Evaluation of the predictions skills and uncertainty of a karst model using short calibration data sets at an Apulian cave (Italy)

T. Leins[1,3] · I. S. Liso[2] · M. Parise[2] · A. Hartmann[1,3]

Abstract

The freshwater resource in karst is subjected to both sea level rise and an increasing pressure caused by the high-water demand. Therefore, developing an understanding of the hydrogeological dynamics of the karst aquifer can be a useful tool for improving protection and management actions. Vora Bosco cave (Apulia, Southern Italy) was instrumented with a multi-parameter probe for groundwater level measurements, aimed at exploring the system behavior within the cave recharge area. To characterize and quantify the natural recharge and discharge behavior of the system, a simple reservoir model was developed, initially also with the intention of predicting groundwater dynamics. Based on the original time-series of water level observations, different calibration datasets were established using different split-sample and bootstrapping approaches, and a regional sensitivity analysis was executed. Furthermore, in addition to the original observation time-series, a 3-month extension was used as a model testing period. Using these analyses, the parameters identifiability and the predictions robustness for the model testing period were evaluated. Results reveal that while the calibration on the whole dataset, as well as the bootstrapping approaches, lead to better performances in the calibration and validation period of the original time-series, and to a higher model precision with smaller uncertainty ranges. their performance in the model testing period becomes very poor and the observed water level data no longer plots within the uncertainty bands. Based on this extensive analysis, the model is finally rejected. Our study therefore also confirms the importance of model validation, especially when only a short time-series of observations are available.

Keywords Karst · Cave · Modeling · Hydrogeology · Apulia

Introduction

About a quarter of the global population derive at least parts of their drinking water from karst aquifers (Ford and Williams 2007; Stevanović 2018; Stevanovic 2019). In the future, many of these karst regions, which account for 7–12% of the world's continental area, will face strong anthropogenic and natural changes. These changes can in turn have negative impacts on water security (Hartmann et al. 2014). This also applies to our study area, Salento (Apulia, southern Italy), where the climate is shifting from Mediterranean to arid or semi-arid, due to climate change (Masciopinto and Liso 2016; Alfio et al. 2020). Further, sea level rise as well as high water demand put pressure on the freshwater resources. Therefore, gaining hydrogeological knowledge is an important measure to sustainably protect and manage karst water resources (Hartmann et al. 2014; Parise et al. 2015, 2018; Stevanovic 2019; Olarinoye et al. 2020).

Understanding the true aquifer hydrodynamic in karst is not a simple issue especially because in such an environment, the hydrological and physical characteristics are extremely variable in space and time. For this, the classical hydrogeological surveys, used for porous media, do not guarantee a good description of the real processes. Studies in karst require specific hydrogeological approaches and methods (Goldscheider and Drew 2007; Ford and Williams 2007) since the karst underground shows various features such as caves, conduits, fractures that are part of a network

✉ I. S. Liso
serenaliso.uniba@gmail.com

1 Chair of Hydrological Modeling and Water Resources, University of Freiburg, 79098 Freiburg, Germany

2 Department of Earth and Environmental Sciences, University Aldo Moro, Bari, Italy

3 Institute of Groundwater Management, Technische Universität Dresden, 01069 Dresden, Germany

of interconnected voids, through which water flows below the surface. This, together with the turbulent regime which is typically experienced in karst conduits, represents the main difficulties in the implementation of traditional hydrogeological models. In addition, the models require a good knowledge of the subsurface characteristics as the starting point to obtain a realistic output. However, underground data in karst are typically obtained through classical geological surveys and field tests (for example pumping and tracer tests); being indirect data, they need to be interpreted, leading to high degrees of uncertainty. The only way to overpass these obstacles and to improve the hydrological forecasts and scenarios is to collect data directly from below, where possible, by means of speleological explorations and observation within the cave systems (Jeannin et al. 2007; Palmer 2007). In this way, the groundwater signals will be free of anthropogenic disturbances (for example wells exploitation) and the geometry of the voids can be well appreciated directly from inside.

Furthermore, to simulate karst aquifer functioning, different types of models can be used (Fleury et al. 2007). They are an essential tool in hydrology and represent the water cycle or parts of it on different scales and in different accuracy levels (Hörmann 2016). Their aim can be to analyze karst aquifer functioning, to test the fit of conceptual models or also to predict flow rates or water quality in order to manage water resources (Fleury et al. 2007). In general, hydrological models can be classified into different groups according to their characteristics. In so-called black box models, the model input is transformed into the model output using only statistical relations. These black-box models cannot be applied outside the value range they were calibrated in and therefore cannot be used for scenario analyses (Kresic 2013; Hartmann et al. 2014; Hörmann 2016).

The opposite of a black box model would be a physical, mathematical model, that is only based on basic physical equations and that represents all physical processes happening in the aquifer. Due to the large heterogeneity of karst aquifers and the lack of sufficient data, purely physical models are almost impossible to establish (Fleury et al. 2007; Mikszewski and Kresic 2015; Hörmann 2016). However, there are also process-based models that do consider processes that take place in karst aquifers. These process-based models can either be distributed or lumped models, where in distributed models the karst system is split into grids with assigned hydraulic parameters, whereas in lumped models the whole karst system is modeled as one system without explicit spatial variability. Both distributed and lumped models are based on physical processes that occur in karst aquifers, but these processes are very simplified. Therefore, the model parameters are mostly determined by calibration instead of being measured in the field (Hartmann et al. 2014). In this model calibration step the model parameters are adjusted, manually or automatically, so that the modeled data match the observed data as well as possible during a specific calibration period (Arsenault et al. 2018; Shen et al. 2022). How similar modeled data and observed data are, and therefore how well the model parameters have been adjusted, is mostly quantified using an objective function (Arsenault et al. 2018). The calibration of the model parameters, however, yields other problems. If too many parameters are determined by calibration, this can lead to equifinality, where the same model output can be achieved by different parameter combinations (Beven 2008). This over-parametrization can lead to an increase of uncertainty for process-based models and the models can become less reliable for predictions (Hartmann et al. 2014). Therefore, the model complexity should always respond to the availability of data (Hörmann 2016).

In addition to the model calibration, it is common in hydrological modeling to further perform a model validation with observed data, which were not used for calibration, with the aims of evaluating the robustness and adequacy of the calibrated model and ensuring parameter transferability (Klemeš 1986; Legates and McCabe 1999; Arsenault et al. 2018; Shen et al. 2022).

In order to do so, the available observation dataset needs to be split into data for calibration and data for validation. This widely used approach is called split-sample test and is one of four levels of model validation proposed by Klemeš (1986). Klemeš (1986) originally suggested a strategy where the available data are split into two sub-periods. One sub-period is used for model calibration and the other sub-period for validation. Then, the sub-periods are switched and the process is repeated. If both ways of calibration and validation lead to acceptable and similar validation results, the model can be considered as acceptable (Klemeš 1986). For a "sufficiently long" data time-series, Klemeš (1986) recommend to use 50% of the data for calibration and 50% for validation, while for shorter data time-series it is recommended to use more data for calibration. Many different variations of this split-sample approach by Klemeš (1986) have been implemented in hydrological modeling (Arsenault et al. 2018). In the most commonly used simplified variation of the split-sample test there is only one single calibration period and one validation period, which are not switched (Shen et al. 2022). In addition to using continuous calibration and validation periods, the calibration and validation data can also be bootstrapped. In this case, a certain proportion of data are randomly sampled out of all available data. This discontinuous dataset is then used for calibration and the remaining data are used for validation. The process is repeated several times to generate more bootstrapping samples. In this way, several parameter sets are calibrated and can be compared (Arsenault et al. 2018).

The split-sample approach has, however, been criticized by different studies (Arsenault et al. 2018; Shen et al. 2022). They state that how to exactly split a data set is not supported by empirical evidence and furthermore not clearly defined, leading to a subjective decision to be made by each modeler (Arsenault et al. 2018; Shen et al. 2022). They further emphasize that the information included in the calibration period is vital for model calibration (Sorooshian et al. 1983). In this context, Shen et al. (2022) especially criticize the widely used approach of splitting calibration and validation data chronologically, using older data for calibration and recent data for validation. In their empirical study, where they analyze the influence of the data splitting technique on the model performance in an independent model testing period, they show that this chronological approach should be avoided. They furthermore come to the conclusion that the most robust choice is to completely skip model validation and to use all available data for model calibration. This is in accordance with the findings of Arsenault et al. (2018), where it is stated that the best strategy is to calibrate on the whole dataset, leading to the best performance in the model testing period and the most robust calibrated parameter sets. In this study, we use a short measured time series of water level measurements in the Vora Bosco Cave to establish a simple reservoir model, which should help to better understand the Apulian karst system. We calibrate and validate this simple model with a short time series using different calibration datasets. We then use an additional model testing period to investigate the performance of the different calibration datasets.

Data and methods

Study site and data

Apulia region is among the most important karst areas in the Mediterranean Basin, with a huge number of surface and underground karst features. Its southernmost sector (Salento) is a flat and elongated peninsula, surrounded by the Ionian Sea to the SW and the Adriatic Sea to the NE (Fig. 1). The most typical karst landforms are small-size dolines/sinkholes (Delle Rose et al. 2004; Festa et al. 2012; Pepe and Parise 2014). Hundreds of caves characterize the Salento coastlines, partly originated or later modified thanks to marine actions (Onorato et al. 2006; Margiotta et al. 2021). Along the shorelines, many karst springs occur (Liso and Parise 2020), including sulfur features of hypogene origin at Santa Cesarea Terme (D'Angeli et al. 2017, 2019, 2021). Inland, many swallow holes act as points of groundwater concentrated recharge; they are locally named "*vore*", a local term meaning hole (Parise et al. 2003).

The physical landscape of Salento is often linked by small scarps related to faults, striking mainly NNW–SSE, and subordinately NW–SE. These faults are related with the evolution of the Apulian Foreland (Funiciello et al. 1991) and were active during Plio-Pleistocene times; they show high-angle planes with opposite dip directions, along which a maximum extensional offset of about 200 m has been estimated (Festa et al. 2015). The main morphological landforms are therefore alternating ridges and depressions, roughly elongated about NNW–SSE. The ridges (*Serre Salentine*) reach a maximum altitude of 199 m a.s.l., while Vora Bosco opens in a topographically depressed area in the middle of the peninsula. As documented also in other sectors of Apulia (Zumpano et al. 2019), a general correlation between karst morphology, development of surface and subsurface features and tectonics is found in Salento (Palmentola and Vignola 1980; Parise 2008; Pepe and Parise 2014).

The Salento groundwater is a freshwater lens fluctuating on seawater, due to marine intrusion coming from both sides of the peninsula, which cause severe problems of saline pollution (Tulipano and Fidelibus 2002; Margiotta and Negri 2005; Cotecchia 2014; Masciopinto and Liso 2016; Masciopinto et al. 2017). Seawater intrusion phenomena are related to natural causes, but are definitely exacerbated by human activities, such as over-exploitation of the groundwater resource, especially occurring during the summer season, due to high tourist vocation of the area. For these reasons, a good knowledge of the aquifer hydrodynamic represents the primary objective to achieve, in order to preserve and protect the precious underground freshwater resource (Olarinoye et al. 2020).

Salento is a remarkable area as concerns biodiversity in cave water environment, hosting many peculiar species and stygofauna of high interest to science (Hollingsworth et al. 2008; Deharveng et al. 2009; Stoch et al. 2009; Cantonati et al. 2020). Studying the stygofauna from wells and caves in Salento is providing very useful information about quality of karst groundwaters, since several species are markers of unpolluted water environment (Masciopinto et al. 2006; Inguscio et al. 2009; Liso et al. 2019, 2020).

The area surrounding Vora Bosco is characterized by scarce or absent surface runoff in normal conditions. During rainstorms, flash floods frequently occur, especially when characterized by concentrated and intense rainfall. On such occasions, rainwater is collected and transported through the ephemeral rivers activated during the rainstorms, which transfer the surface runoff toward the many swallow holes and the different typologies of sinkholes (see Gutierrez et al. 2014; Parise 2019, 2022) present in the area. These karst features contribute therefore to the natural recharge of groundwater in a concentrate way. When they are not able to absorb the amount of water, an additional contribution to groundwater recharge occurs in a diffuse way.

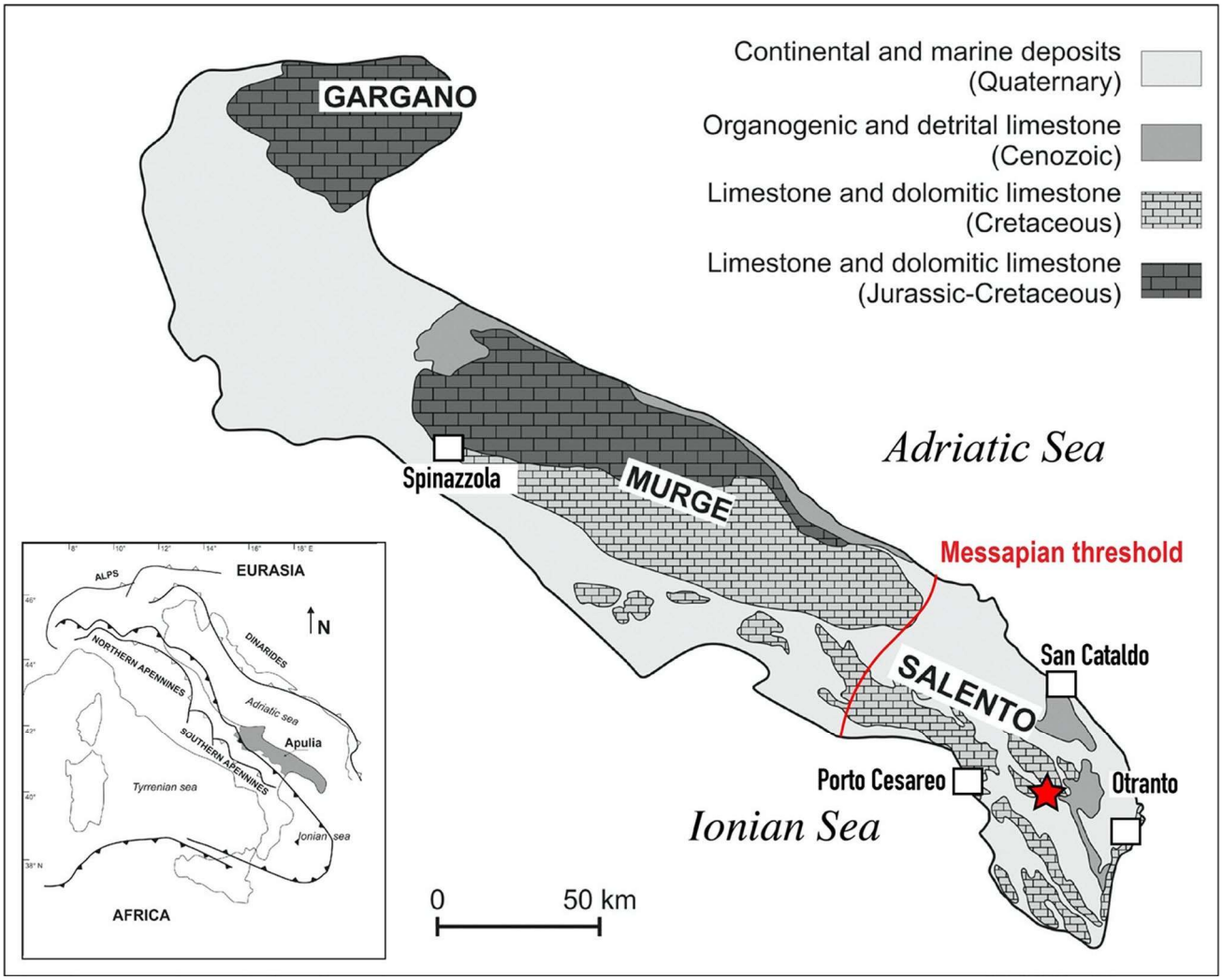
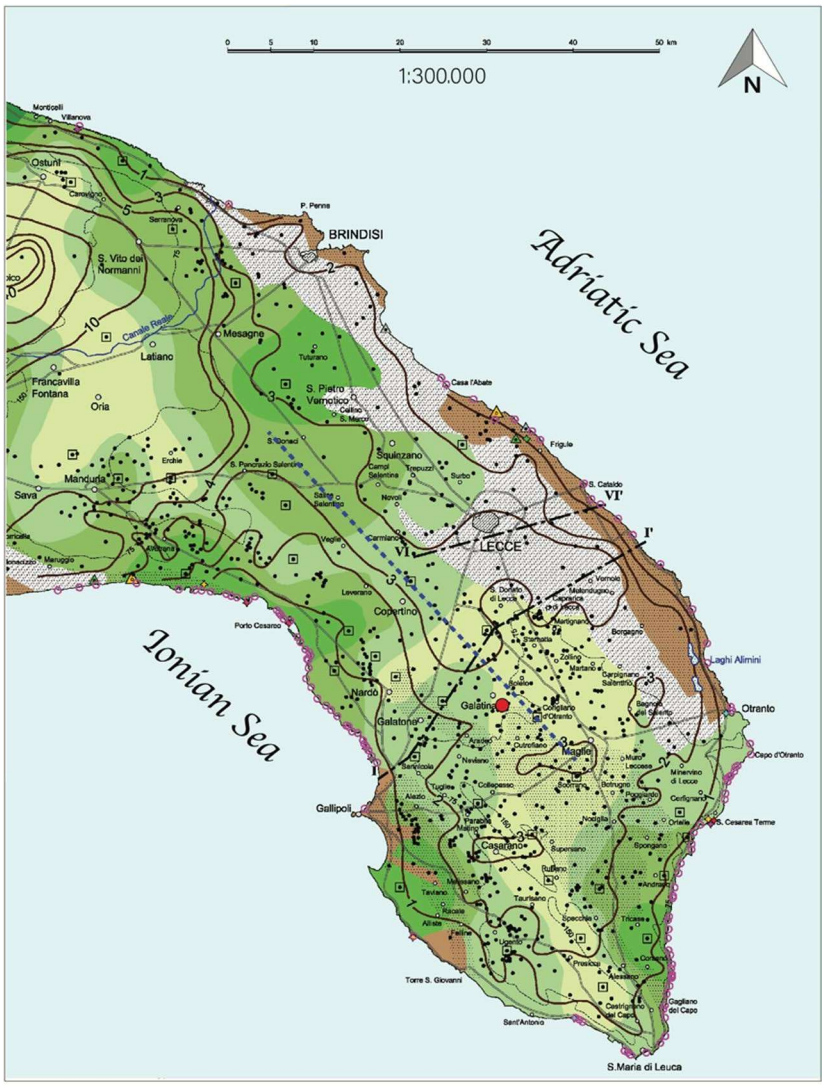


Fig. 1 Map of Italy and schematic geological map of Apulia Region. The red star identifies the Vora Bosco location, and the hydrogeological water divide, named Messapian threshold, is highlighted in red (mod. after Mongelli et al. 2015)

Based upon the present knowledge of Salento hydrogeology, Vora Bosco is approximately located in correspondence of the main hydrogeological divide (Figs. 2, 3), separating the freshwater flux into two main directions: to the east, toward the Adriatic Sea, and to the west, in direction of the Ionian Sea. At a greater detail, the cave is within the buffer zone of the divide, that is the area where this moves in function of the freshwater volumes entering (recharge) and exiting (withdrawals) the aquifer system. Because of the huge number of extraction wells, especially in the central portion of the peninsula, a disorder in the underground waterflow dominates, making difficult the precise identification of the true groundwater flow direction.

Vora Bosco (Lat. N 40.15049670; Long. E 18.17463107) is the only cave in Salento where direct explorations by cavers were able to reach the groundwater level. The cave (entrance at 64 m a.s.l.) is located at the outskirts of the small village of Noha (Galatina municipality), at the bottom of a W–E elongated solution doline. The karst system has overall development of 161 m, and the water table is reached at about 60 m-depth. The cave passages consist in alternating sub-horizontal meandering galleries and vertical shafts controlled by the main discontinuity systems (Fig. 4).

Differently than in the rest of Apulia, in Salento the Cretaceous limestone bedrock is overlain by a complex sequence of Tertiary and Quaternary clastic carbonate deposits (Funiciello et al. 1991; Bosellini et al. 1999). Moving through Vora Bosco cave, the whole stratigraphic sequence can be appreciated, starting with Marine Terraced deposits (Middle–Upper Pleistocene), followed by the Lower Pleistocene "*Calcarenite di Gravina*" where the first sector of shafts develops. This latter stratigraphic unit is characterized by medium permeability value of 10^{-5} m/s (Andriani and



Main Spring Flow rate and Salinity

Flow rate (l/s)	Salinity (g/l)					
	0,4÷1	1÷3	3÷6,5	6,5÷11,5	11,5÷15,5	non nota
0 ÷ 10	–	▲	▲	–	–	▲
10 ÷ 100	▲	▲	▲	–	–	▲
100 ÷ 300	–	▲	▲	▲	–	▲
300 ÷ 900	–	▲	▲	▲	▲	▲
> 900	–	▲	▲	▲	–	–
non nota	◆	◆	◆	◆	◆	–

Fig. 2 Hydrogeological map of Salento, with location of Vora Bosco (red circle) and the main Salento hydrogeological divide (blue dotted line) (mod. after Cotecchia 2014)

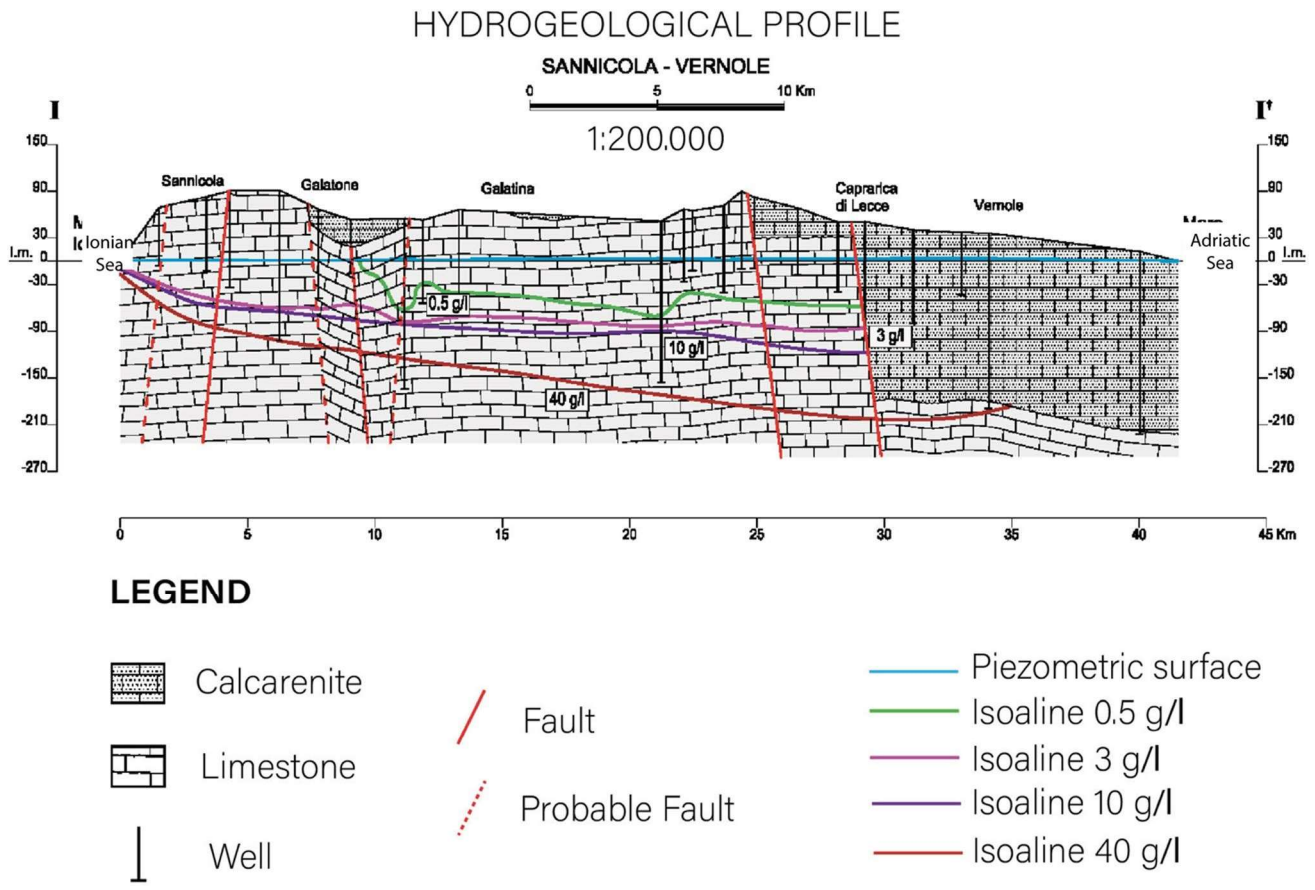


Fig. 3 Salento cross section I-I'; trace of section shown in Fig. 2 (mod. after Cotecchia 2014)

Walsh 2010; Margiotta et al. 2010; Andriani et al. 2021), in function of both grain size and litho-stratigraphic conditions. The Pleistocene deposits unconformably cover the Miocene formations, differentiated in *Pietra Leccese* and *Calcarenite di Andrano,* both consisting of fine and medium-fine grain marls and calcarenites, with local presence of more friable layers. They are characterized by very low porosity and permeability linked essentially to fracturing, leading to values of 10^{-6} m/s. These deposits mark the mostly vertical development of the cave, through a series of shafts ending with a very narrow passage leading to the final drop of about 7 m, down to the water table. Sub-horizontal bedded Cretaceous limestones and dolostones, with high degree of fracturing, and intensively karstified, crop out in this last part of the cave. High permeability values (10^{-3} m/s) are typical of this rock formation (Cotecchia 2014, and references therein), which represent the Mesozoic carbonate bedrock of Salento, and of the whole Apulia as well (Festa et al. 2015). Eventually, as the last rock type cropping out within the karst system, pockets of well cemented calcareous breccia have been recognized at several locations.

The closeness to the surface allows cavers to physically reach the water table. Taking advantage of this opportunity, in the timespan November 2017–October 2018 the cave was instrumented with a multi-parameter probe for groundwater monitoring and sensors for cave microclimate measurements. In detail, the multi-parameter probe is the OTT Orpheus Mini (OTT-Corr-Tek), and was installed about 2 m below the water table and, at installation, the probe was set to register one measure per hour. It measures groundwater temperature (°C, °F), conductivity (mS/cm o µS/cm) and level/pressure (m, cm/bar, psi). The probe is able to return the salinity (PSU, ppt) e and TDS, total dissolved solid, (mg/l) thanks to specific automated algorithm using the EC values. This dataset from November 2017 until July 2018 is further referred to as the original dataset.

Data from the closest rain-gauge station to Vora Bosco, at Galatina, indicate a mean annual temperature around 16 °C, with annual rain values of 663 mm. As concerns the time of interest of this study (years 2017–2018), the 2017 rain amount is in agreement with the mean annual value, and precisely 600 mm; on the other hand, the 2018 rain shows a significant different amount, reaching about 1100 mm. For the same time span, the temperature follows the seasonal trend with lower values from November 2017 (8.9 °C) to February 2018 (5.9 °C), and maximum between July and August

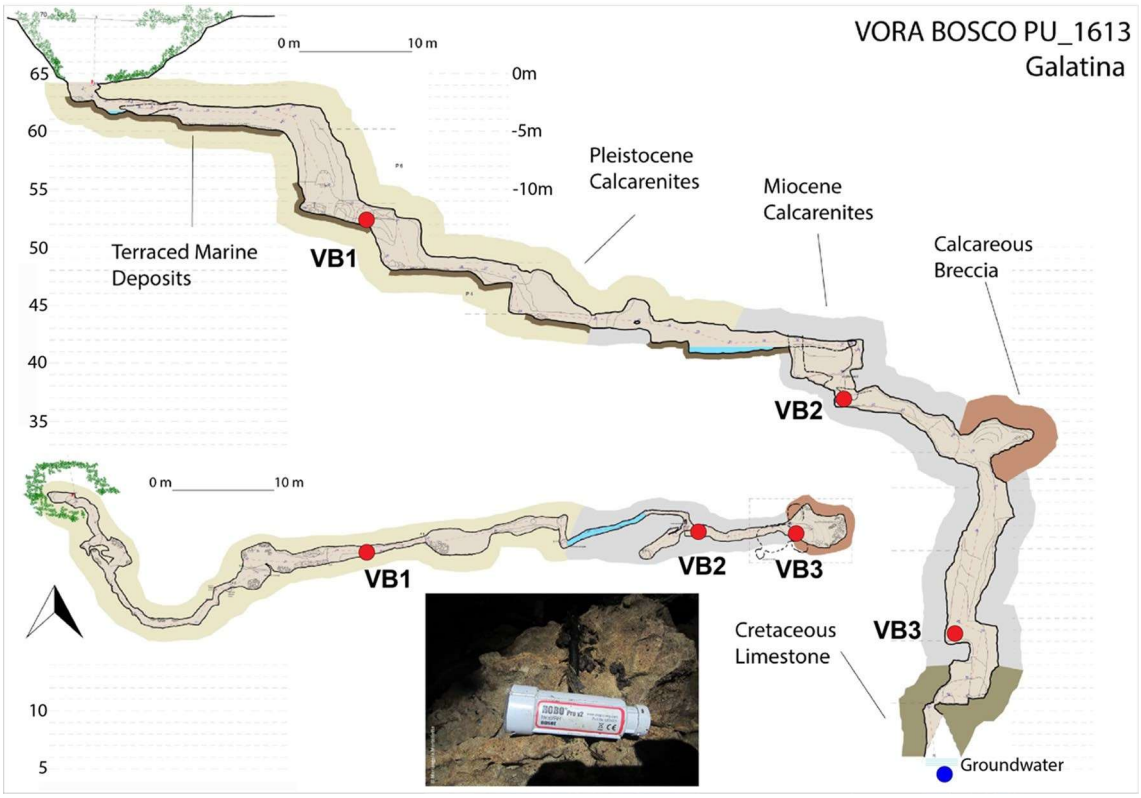


Fig. 4 Schematic map and section of Vora Bosco cave (survey by the Apulian Speleological Federation), showing location of the Hobo sensors (red circles). The multi-parameter probe was installed below the water table, in correspondence of the red blue circle

2018 (31.8 °C). The rainiest months were February 2018 (217 mm over 14 rainy days) and October 2018 (351 mm over 8 rainy days). The most significant rain events were registered between February 14 and 26 (total rain 144 mm) and on October 22–23, 2018 (total rain 210.60 mm). In particular, on the occasion of the February 25–26, 2018 rainstorm, during which about 60 mm of cumulative rainfall were recorded, the water table rose more than 10 m in less than one day.

All the cave equipment was installed within the framework of a project funded by Apulia Region (Parise et al. 2020). Unfortunately, after the severe October 23, 2018, rainstorm, and the consequent flash flood, all instrumentations installed inside the cave were lost, due to the huge amount of water and rocks that invaded the underground spaces. The whole area at the surface was flooded (Fig. 5) for some days, before being naturally drained by Vora Bosco and other nearby swallow holes.

As registered by the probe, the piezometric head at Vora Bosco fluctuates generally between 4 and 4.5 m. In some specific conditions, during and after severe rainstorms, the freshwater level inside the cave can significantly move up. On the occasion of the February 25–26, 2018 rainstorm (grey line in Fig. 6), during which about 60 mm of cumulate rainfall were recorded, the water table rose more than 10 m (Fig. 6). In addition to this original data time-series, 3 more months of water level data became available later on. This additional data is used as a testing period. Figure 6 shows the extended time-series of observed water level at Vora Bosco. The dotted line marks the division between the original period and the testing period, the grey line marks the large event on February 25–26 2018. Hourly precipitation and potential evapotranspiration data from the nearby Galatina gauge station are used for modeling.

The simulation model

A model similar to Fleury et al. (2007) was developed, based on the most dominant processes of the karst system. Its lumped, process-oriented structure consists of three reservoirs, filled and emptied depending upon specific thresholds, which is based on the comparison of the water level at Vora Bosco with the simulated groundwater storage. Likely to Fleury et al. (2007), water storage in the reservoirs is represented as a water column. The basic model structure is illustrated in Fig. 7 and the parameters of the model are displayed in Table 1. Due to the limited data availability, a modeling approach with a low degree of freedom was chosen and the

Fig. 5 Vora Bosco area **a** in normal conditions, and **b** totally flooded after the rainstorm which occurred on October 23, 2018

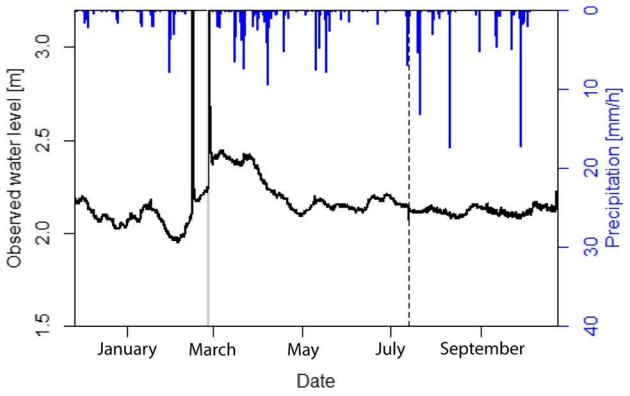


Fig. 6 Available time-series of water level observations and precipitation, dashed line marks the end of the original time-series and the beginning of the model testing period, grey line marks the big event on February 25–26

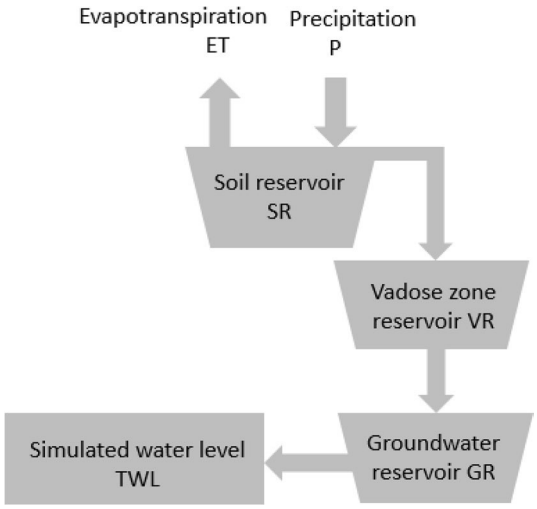


Fig. 7 Model structure with the overflow soil reservoir and the bucket vadose zone and groundwater reservoirs

model only uses 3 calibration parameters. In a preliminary test other model structures were tested, too, including separate conduit reservoirs, but they did not provide superior results than this parsimonious model.

Hourly data of precipitation and potential evapotranspiration are used as the model inputs. For a 5 year warm-up period of the model as well as for the testing period, evaporation data from the GLEAM dataset (Miralles et al. 2011; Martens et al. 2017; Gleam 2021) and precipitation data from MSWEP (multi-source weighted-ensemble precipitation) (Beck et al. 2019; GloH2O 2021) are used.

Precipitation input P [mm/h] goes to the soil reservoir SR [mm], which is an overflow reservoir that stores water until a threshold, the overflow capacity OC [mm], is exceeded. The surplus water then drains to the vadose zone reservoir VR [mm]. Furthermore, water leaves the reservoir with evapotranspiration. To consider soil saturation

Table 1 Model parameter descriptions and units

Parameter		Unit
OC	Overflow capacity	[mm]
VC	Vadose zone storage constant	[d]
GC	Groundwater storage constant	[d]

in this process, potential evapotranspiration ET_p [mm/h] is transformed to an estimate of actual evapotranspiration ET_a [mm/h] according to Eq. 2. The amount of water that is stored in the soil reservoir SR [mm] is described by Eq. 1 and the recharge from the soil reservoir to the vadose zone reservoir VR [mm] is calculated according to Eq. 3:

$$SR_{(t)} = \max \left[\min \left[SR_{(t-1)} + P_{(t)} - ET_{a(t)} \right], OC \right], \quad (1)$$

$$ET_{a(t)} = ET_{p(t)} * \frac{SR_{(t-1)}}{OC}, \quad (2)$$

$$RT_{(t)} = \max \left[SR_{(t-1)} + P_{(t)} - ET_{a(t)} - OC, 0 \right]. \quad (3)$$

The vadose zone is represented in the model as a linear reservoir. In this simple and commonly used approach, discharge from the reservoir is proportional to the stored water volume in the reservoir (Maillet 1905; Dewandel et al. 2003).

The vadose zone reservoir VR is fed by the soil reservoir and drains to the groundwater reservoir GR [mm] depending on the storage coefficient of the vadose zone VC [d]. The storage of water in the vadose zone reservoir VR is described in Eq. 4:

$$VR_{(t)} = VR_{(t-1)} + RT_{(t)} - \frac{VR_{(t-1)}}{VC * 24}. \quad (4)$$

The groundwater reservoir GR (Eq. 5) represents the long-term storage of the system. It is also modeled as a linear reservoir and drains according to the groundwater storage coefficient GC [d]:

$$GR_{(t)} = GR_{(t-1)} + \frac{VR_{(t-1)}}{VC * 24} - \frac{GR_{(t-1)}}{GC * 24}. \quad (5)$$

Since the model simulates stored water volume in the groundwater reservoir GW [mm] and the groundwater level observations are given in level [m], a direct comparison is not possible. Instead, the similarity of simulations and observations are compared through their linear correlation. The objective function to evaluate the goodness of fit of modeled groundwater storage and observed water level is the squared correlation coefficient R^2, based on the Pearson correlation coefficient r (Eq. 6) (Dormann 2013):

$$r = \frac{\sum_{i=1}^{n} \left(x_{i1} - \bar{x}_1 \right) \left(x_{i2} - \bar{x}_2 \right)}{\sqrt{\sum_{i=1}^{n} \left(x_{i1} - \bar{x}_1 \right)^2 \sum_{i=1}^{n} \left(x_{i2} - \bar{x}_2 \right)^2}}. \quad (6)$$

With x_{i1} and x_{i2} being the simulated and observed values and $\bar{x}_1$ and $\bar{x}_2$ being their mean values, a correlation coefficient near zero means no correlation, whereas a value near one indicates a strong correlation between the two variables (Hörmann 2016).

In order to analyze the uncertainty of the model predictions, a Generalized Likelihood Uncertainty Estimation (GLUE) after Beven and Binley (1992) is performed, using the previously defined behavioral Monte Carlo runs. For all behavioral parameter combinations, their model outputs are computed and out of those, 95% uncertainty bands are calculated.

Evaluation of the prediction skills of the model using different calibration datasets

In order to test the influence of the calibration period on the prediction skills of the model, different calibration–validation approaches are compared and evaluated. The model is calibrated on selected parts of the original dataset while the remaining part is then used for validation. As the available time-series does not even comprise one year, it was decided that the dataset is not sufficiently long in order to be split in half for calibration and validation (Klemeš 1986). Therefore, two-thirds of the original dataset are used as a calibration dataset, while the validation dataset includes the other third of the data. The different datasets used for calibration can be distinguished in 2 groups, the continuous calibrations (c1–c4), where different continuous time-series are used for model calibration, and the bootstrapping calibrations (b1–b10). In the bootstrapping calibrations, two-thirds of the dataset are randomly chosen for calibration and the remaining one third is used for validation. The random division into calibration and validation is executed 10 times, leading to 10 different calibration and validation results, which can then be compared. Table 2 shows an overview of the different calibration datasets.

After being calibrated and validated on the original time-series, the calibrated model is additionally tested on the testing period (see Section "Study site and data"), which was used neither for calibration nor for validation.

Table 2 Calibration datasets

Calibration dataset	Description
c1	Calibration on the first two-thirds of the original time series
c2	Calibration on the last two-thirds of the original time series
c3	Calibration on the first and last third of the original time series
c4	Calibration on the whole original time series
b1–b10	Ten bootstrapping rounds with random pick of twothirds of the original time series

Parameter and uncertainty estimation

For each of the calibration datasets (see previous section), a calibration of the model parameters is performed by executing a Monte Carlo analysis with 10,000 runs. In each run, the three model parameters are randomly picked from a uniform distribution between defined parameter ranges (Table 3). For each run, the correlation between simulated groundwater storage and observed water level at Vora Bosco Cave is quantified by calculating the objective function R^2 for the calibration period (Beven 2008). The parameter sets that provided the best objective functions for the calibration period are selected. The two main water level peaks of the time series (Fig. 6) are excluded from calibration as during these events rainwater enters the karst system through the cave opening and reaches the water table in short time.

To analyze the influence of the parameters on the model output and the model performance, the results of the Monte Carlo analysis are used to perform a sensitivity analysis on the parameters. This is done according to the approach of Hornberger, Spear & Young (HSY, Spear and Hornberger 1980; Hornberger and Spear 1981). The Monte Carlo runs are divided into a behavioral and a non-behavioral group, which are separated at the threshold of the correlation coefficient $r = 0.8$ and therefore, $R^2 = 0.64$ based on the calibration period. The cumulative probability density function of the behavioral group can be plotted over the respective parameter calibration range, indicating whether the model parameters can be considered sensitive parameters (Beven 2008). A strong deviation of a parameter's cumulative probability density function from the 1:1 line in this plot indicates a strong sensitivity of the parameter. When the cumulative probability density function plots close to the 1:1 line on the other hand, low parameter sensitivity is indicated.

Table 3 Parameter ranges for model calibration

Parameter	Calibration range	Unit
Overflow capacity *OC*	0–50	[mm]
Vadose zone storage constant *VC*	1–35	[d]
Groundwater storage constant *GC*	1–500	[d]

Results and discussion

Estimated parameters and their uncertainty

Table 4 shows the results of the analysis for the different calibration datasets. The mean and standard deviation of the calibrated parameter values and achieved model efficiencies in different periods are calculated based on all behavioral runs of the respective calibration dataset. For calibration dataset c3, no parameter combination of the Monte Carlo

Table 4 Calibration results for parameter values and model efficiencies

Parameter	*OC*	*VC*	*GC*	R^2_cal	R^2_val	R^2_test
Unit	mm	d	d	–	–	–
Description	Overflow capacity	Vadose zone storage constant	Groundwater storage constant	R^2 efficiency calibration period	R^2 efficiency validation period	R^2 efficiency testing period
Calibration range	0—50	1—35	1—500	–	–	–
Calibrated parameter values or efficiencies for different calibration datasets						
c1	25.06 ± 14.53	7.64 ± 6.93	189.89 ± 136.49	0.72 ± 0.07	0.24 ± 0.04	0.06 ± 0.05
c2	20.47 ± 14.61	13.38 ± 9.55	30.81 ± 28.73	0.70 ± 0.05	0.38 ± 0.09	0.14 ± 0.03
c3	NA	NA	NA	NA	NA	NA
c4	17.02 ± 13.11	3.01 ± 1.69	158.97 ± 38.77	0.68 ± 0.03	NA	0.06 ± 0.02
b1	15.01 ± 12.51	3.18 ± 1.80	150.74 ± 37.30	0.68 ± 0.04	0.70 ± 0.03	0.06 ± 0.02
b2	17.50 ± 13.31	2.75 ± 1.42	164.14 ± 40.07	0.67 ± 0.03	0.68 ± 0.03	0.06 ± 0.03
b3	18.35 ± 12.90	3.02 ± 1.63	162.35 ± 39.88	0.67 ± 0.03	0.67 ± 0.03	0.06 ± 0.02
b4	18.31 ± 13.37	3.17 ± 1.70	161.64 ± 39.40	0.67 ± 0.03	0.67 ± 0.03	0.06 ± 0.02
b5	14.51 ± 11.84	3.27 ± 1.79	155.35 ± 39.92	0.68 ± 0.03	0.68 ± 0.03	0.06 ± 0.03
b6	17.80 ± 13.51	2.95 ± 1.79	158.02 ± 35.73	0.68 ± 0.03	0.67 ± 0.03	0.06 ± 0.03
b7	19.63 ± 13.23	3.22 ± 1.78	159.29 ± 39.19	0.67 ± 0.03	0.66 ± 0.03	0.05 ± 0.02
b8	17.15 ± 12.61	3.11 ± 1.75	161.70 ± 42.77	0.68 ± 0.03	0.67 ± 0.04	0.06 ± 0.03
b9	17.40 ± 13.26	3.02 ± 1.53	156.62 ± 38.76	0.68 ± 0.03	0.66 ± 0.04	0.06 ± 0.03
b10	18.05 ± 13.32	3.03 ± 1.61	156.89 ± 36.16	0.68 ± 0.03	0.66 ± 0.04	0.06 ± 0.03

analysis could achieve a sufficient model efficiency. Therefore, there are no behavioral runs for calibration dataset c3, which will not be shown in the following results.

The analysis shows heterogeneous results for the efficiency values of c1, c2 and c4. For the calibration period, c1 and c2 provide the best performance. The efficiencies for the calibration period of c4 and b1 to b10 are slightly lower. While there is no efficiency value for the validation period of c4, as all data were used for calibration, the validation period efficiencies for b1–b10 are at similar values as their calibration period efficiencies. For c1 and c2, the performance in the validation period is drastically lower than the performance in the calibration period. In the testing period, all calibration datasets show very low performances. While c2 shows the best testing period efficiency, the performance is still comparably poor. Compared to c4 and b1 to b10, c1 and c2 also show strongly different values for the calibrated model parameters. The results of calibration datasets b1 to b10 resemble the results of c4, both in terms of performances in the calibration and validation period and in the calibrated parameter values.

Figure 8 shows the cumulative distributions of the different calibration datasets derived from the HSY method (see Section "Parameter and uncertainty estimation"). For the bootstrapping calibrations, the range of all bootstrapping rounds is illustrated as a grey band. For all three parameters, especially for the vadose zone storage constant *VC*, the range of the different bootstrapping rounds is very narrow and close to the cumulative distribution of c4. The cumulative distributions of the continuous calibration datasets in Fig. 8 differ a lot from each other. For the overflow capacity *OC* the cumulative distribution of c1 hardly deviates from the 1:1 line, indicating low parameter sensitivity. The cumulative distribution of c4 deviates most from the 1:1 line for *OC* . The vadose zone storage constant *VC* appears to be the most sensitive using calibration dataset c4 compared to c1 and c2. For the groundwater storage constant *GC* , while calibration method c1 deviates the least from the 1:1 line, both c2 and c4 indicate high parameter sensitivity, but there is a large offset between the two curves. This shows that parameter *GC* can be considered sensitive in both calibration methods, but with another ideal parameter value. In general, the cumulative distribution range of the bootstrapping calibrations most closely resembles the cumulative distribution of the continuous dataset c4, where the whole original time-series is used for calibration.

The developed model is a very simple model, yet its reservoir structure is based on actual processes that occur in karst aquifers. It can therefore be described as a simple lumped process-oriented reservoir model (Hartmann et al. 2014). In most cases, the sensitivity analysis show at least some distinction of the parameters' cumulative distributions from the 1:1 line, indicating sensitivity (Hornberger and Spear 1981). Good model results can thus only be achieved if all model

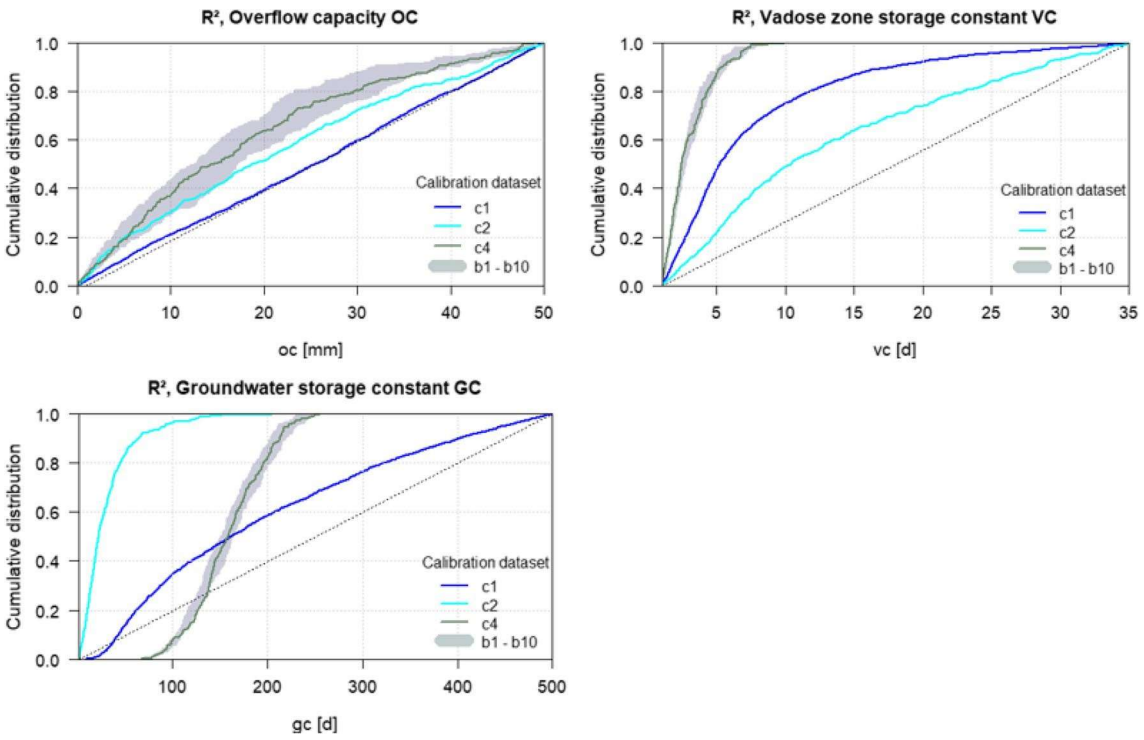


Fig. 8 HSY parameter sensitivity analysis of all model parameters for different calibration datasets

parameters are within a certain range, so with the simple model structure equifinality could be avoided (Beven 2008).

In the calibration of the three model parameters using the different calibration datasets, the parameter *OC* appears to be the least sensitive. However, for c2, c4 and b1 to b10 there is a tendency towards lower values for this parameter. The overflow capacity is the model parameter of the soil storage. Lower values of *OC* therefore indicate that there is only a very small storage function of the soil in the system. This means that even at very small precipitation events, recharge can be generated. This result of the model is reflected in the system characteristics of the Vora Bosco Cave surrounding territory. The soil thickness is approximately 2 m; immediately below, Marine Terraced deposits are present. The Tertiary and Quaternary clastic carbonate deposits, affected by karst processes, together with the calcarenite permeability, do not allow water to be stored within the soil and carbonate succession, and a rapid transferring of water volume occurs. The calibrated storage constants for the vadose zone and the groundwater are overall in a reasonable magnitude. However, the different calibration datasets lead to very different calibrated parameter values. This discrepancy is related to the difference in calibration data and indicates that different information about the karst system are contained in different parts of the groundwater level observation time-series. This is also reflected in the different parameter sensitivities.

The different parameters represent different processes taking place in the karst system. For instance, the sensitivity analysis of c1 and c2 show that the vadose zone storage constant *VC* is more sensitive for c1 than for c2. The groundwater storage constant *GC* on the other hand is more sensitive for c2 than for c1. This sensitivity reflects the relevance of the parameters, and therefore the processes they represent for the respective calibration time-series. For example, parameter *GC* represents the long-term storage in the system. While the calibration dataset c1 contains the winter months and therefore the wet season, c2 contains the summer months and therefore the dry season. During the dry season, the long-term storage becomes a more important component of the system and thus, the sensitivity of the representative parameter increases. The available dataset was split into three thirds for the continuous calibration dataset variations. From the results of the respective calibrations it seems that the chronologically 2nd third is the most informative part of the time-series as there are no behavioral Monte Carlo runs for c3, where this 2nd third is not considered for calibration. Furthermore, the calibration with the c2 dataset leads to the best performance for the model testing period. For c2, the 2nd and 3rd third of the original time-series are included in the calibration. Therefore, the data included in c2 are closest to the testing period, which might be beneficial. This also relates to the findings of Shen et al. (2022), where it is shown that the widely used practice of using older data for calibration leads to poorer performances in the model testing period. In addition, this could indicate that the 1st third of the original time-series is less informative and could be related to problems in identifying the correct initial conditions.

The calibration dataset c4 shows the most sensitive results in the sensitivity analysis for all parameters. As the whole original period is included in this calibration dataset, the most information is contained in it, leading to sensitive parameters. The similarity of the results of the b1–b10 calibrations to the c4 results shows that the bootstrapping datasets reflect the information of the whole dataset.

Prediction skills of the model using the different calibration datasets

Figures 9, 10, and 11 show the standardized observed water level and the standardized model results of the calibration datasets c1, c2 and b1. The modeled time-series of calibration datasets c4 and b2 to b10 can be found in the appendix. The model run with the highest objective function is displayed in blue and the whole GLUE uncertainty band as well as the GLUE 95% confidence intervals are shown in light grey and dark grey. For the continuous calibration datasets, the calibration period is highlighted in blue and the validation period in yellow. The dotted line marks the end of the original time-series and the beginning of the testing period. Even though the full extended time-series is shown, the observed and simulated time-series are standardized based only on the original time series.

The modeled time-series of calibration dataset b1 (Fig. 11) shows the narrowest GLUE uncertainty bands compared to calibration dataset c1 (Fig. 9) and calibration dataset c2 (Fig. 10). During the original time-series, most of the observed data lie within the uncertainty bands. The performance or accuracy of the model based on calibration

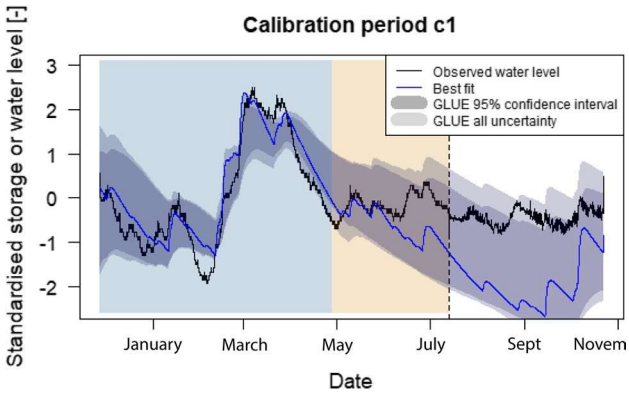


Fig. 9 Time-series of standardized observed and simulated water levels using the c1 calibration dataset including GLUE uncertainty bands (calibration period in blue, validation period in yellow, model testing period in white)

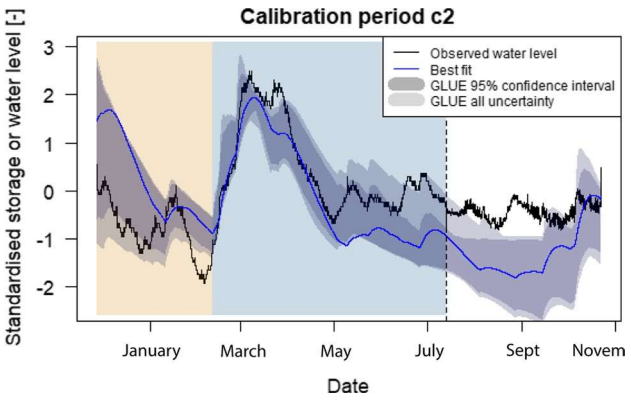


Fig. 10 Time-series of standardized observed and simulated water levels using the c2 calibration dataset including GLUE uncertainty bands (calibration period in blue, validation period in yellow, model testing period in white)

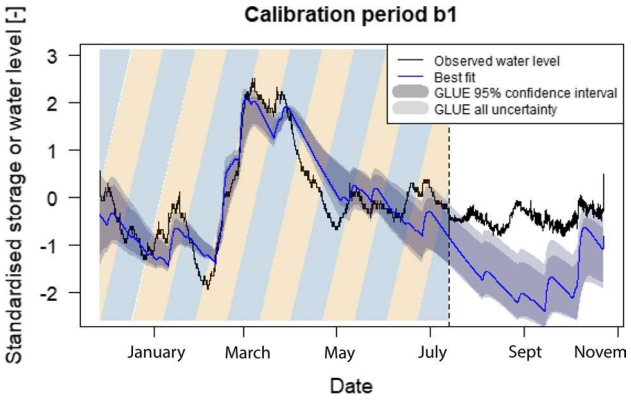


Fig. 11 Time-series of standardized observed and simulated water levels using the b1 calibration dataset including GLUE uncertainty bands (bootstrapping period in blue/yellow, model testing period in white)

dataset b1 is therefore high in the calibration and validation period. Further, the precision of b1 is higher compared to the models based on c1 and c2. To some extent this supports the findings of Shen et al. (2022), as the model precision is improved by the use of more information for calibration. For the model testing period however, the observed data lies outside of the uncertainty bands. The time-series of calibration datasets c4 and b2 to b10 (Online Appendix) show very similar results to calibration dataset b1. For c2, the performance in the original period is a bit poorer and the precision in the calibration and validation period is comparably low. However, the modeled uncertainty describes the observed data in the model testing period much better, indicating more robust calibrated parameters. The achieved lower model precision by the split-sample approach therefore turns out to be more realistic here. It therefore seems that c2 with a shorter calibration time period, which envelops the highest peak, provides the more robust predictions. Regarding model parameterization, already Seibert and Beven (2009) have shown that a lot of the information content of a hydrological time-series can be contained in only a few measurements. In the available data time-series, it can therefore be assumed that most of the important information for model calibration is contained in the second third of the original time-series, which includes the main peak. As mentioned before, calibration dataset c3, which does not include this part of the time-series, does not lead to behavioral parameter sets at all. Also Reynolds et al. (2020) showed in the context of flood prediction that predictions improved already when only one event was included in the calibration. They further state that prediction accuracy improves considerably when two to four events are included. However, they found that beyond this number the increase in model performance was only small. This relates to the short calibration dataset c2, which however envelops the most important information of the time-series.

Another advantage of the c2 calibration dataset is that it does not contain data from the first part of the time-series and therefore is not subjected to problems related to antecedent conditions and model warm-up. Since the observed data record begins on the falling limb of a medium size event, it can be difficult for the model to start with the correct initial conditions despite the warm-up period. This phenomenon was also described by Mazzilli et al. (2012), where they show that despite the warm-up period, initialization bias can have a significant impact on calibration results. The poorer testing period performance of calibration datasets, which include the first part of the original time-series, might therefore be partly related to problems with initial conditions.

Considering the full analysis including the evaluation in the testing period, despite the parsimony of the model none of the calibration datasets could lead to a combination of parameters that can provide a robust prediction of groundwater dynamics. The model is therefore rejected and not used for future predictions of groundwater dynamics. Only few modeling studies, and none in karst hydrology, have documented such a model rejection process before (Choi and Beven 2006; Chang et al. 2021; Hartmann et al. 2013).

Conclusions

In this study, a very simple reservoir model was developed to simulate the groundwater level at Vora Bosco Cave, and different calibration datasets were used for its calibration. Even though the overall dataset is limited, useful considerations may be drawn from this work, which might be further validated in the future, taking advantage of the availability of more extensive datasets covering longer periods. In our case study, the resulting calibrated models were

compared not only in the calibration and validation period, but also in an additional model testing period. By combining the model performance evaluation with regional sensitivity analysis, additional information on the identifiability of the parameters and the robustness of the predictions for the model testing period could be gained. The parameters are uncertain, however some interpretations are possible due to the more or less pronounced confinement of the established uncertainty ranges. For the original time-series, all calibration datasets could lead to good performances in the calibration period. For the validation period of the original time-series, the best performances could be achieved using the bootstrapped calibration datasets or the whole original time-series for calibration. The continuous split-sample datasets led to much poorer validation performances and also showed overall a higher model uncertainty and therefore a lower model precision. However, due to the more moderate precision and the wider uncertainty ranges, the continuous split-sample calibration datasets provided more robust predictions with realistic uncertainty ranges in the model testing periods. Furthermore, it was found that problems in finding the initial conditions are most likely the reason why the calibration dataset containing the complete original time-series provided poorer results despite seemingly high precision.

While the initial intentions of this modeling project included using the model for the prediction of groundwater dynamics, the evaluation of the model showed that no parameter combination could be found that provides robust predictions. The model was therefore rejected based on this extensive analysis.

Overall, the results of the study also show that, especially for short datasets, the split-sample test is a justified concept. The poorer validation performances and the wider uncertainty ranges achieved by the split-sample calibration datasets may seem like a disadvantage regarding only the original time-series. However, taking the model testing period into consideration, they prove to be valid. In this context, only using all data for calibration, as it is sometimes recommended in literature, may be misleading and prove fatal to models that should be used for prediction. In addition, especially if the split-sample test is executed with switched calibration and validation periods, potential problems with the initialization bias can be identified.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12665-023-10984-2.

Acknowledgements Support to T.L. and A.H. was provided by the Emmy Noether Programme of the German Research Foundation (DFG; grant no. HA 8113/1-1; project "*Global Assessment of Water Stress in Karst Regions in a Changing World*"). The work by I.S. Liso and M. Parise was partly funded through the *Protocollo d'intesa con Regione Puglia per l'attuazione dell'art. 45 "Interventi per esplorazione dei fenomeni carsici," comma 1 della L.R. n. 45 del 30/12/2013*. We also thank the cavers Mariangela Martellotta and Francesco De Salve for cave survey and support during the work within the karst system. fds

Author contributions TL and AH developed the model, and contributed to writing the manuscript; ISL and MP provided data, performed initial analyses, and contributed to writing the manuscript; MP and AH supervised the work. All authors reviewed the manuscript.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests. Research data policy: The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alfio MR, Balacco G, Parisi A, Totaro V, Fidelibus MD (2020) Drought Index as Indicator of Salinization of the Salento Aquifer (Southern Italy). Water 12:1927. https://doi.org/10.3390/w12071927

Andriani GF, Walsh N (2010) Petrophysical and mechanical properties of soft and porous building rocks used in Apulian monuments (south Italy). In: Prikryl R, Torok A (eds) Natural stone resources for historical monuments, vol Special Publication 333. Geological Society of London, pp 129–141

Andriani GF, Pastore N, Giasi CI, Parise M (2021) Hydraulic properties of unsaturated calcarenites by means of a new integrated approach. J Hydrol 602:126730

Arsenault R, Brissette F, Martel JL (2018) The hazards of split-sample validation in hydrological model calibration. J Hydrol 566(31):346–362

Beck HE, Wood EF, Pan M, Fisher CK, Miralles DM, van Dijk AIJM, McVicar TR, Adler RF (2019) MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull Am Meteorol Soc 100(3):473–500

Beven K (2008) Rainfall-runoff modelling: the primer. Wiley (**reprint**)

Beven K, Binley A (1992) The future of distributed models: Model calibration and uncertainty prediction. Hydrol Process 6(3):279–298. https://doi.org/10.1002/hyp.3360060305

Bosellini A, Bosellini FR, Colalongo ML, Parente M, Russo A, Vescogni A (1999) Stratigraphic architecture of the Salento coast from Capo d'Otranto to S. Maria di Leuca (Apulia, southern Italy). Riv Ital Paleontol Stratigr 105:397–416

Cantonati M, Poikane S, Pringle CM, Stevens LE, Turak E, Heino J, Richardson JS, Bolpagni R, Borrini A, Cid N, Ctvrtlikova M, Galassi DMP, Hajek M, Hawes I, Levkov Z, Naselli Flores L, Saber AA, Di Cicco M, Fiasca B, Hamilton PB, Kubecka J, Segadelli S, Znachor P (2020) Characteristics, main impacts, and stewardship of natural and artificial freshwater environments: consequences for biodiversity conservation. Water 12(1):260. https://doi.org/10.3390/w12010260

Chang Y, Hartmann A, Liu L, Jiang G, Wu J (2021) Identifying more realistic model structures by electrical conductivity observations of the karst *spring*. Water Resour Res. https://doi.org/10.1029/2020WR028587

Choi HT, Beven K (2006) Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of

TOPMODEL within the GLUE framework. J Hydrol 332:316–336. https://doi.org/10.1016/j.jhydrol.2006.07.012
Cotecchia V (2014) Le acque sotterranee e l'intrusione marina in Puglia: dalla ricerca all'emergenza nella salvaguardia della risorsa. Mem Descr Carta Geol It 92(1):1–511 (**ISBN: 978-88-9311-003-7**)
D'Angeli IM, Vattano M, Parise M, De Waele J (2017) The coastal sulfuric acid cave system of Santa Cesarea Terme (Southern Italy). In: Klimchouk A et al (eds) Hypogene karst regions and caves of the world. Springer, Cave and Karst Systems of the World. https://doi.org/10.1007/978-3-319-53348-3_9
D'Angeli IM, Ghezzi D, Leuko S, Firrincieli A, Parise M, Fiorucci A, Vigna B, Addesso R, Baldantoni D, Carbone C, Mill AZ, Jurado V, Saiz-Jimenez C, De Waele J, Cappelletti M (2019) Geomicrobiology of a seawater-influenced active sulfuric acid cave. PLoS ONE 14(8):e0220706. https://doi.org/10.1371/journal.pone.0220706
D'Angeli IM, De Waele J, Fiorucci A, Vigna B, Bernasconi SM, Florea LJ, Liso IS, Parise M (2021) Hydrogeology and geochemistry of the sulfur karst springs at Santa Cesarea Terme (Apulia, southern Italy). Hydrogeol J 29:481–498. https://doi.org/10.1007/s10040-020-02275-y
Deharveng L, Stoch F, Gibert J, Bedos A, Galassi D, Zagmajster M, Brancelj A, Camacho A, Fiers F, Martin P, Giani N, Magniez G, Marmonier P (2009) Groundwater biodiversity in Europe. Freshw Biol 54:709–726. https://doi.org/10.1111/j.1365-2427.2008.01972.x
Delle Rose M, Federico A, Parise M (2004) Sinkhole genesis and evolution in Apulia, and their interrelations with the anthropogenic environment. Nat Hazards Earth Syst Sci 4:747–755
Dewandel B, Lachassagne P, Bakalowicz M, Weng P, Al-Malki A (2003) Evaluation of aquifer thickness by analysing recession hydrographs. Application to the Oman ophiolite hard-rock aquifer. J Hydrol 274(1–4):248–269
Dormann CF (2013) Parametrische Statistik: Verteilungen, maximum likelihood und GLM. In: R. Statistik und ihre Anwendungen. Springer. https://doi.org/10.1007/978-3-642-34786-3
Festa V, Fiore A, Parise M, Siniscalchi A (2012) Sinkhole evolution in the Apulian Karst of Southern Italy: a case study, with some considerations on Sinkhole Hazards. J Cave Karst Stud 74(2):137–147
Festa V, Fiore A, Miccoli MN, Parise M, Spalluto L (2015) Tectonics versus karst relationships in the Salento peninsula (Apulia, Southern Italy): implications for a comprehensive land-use planning. In: Lollino G, Manconi A, Guzzetti F, Bobrowsky MCP, Luino F (eds) Engineering geology for society and territory, vol 5, Urban geology, sustainable planning and landscape exploitation. Springer. https://doi.org/10.1007/978-3-319-09048-1
Fleury P, Plagnes V, Bakalowicz M (2007) Modelling of the functioning of karst aquifers with a reservoir model: application to Fontaine de Vaucluse (South of France). J Hydrol 345(1–2):38–49. https://doi.org/10.1016/j.jhydrol.2007.07.014
Ford D, Williams P (2007) Karst hydrogeology and geomorphology. Wiley (**ISBN: 978-0-470-84997-2**)
Funiciello R, Montone P, Parotto M, Salvini F, Tozzi M (1991) Geodynamical evolution of an intra-orogenic foreland: the Apulia case history (Italy). Boll Soc Geol Ital 110(3/4):419–425
Gleam (2021) Downloads. https://www.gleam.eu/. Accessed 9 Oct 2022
GloH2O (2021) MSWEP. Multi-source weighted-ensemble precipitation. http://www.gloh2o.org/mswep/. Accessed 9 Oct 2022
Goldscheider N, Drew D (2007) Methods in karst hydrogeology. Taylor and Francis Group, International contribution to hydrogeology, London, p 26 (**ISBN 13:978-0-415-428723-6**)
Gutierrez F, Parise M, De Waele J, Jourde H (2014) A review on natural and human-induced geohazards and impacts in karst. Earth-Sci Rev 138:61–88. https://doi.org/10.1016/j.earscirev.2014.08.002
Hartmann A, Wagener T, Rimmer A, Lange J, Brielmann H, Weiler M (2013) Testing the realism of model structures to identify karst system processes using water quality and quantity signatures. Water Resour Res 49:3345–3358. https://doi.org/10.1002/wrcr.20229,2013
Hartmann A, Goldscheider N, Wagener T, Lange J, Weiler M (2014) Karst water resources in a changing world: Review of hydrological modeling approaches. Rev Geophys 52:218–242. https://doi.org/10.1002/2013RG000443
Hollingsworth E, Brahana V, Inlander E, Slay M (2008) Karst Regions of the World (KROW), Global karst datasets and maps to advance the protection of karst species and habitats worldwide. USGS Scientific Investigations Report 2008–5023
Hörmann G (2016) Hydrologische Modelle. In: Fohrer N, Bormann H, Miegel K, Casper M, Bronstert A, Schumann A, Weiler M (eds) Utb basics: Vol. 4513 Hydrologie, 1st edn. Haupt Verlag
Hornberger GM, Spear RC (1981) An approach to preliminary analysis of environmental systems. J Environ Manag 12:7–18
Inguscio S, Rossi E, Parise M (2009) Biogeographical distribution of subterranean fauna in Apulia (Italy) in the context of the palaeogeographic evolution of the area. In: Proc 15th Int Congr Speleology, Kerrville (Texas, USA), 2:749–754
Jeannin PY, Groves C, Hauselmann P (2007) Speleological investigations. In: Goldscheider N, Drew D (eds) Methods in karst hydrogeology, vol 26. Taylor and Francis, Int Contributions to Hydrogeology, pp 25–44
Klemeš V (1986) Operational testing of hydrological simulation models. Hydrol Sci J 31(1):13–24. https://doi.org/10.1080/02626668609491024
Kresic N (2013) Water in karst: management, vulnerability, and restoration. McGraw Hill, New York
Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241. https://doi.org/10.1029/1998WR900018
Liso IS, Parise M (2020) Apulian karst springs: a review. J Environ Sci Eng Technol 8:63–83. https://doi.org/10.12974/2311-8741.2020.08.7
Liso IS, Loiotine L, Andriani GF, Parise M (2019) Apulian caves as natural hydrogeological laboratories. Ital J Engng Geol Environ 1:67–72
Liso IS, Chieco M, Fiore A, Pisano L, Parise M (2020) Underground geosites and caving speleotourism: some considerations, from a case study in Southern Italy. Geoheritage 12:13. https://doi.org/10.1007/s12371-020-00424-z
Maillet E (1905) Essais d'hydraulique souterraine et fluviale Librairie Sci. A. Hermann, Paris, p 218
Margiotta S, Negri S (2005) Geophysical and stratigraphical research into deep groundwater and intruding seawater in the Mediterranean area (the Salento peninsula, Italy). Nat Hazards Earth Syst Sci 5:127–136
Margiotta S, Mazzone F, Negri S (2010) Stratigraphic revision of Brindisi-Taranto Plain: hydrogeological implications. Mem Descr Carta Geol It 90:165–180
Margiotta S, Marini G, Fay S, D'Onghia FM, Liso IS, Parise M, Pinna M (2021) Hydro-stratigraphic conditions and human activity leading to development of a sinkhole cluster in a Mediterranean water ecosystem. Hydrology 8:111. https://doi.org/10.3390/hydrology8030111
Martens B, Miralles DG, Lievens H, van der Schalie R, de Jeu RAM, Fernández-Prieto D, Beck HE, Dorigo WA, Verhoest NEC (2017) GLEAM v3: satellite-based land evaporation and

root-zone soil moisture. Geosci Model Dev 10:1903–1925. https://doi.org/10.5194/gmd-10-1903-201
Masciopinto C, Liso IS (2016) Assessment of the impact of sea-level rise due to climate change on coastal groundwater discharge. Sci Total Environ 569–570:672–680. https://doi.org/10.1016/j.scitotenv.2016.06.183
Masciopinto C, Semeraro F, La Mantia R, Inguscio S, Rossi E (2006) Stygofauna abundance and distribution in the fissures and caves of the Nardò (Southern Italy) fractured aquifer subject to reclaimed water injections. Geomicrobiology J 23:267–278. https://doi.org/10.1080/01490450600760690
Masciopinto C, Liso IS, Caputo MC, De Carlo L (2017) An integrated approach based on numerical modelling and geophysical survey to map groundwater salinity in fractured coastal aquifers. Water 9:875. https://doi.org/10.3390/w9110875
Mazzilli N, Guinot V, Jourde H (2012) Sensitivity analysis of conceptual model calibration to initialisation bias. Application to karst spring discharge models. Adv Water Resour 42(2):1–16
Mikszewski A, Kresic N (2015) Mathematical modeling of karst aquifers. In: Stevanovic Z (ed) Karst aquifers—characterization and engineering. Springer, Professional Practice in Earth Sciences, pp 283–298
Miralles DG, Holmes TRH, de Jeu RAM, Gash JH, Meesters AGCA, Dolman AJ (2011) Global land-surface evaporation estimated from satellite-based observations. Hydrol Earth Syst Sci 15:453–469. https://doi.org/10.5194/hess-15-453-2011
Mongelli G, Buccione R, Sinisi R (2015) Genesis of autochthonous and allochthonous Apulian karst bauxites (Southern Italy): Climate constraints. Sedimen Geol 325:168–176. https://doi.org/10.1016/j.sedgeo.2015.06.005
Olarinoye T, Gleeson T, Marx V, Seeger S, Adinehvand R, Allocca V, Andreo B, Apaéstegui J, Apolit C, Arfib B, Auler A, Barberá JA, Batiot-Guilhe C, Bechtel T, Binet S, Bittner D, Blatnik M, Bolger T, Brunet P, Charlier JB, Chen Z, Chiogna G, Coxon G, De Vita P, Doummar J, Epting J, Fournier M, Goldscheider N, Gunn J, Guo F, Guyot JL, Howden N, Huggenberger P, Hunt B, Jeannin PY, Jiang G, Jones G, Jourde H, Karmann I, Koit O, Kordilla J, Labat D, Ladouche B, Liso IS, Liu Z, Massei N, Mazzilli N, Mudarra M, Parise M, Pu J, Ravbar N, Sanchez LH, Santo A, Sauter M, Sivelle V, Skoglund RØ, Stevanovic Z, Wood C, Worthington S, Hartmann A (2020) Global karst springs hydrograph dataset for research and management of the world's fastest- flowing groundwater. Sci Data 7:59. https://doi.org/10.1038/s41597-019-0346-5
Onorato R, Belmonte G, Costantini A (2006) Le grotte sommerse della costa neretina (Salento, SE Italia). Thalassia Salentina 29:39–54
Palmentola G, Vignola N (1980) Dati di neotettonica sulla Penisola salentina. Progetto Finalizzato Geodinamica CNR 356:173–202
Palmer AN (2007) Cave geology. Cave Books, p 454
Parise M (2008) I sinkholes in Puglia. Mem Descr Carta Geol Ital 85:309–334
Parise M (2019) Sinkholes. In: White WB, Culver DC, Pipan T (eds) Encyclopedia of caves, 3rd edn. Academic Press, Elsevier, pp 934–942 (**ISBN 978-0-12-814124-3**)
Parise M (2022) Sinkholes, subsidence and related mass movements. In: Shroder JJF (ed) Treatise on geomorphology, vol 5. Elsevier. Academic Press, pp 200–220. https://doi.org/10.1016/B978-0-12-818234-5.00029-8
Parise M, Federico A, Delle Rose M, Sammarco M (2003) Karst terminology in Apulia (southern Italy). Acta Carsolog 32(2):65–82
Parise M, Ravbar N, Živanovic V, Mikszewski A, Kresic N, Mádl-Szőnyi J, Kukuric N (2015) Hazards in karst and managing water resources quality. In: Stevanovic Z (ed) Karst aquifers—characterization and engineering. Professional practice in earth sciences. Springer, Berlin, pp 601–687. https://doi.org/10.1007/978-3-319-12850-4_17
Parise M, Gabrovsek F, Kaufmann G, Ravbar N (2018) Recent advances in karst research: from theory to fieldwork and applications. In: Parise M, Gabrovsek F, Kaufmann G, Ravbar N (eds) Advances in karst research: theory, fieldwork and applications, vol special publ. 66. Geological Society, London, pp 1–24
Parise M, Benedetto L, Chieco M, Fiore A, Lacarbonara M, Liso IS, Masciopinto C, Pisano L, Riccio A, Vurro M (2020) First outcomes of a project dedicated to monitoring groundwater resources in Apulia, Southern Italy. In: Bertrand C, Denimal S, Steinmann M, Renard P (eds) Eurokarst 2018, Besançon. Springer, Advances in Karst Science, pp 243–249. https://doi.org/10.1007/978-3-030-14015-1_27
Pepe M, Parise M (2014) Structural control on development of karst landscape in the Salento Peninsula. Acta Carsolog 43(1):101–114
Reynolds JE, Halldin S, Seibert J, Xu CY, Grabs T (2020) Robustness of flood-model calibration using single and multiple events. Hydrol Sci J 65(5):842–853. https://doi.org/10.1080/02626667.2019.1609682
Seibert J, Beven KJ (2009) Gauging the ungauged basin: how many discharge measurements are needed? Hydrol Earth Syst Sci 13:883–892
Shen H, Tolson BA, Mai J (2022) Time to update the split-sample approach in hydrological model calibration. Water Resour Res 58(3):135
Sorooshian S, Gupta VK, Fulton JL (1983) Evaluation of Maximum Likelihood Parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility. Water Resour Res 19:251–259. https://doi.org/10.1029/WR019i001p00251
Spear RC, Hornberger GM (1980) Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis. Water Res 14(1):43–49. https://doi.org/10.1016/0043-1354(80)90040-8
Stevanović Z (2018) Global distribution and use of water from karst aquifers. In: Parise M, Gabrovsek F, Kaufmann G, Ravbar N (eds) Advances in karst research: theory, fieldwork and applications, vol sp publ 66. Geological Society, London, pp 217–236. https://doi.org/10.1144/SP466.26
Stevanovic Z (2019) Karst water resources in a changing world: Karst waters in potable water supply: a global scale overview. Environ Earth Sci 78:662. https://doi.org/10.1007/s12665-019-8670-9
Stoch F, Artheau M, Brancelj A, Galassi DMP, Malard F (2009) Biodiversity indicators in European ground waters: towards a predictive model of stygobiotic species richness. Freshw Biol 54(4):745–755
Tulipano L, Fidelibus MD (2002) Mechanisms of groundwater salinisation in a coastal karstic aquifer subject to over-exploitation. Proc 17th SWIM, Delft (The Netherlands). ISBN 90-800089– 8–2, pp 262–272
Zumpano V, Pisano L, Parise M (2019) An integrated framework to identify and analyze karst sinkholes. Geomorphology 332:213–225