, Hydrol. Earth Syst. Sci., 25,  
2617–2648, <https://doi.org/10.5194/hess-25-2617-2021>, 2021.
- Liñán, C., Benavente, J., del Rosal, Y., Vadillo, I., Ojeda, L. and Carrasco, F.: Condensation water in heritage  
touristic caves: Isotopic and hydrochemical data and a new approach for its quantification through image  
analysis, Hydrol. Process, 35, e14083. <https://doi.org/10.1002/hyp.14083>, 2021.
- 445 Lipton, Z. C.: Berkowitz, J. and Elkan, C.: A critical review of recurrent neural networks for sequence learning,  
arXiv [preprint], arXiv: 1506.00019, <https://doi.org/10.48550/arXiv.1506.00019>, 2015.
- Liu, G. J., Tang, Z. Y., Qin, H., Liu, S., Shen, Q., Qu, Y. H., Zhou, J. Z.: Short-term runoff prediction using deep  
learning multi-dimensional ensemble method, 609, 127762, <https://doi.org/10.1016/j.jhydrol.2022.127570>,  
2022.
- 450 Lundberg, S. M. and Lee, S.: A unified approach to interpreting model predictions, in: Advances in neural  
information processing systems 30 (NIPS 2017), edited by: Guyon, I., Luxburg, U. V., Bengio, S., Wallach,



- H., Fergus, R., Vishwanathan, S. and Garnett, R., Neural information processing systems (NIPS), La Jolla, 2017.
- Matsuoka, N., Murton, J.: Frost weathering: recent advances and future directions. *Permafrost Periglac.*, 19, 195-455 210, <https://doi.org/10.1002/ppp.620>, 2008.
- Meinzer, O. E.: The history and development of ground–water hydrology. *J Washington Academy of Sciences*, 24, 6-32, <http://www.jstor.org/stable/24532897>, 1934.
- Mollo, L. and Greco, R.: Moisture measurements in masonry materials by time domain reflectometry. *J. Mater. Civil Eng.*, 23, 441-444, 2011.
- 460 Nguyen, J. L., Schwartz, J. and Dockery, D. W.: The relationship between indoor and outdoor temperature, apparent temperature, relative humidity, and absolute humidity, *Indoor Air*, 24, 103–112, <https://doi.org/10.1111/ina.12052>, 2013.
- Rempe, D. M. and Dietrich, W. E.: Direct observations of rock moisture, a hidden component of the hydrologic cycle, *P. Natl. Acad. Sci. USA*, 115, 2664–2669, <https://doi.org/10.1073/pnas.1800141115>, 2018.
- Sakaki, T. and Rajaram, H.: Performance of different types of time domain reflectometry probes for water content measurement in partially saturated rocks, *Water Resour. Res.*, 42, W07404, 465 <https://doi.org/10.1029/2005WR004643>, 2006.
- Salve, R., Rempe, D. M. and Dietrich, W. E.: Rain, rock moisture dynamics, and the rapid response of perched groundwater in weathered, fractured argillite underlying a steep hillslope, *Water Resour. Res.*, 48, 1–25, <https://doi.org/10.1029/2012WR012583>, 2012.
- Sass, O.: Rock moisture measurements: techniques, results, and implications for weathering. *Earth Surf. Proc. Land*, 470 30, 359–374, <https://doi.org/10.1002/esp.1214>, 2005.
- Schmidt, L. and Rempe, D.: Quantifying dynamic water storage in unsaturated bedrock with borehole nuclear magnetic resonance, *Geophys. Res. Lett.*, 47, e2020GL089600, <https://doi.org/10.1029/2020GL089600>, 2020.
- Schoups, G., Hopmans, J. W., Young, C. A., Vrugt, J. A. and Wallender, W.W.: Multi-criteria optimization of a regional spatially-distributed subsurface water flow model, *J. Hydrol.*, 31, 20–48, 475 <https://doi.org/10.1016/j.jhydrol.2005.01.001>, 2005.
- Shapley, L.S., 1953. A value for n-person games, in: *Contributions to the Theory of Games II*, editd by: Kuhn, H. and Tucker, A., Princeton University Press, Princeton, 307–317, <https://doi.org/10.1515/9781400881970-018>, 1953.



- Štrumbelj, E. and Kononenko, I.: Explaining prediction models and individual predictions with feature  
480 contributions, *Knowl. Inf. Syst.* 41, 647–665, <https://doi.org/10.1007/s10115-013-0679-x>, 2014.
- Sun, Z. H. and Scherer, G. W.: Measurement and simulation of dendritic growth of ice in cement paste, *Cement  
Concrete Res.*, 40, 1393–1402, <https://doi.org/10.1016/j.cemconres.2010.03.008>, 2010.
- Tindall, J., Kunkel, J. and Anderson, D. E.: *Unsaturated Zone Hydrology for Scientists and Engineers*, Pearson  
Education Inc., New York, 1998.
- 485 Vereecken, H., Huisman, J. A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogen, H., Weihermuller, L., Herbst,  
M., Martinez, G. and Vanderborght, J.: On the spatio-temporal dynamics of soil moisture at the field scale, *J.  
Hydrol.*, 516, 76–96, <https://doi.org/10.1016/j.jhydrol.2013.11.061>, 2014.
- Vinnikov, K. Y., Robock, A., Speranskaya, N. A., and Schlosser, C. A.: Scales of temporal and spatial variability  
of midlatitude soil moisture, *J. Geophys. Res.-Atmos.*, 101, 7163–7174, <https://doi.org/10.1029/95JD02753>,  
490 1996.
- Wang, X. S., Wan, L., Peng, T., Dong, P. and Qian, J.: Infiltration water in Yungang Grottoes and its movement,  
*Geotechnical Investigation & Surveying*, 40, 12–16, 2012. (in Chinese)
- Xiang, Z. R., Yan, J. and Demir I.: A rainfall-runoff model with LSTM-based sequence-to- sequence learning,  
*Water Resour. Res.*, 56, e2019WR025326, <https://doi.org/10.1029/2019WR025326>, 2020.
- 495 Xie, H. Y., Jiang, X. W., Tan, S. C., Wan, L., Wang, X. S., Liang, S. H. and Zeng, Y. J.: Interaction of soil water  
and groundwater during the freezing–thawing cycle: field observations and numerical modeling, *Hydrol. Earth  
Syst. Sci.*, 25, 4243–4257, <https://doi.org/10.5194/hess-25-4243-2021>, 2021.
- Xu, L., Lio, A., Hu, J., Ogletree, A. D. F. and Salmeron, M.: Wetting and capillary phenomena of water on mica, *J.  
Phys. Chem. B*, <https://doi.org/10.1021/JP972289L>, 1998.
- 500 Yu, L. Y., Zeng, Y. J., Su, Z. B., Cai, H. J. and Zheng, Z.: The effect of different evapotranspiration methods on  
portraying soil water dynamics and ET partitioning in a semi-arid environment in Northwest China. *Hydrol.  
Earth Syst. Sci.*, 20, 975–990, <https://doi.org/10.5194/hess-20-975-2016>, 2016.
- Yu, L., Zeng, Y., Wen, J. and Su, Z.: Liquid-Vapor-Air Flow in the Frozen Soil. *J. Geophys. Res.-Atmos.*, 123, 7393-  
7415, <https://doi.org/10.1029/2018jd028502>, 2018.
- 505 Zhang, Z. Y., Wang, W., Gong, C. C., Wang, Z. F., Duan, L., Yeh, T. J. and Yu, P. Y.: Evaporation from seasonally  
frozen bare and vegetated ground at various groundwater table depths in the Ordos Basin, Northwest China.  
*Hydrol. Process.*, 33, 1338-1348, <https://doi.org/10.1002/hyp.13404>, 2019.
- Zhao, H. F., Li, J., Yuan, Q. Q., Lin, L. P., Yue, L. W. and Xu, H. Z.: Downscaling of soil moisture products using



- deep learning: Comparison and analysis on Tibetan Plateau, *J. Hydrol.*, 607, 127570,  
510 <https://doi.org/10.1016/j.jhydrol.2022.127570>, 2022.
- Zhao, Z., Chen, W. H., Wu, X. M., Chen, P. C. Y. and Liu, J. M.: LSTM Network: a deep learning approach for  
short-term traffic forecast, *IET Intell. Transp. Sy.*, 11, 68–75, <https://doi.org/10.1049/IET-ITS.2016.0208>,  
2017.
- Zhu, Y., Irmak, S., Jhala, A. J., Vuran, M. C. and Diotto, A.: Time-domain and frequency-domain reflectometry  
515 type soil moisture sensor performance and soil temperature effects in fine- and coarse-textured soils, *Appl.*  
*Eng. Agric.*, 35, 117–134, <https://doi.org/10.13031/aea.12908>, 2019.