

Suitable sites for groundwater development: a capture map-based approach integrated with weighted overlay analysis

Mayank Bajpai^a, Shishir Gaur^a, Ranveer Kumar ^{a,*}, Anurag Ohri^a and Hervé Piégay^b

^a Department of Civil Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh 221005, India

^b ENS-Lyon 15 Parvis René Descartes BP 7000, Lyon 69342, France

*Corresponding author. E-mail: ranveerkumar.rs.civ20@iitbhu.ac.in

 RK, 0000-0001-7530-3581

ABSTRACT

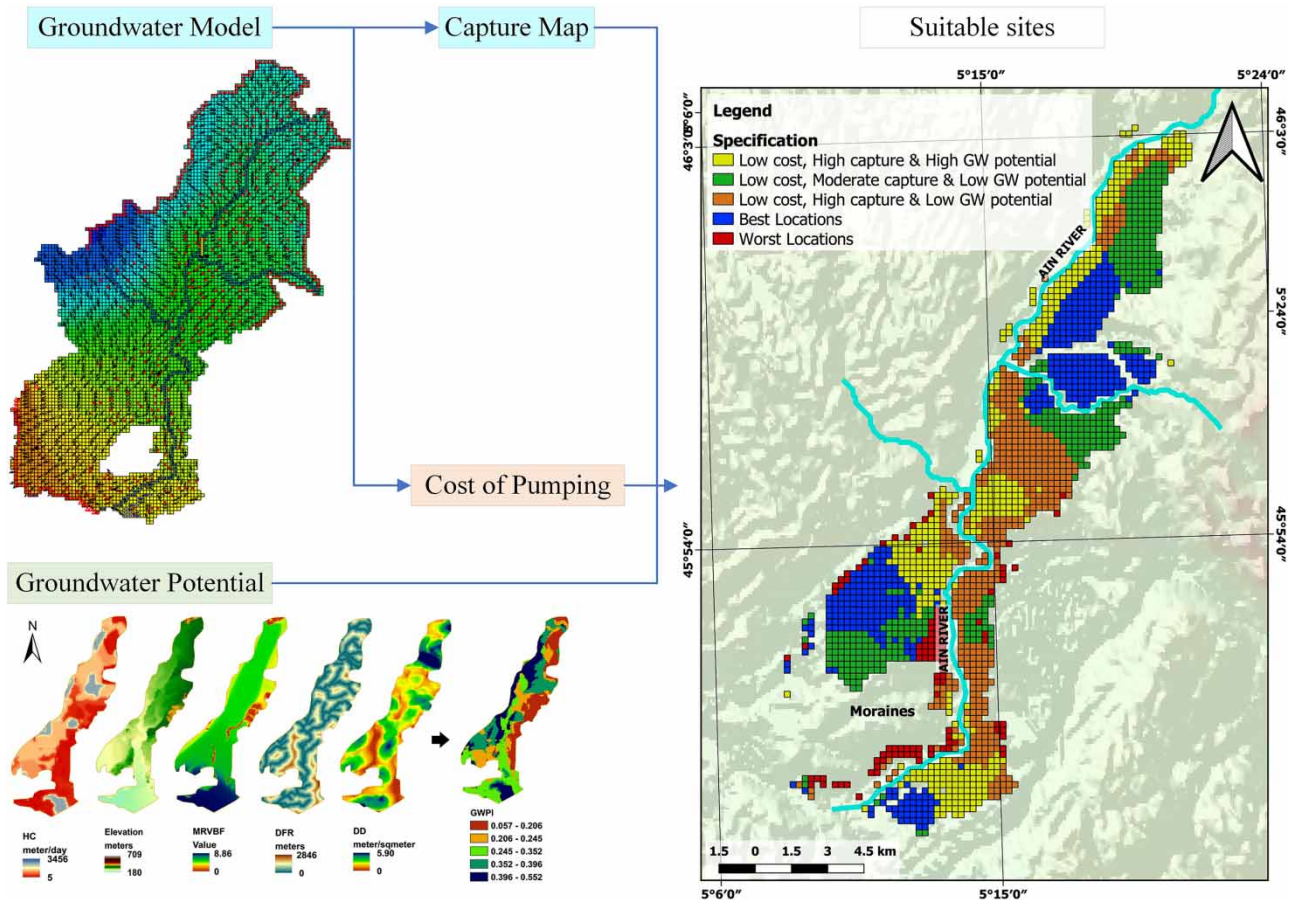
A novel methodology for suitable site selection for groundwater development based on river capture, pumping cost, and groundwater potential has been proposed for better groundwater utilisation. River capture and cost map have been generated from a calibrated groundwater model, simulated with forecasted hydrological time series data. The groundwater potential has been calculated with weighted overlay analysis. These three variables have been used to classify the model domain into five zones of groundwater development by K-Means clustering. The area with lower river capture, low cost of pumping, and high groundwater potential is found to be the best location for groundwater extraction. The methodology has been applied to the lower Ain River basin, in France.

Key words: forecasting, groundwater potential, LSTM, NAR, river capture fraction, water resources management

HIGHLIGHTS

- LSTM is a superior algorithm as compared to NAR for hydrological time-series forecasting.
- The capture map has been utilised as an aquifer response to pumping in groundwater potential zonation.
- Low proximity to streams does not justify high groundwater potential.

GRAPHICAL ABSTRACT



1. INTRODUCTION

One of the most significant and essential natural resources is groundwater (GW), which is kept in the underground geological formations of the earth's crust. It provides water for home, commercial, industrial, agricultural, and other development projects (Ayazi *et al.* 2010; Nampak *et al.* 2014). The presence and spread of GW are influenced by various anthropogenic and natural variables. Therefore, studying potential GW resources is essential for their wise development and usage (Bai *et al.* 2022). Studies on the effects of climate change indicate that there won't be much change in GW recharge in the foreseeable future (Crosbie *et al.* 2013; Taylor *et al.* 2013). The comparison of anthropogenic pumping and climate-driven factors showed that the decrease in GW resources is mainly caused by the combined effects of excessive pumping and climatic changes, albeit the contribution of pumping could potentially be much more significant than the natural replenishment (Wu *et al.* 2020). As the land and water resource development intensifies, it is becoming clear that the development of either GW or surface water affects the other. Therefore, it is crucial to identify the best locations for GW development projects to lessen the strain on surface water resources and GW storage.

GW potential zonation (determination of best locations for GW extraction) methodologies have come a long way with the use of physical methods (like borehole drilling and pumping tests), weighted overlay analysis (WOA) (Srivastava & Bhattacharya 2006; Pandey *et al.* 2014) and machine learning methods (Kumar *et al.* 2021; Bai *et al.* 2022). All procedures are based on the data collected by remote sensing, except for physical methods, which are expensive and time-consuming. These approaches do not consider the aquifer's response to the anticipated GW development (WOA and Machine Learning). If the aquifer's parameters are known, modelling the GW flow process can analyse the impact of pumping on the aquifer's capacity to support the expected GW development.

A combined effect of GW pumping on the aquifer and the river has been determined using the concept of a capture fraction map/capture map. The term ‘capture map’ was first articulated by Theis (1940), where he mentioned that water withdrawal by a well is balanced by a loss somewhere. He also mentioned that the cone of depression around a well could expand to recharge and discharge areas, thus altering the inflow and outflow patterns. Later, in 1954, Robert E. Glover and Glenn G. Balmer developed an analytical solution to study the transient capture of surface water from a river by a constant rate withdrawal well (Glover & Balmer 1954). In the last two decades, hypothetical models (Leake *et al.* 2010; Nadler *et al.* 2018), transient models based on previous data (Barlow & Dickerman 2001; Cosgrove & Johnson 2004), and steady-state models (Cosgrove & Johnson 2004) were used to generate capture maps. Leake *et al.* (2010) described the capture as fractions in terms of both volume and rates, which can be instantaneous or cumulative.

A novel methodology has been proposed to delineate suitable locations for pumping based on the river capture fraction, cost of pumping, and groundwater potential index (GWPI) (Figure 1). A calibrated GW forecasting model was simulated under future stress scenarios to calculate the river capture fraction and pumping cost. Additionally, using WOA, the

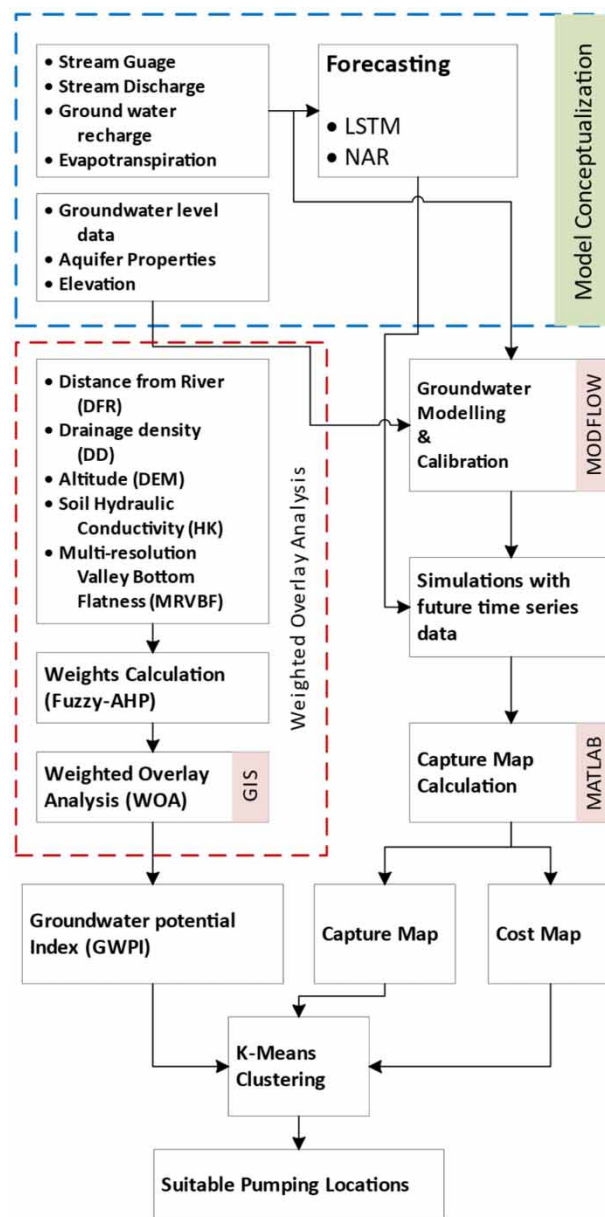


Figure 1 | Flow chart of the different steps of the methodology.

GWPI has been estimated with five geohydrological parameters: elevation, soil hydraulic conductivity (HC), multi-resolution valley bottom flatness (MRVBF), drainage density (DD), and distance from the river (DFR). Until this date, it was suggested to generate river capture either in steady-state conditions (Cosgrove & Johnson 2004; Leake *et al.* 2010) or transient model with historical data (Barlow & Dickerman 2001; Cosgrove & Johnson 2004; Leake *et al.* 2008, 2010). The river capture map has never been derived with the forecasted data nor used as an aquifer response to pumping to determine future GW development locations. This study uses the forecasted data to generate a river capture map for the future scenario. The time series data have been forecasted with Long Short-Term Memory (LSTM) and Nonlinear Auto-Regressive (NAR) neural networks. Based on the river capture, the cost data, and GWPI, the suitable locations for pumping have been delineated.

The objectives of this research work are as follows:

- To forecast the input time-series data up to 40 years with LSTM and NAR.
- To generate the river capture and cost map in the problem domain from the outputs of a calibrated GW flow model, simulated for the forecasted time-series data.
- The calculation of GWPI with WOA.
- To find the optimum pumping locations based on river capture, pumping cost, and GWPI.

2. METHODOLOGY

2.1. Data forecasting

A forecasting simulation model must have future time-series input data as its primary component. The accuracy of anticipated time series for various hydrological parameters, such as rainfall, evapotranspiration, and river discharge is always a determining factor in the capture map's accuracy. This study uses LSTM and NAR to predict stream discharge, rainfall, and stream water level. LSTM is a special type of Recurrent Neural Network (RNN) in which repeating unit (memory units) has a complex structure to 'remember' large sequences. These memory units also eliminate the problem of vanishing gradient encountered during backpropagation to obtain the weights of the neural network (Jozefowicz *et al.* 2015). NAR neural network applies a discrete nonlinear autoregressive network model for forecasting data (Ruiz *et al.* 2016). The network consists of multiple hidden layers (n-layers) that perform the nonlinear operations and an output layer that performs the regression operation. The hyperparameters of LSTM and NAR have been tuned with Linear Bayesian Optimisation (LBO) to achieve the best prediction accuracy. The Root-Mean-Square Error (RMSE) has been used as the accuracy metric as it combines the size of prediction errors for different data points into a single indicator of prediction accuracy. Each error has an impact on RMSE proportional to the size of the squared error; as a result, more significant errors have an outsized influence on RMSE. After comparing the RMSE of LSTM and NAR on test data, the most accurate method has been utilised to forecast the time series data.

2.2. GW modelling and simulation

A transient GW flow model is required to generate the river capture and cost maps. The GW model has been conceptualised in GMS (Aquaveo) environment with MODFLOW 2005 (Harbaugh 2005). The model parameters have been calibrated with the Parameter ESTimation (PEST) program (Doherty 2015). After calibration, the forecasted time-series data have been used to simulate the GW flow process for future stress periods.

2.3. Capture map and cost map

Leake *et al.* (2010) described the capture as the withdrawal-induced changes in inflow to or outflow from a hydrological system. In Nadler *et al.* (2018), the capture map bias in terms of pumping rate and time for nonlinear models was discussed. This bias limits the superposition of two capture maps. The map was created for a specific pumping and time to avoid this bias. In addition, instead of the ratio of discharge rates, the ratio of change in cumulative water discharge in the feature (river) to cumulative water consumption is considered to study the overall effect of pumping on the river up to the targeted time. To simulate the effect of pumping at different locations and times, a dummy well with a discharge of $-1,000 \text{ m}^3/\text{s}$ was added for each cell in the study domain. The depletion in the river discharge has been cumulated for all time steps. By changing the location of the dummy well (row (r), column (c)) and running the MODFLOW simulation, cost and capture data were

calculated as:

$$\text{Cost}(r, c, i) = d(r, c, j) \times p(r, c, j) \quad (1)$$

$$\text{Capture}(r, c, i) = \frac{\sum_{j=0}^i (w(j) - w(0)) \times t(j)}{\sum_{j=0}^i q(j) \times t(j)} \quad (2)$$

where r is the row number; c is the column number; j is the time step; $t(j)$ is the time interval length at j th time step; d is the depth of GW from the surface; p is the price of pumping for 1,000 m³/s of discharge for unit depth movement; $w(j)$ is the discharge from the feature at time step j ; $w(0)$ is the discharge from the feature without the dummy well; $q(j)$ is the discharge of the well. For this study, we assumed the price of pumping (p) at a certain time step to be constant over space. The procedure to calculate the capture fraction of the river to the pumping well has been modified by [Leake et al. \(2010\)](#) to incorporate the MODFLOW models which have been developed with the HDF5 file compression system.

- **Read the .h5 input file & extract the well data from .h5 file**
- **Set the discharge of dummy well as -1000 m³/s.**
- **For each interested location in model grids**
 - Set the position of dummy well to that location
 - Write dummy well to .h5 file
 - Run MODFLOW with updated .h5 file
 - Read the .ccf output file for flow data
 - Calculate the change in volume of water for calculation of capture using equation (2)
 - Read the .hed output file for head data
 - Get the drawdown for cost map from .hed file
 - Calculate the cost using equation (1)
- **End for**
- **Output the grid of cost map and capture map.**

2.4. Groundwater potential index

GWPI is a dimensionless quantity representing GW availability in an area. It has been determined using a weighted linear arrangement, as suggested by [Venkatramanan et al. \(2019\)](#).

$$\text{GWPI} = \sum_{j=1}^n \sum_{i=1}^m (\widetilde{\text{NW}}_j \times R_i) \quad (3)$$

Here, $\widetilde{\text{NW}}_j$ is the normalised weight of GW-influencing factors calculated using Saaty's AHP method ([Saaty 2014](#)) and R_i is the weight assigned to each subclass of the GW-influencing factor. The thematic raster for different hydrological parameters has been prepared in the GIS environment. Initially, 10 GW-influencing parameters (elevation, Topographic Wetness Index, slope, distance from river, DD, land use map, surface geomorphology, soil HC, multi-resolution ridge top flatness, and MRVBF) have been selected for WOA. Expert analysis and previous research on the study area have shown that the five parameters (elevation, HC, MRVBF, DFR, and DD) have more importance than others. Therefore, these five hydrological parameters have been used to calculate GWPI.

The topography significantly influences the GW recharge and its movement, as the GW table frequently reflects the local terrain ([Weiss 2001](#)). TWI is used to quantify the topographical control of the hydrological processes. It is defined as $\log(A_s/\tan(\beta))$, where A_s is the local upslope area per unit contour length and $\tan\beta$ is the steepness (slope) in degree

(Moore *et al.* 1991). GW potential is greatly influenced by soil HC, which is a primary driver of GW recharge. The MRVBF index represents the flatness and lowness of the valley bottoms. When compared to channels, valley bottoms offer a good indication of the breadth of a subsurface flow path in flat terrain and hence a good GW potential (Gallant & Dowling 2003). Since the active river aquifer exchanges can replenish the withdrawn GW, proximity to rivers and streams increases GW potential.

2.5. K-Means clustering

K-Means clustering (MacQueen 1967) is an unsupervised learning algorithm that groups the unlabelled dataset into different clusters, such that the sum of distances between the data point and their corresponding cluster centres is minimised. The three independent variables (dimension), namely river capture, pumping cost, and GWPI have been used to cluster the corresponding locations into five clusters, representing areas in different zones for GW development. The elbow method is utilised to determine the optimal number of five clusters. The elbow method is an iterative process in which the WCSS (within-cluster sum of the square) is minimised as per the number of clusters. The optimum number of clusters is selected corresponding to the highest change of WCSS, represented by a drastic change in the WCSS vs. the number of cluster plots (the elbow). The K-Means clustering has been implemented with the sci-kit learn library (Pedregosa *et al.* 2011) in a Python environment.

3. IMPLEMENTATION

3.1. Study area

The presented methodology has been implemented in the lower Ain Valley, France. This study area has been selected because topography, water budget, well discharge, riverbed elevation, evaporation rate, and soil properties are available. In addition, a few publications (Lejot *et al.* 2007; Gaur *et al.* 2011a, 2011b; Lalot *et al.* 2012; Graillot *et al.* 2014; Dole-Olivier *et al.* 2019) have also contributed to a better understanding of the considered domain and its vicinity. The Ain River drains a basin of around 3,640 km² and has an oceanic hydrologic regime. The majority of its path is situated in France's southern Jura Mountains. The Ain River is a right-bank tributary of the Rhône River and has a length of roughly 200 km. The lower Ain River is in a very expansive floodplain between the Allement dam and its confluence with the Rhône River (Figure 2(a)).

3.2. Time series data forecasting

Apart from the aquifer properties, river discharge, and river stage have been forecasted for the next 40 years with LSTM, stacked LSTM, and NAR (Figure 3). Hyperparameter tuning with LBO was performed to reduce the overfitting and underfitting. Among the three techniques – LSTM, stacked LSTM, and NAR – the one with the least RMSE was selected for final forecasting (Table 1). The recharge and evapotranspiration for the future time series have been taken as the 10-year moving average.

3.3. GW modelling

The model conceptualisation has been done using available geological and hydrological data. The modelled aquifer thickness varies from 240 to 550 m, as observed in the lithology data by BRGM (Bureau de Recherches Géologiques et Minières), where SRTM data were used to get the surface elevation of the area. For the underlying sediments, a two-layered model was created, with each layer considered homogenous, horizontal, and isotropic. For the older sediment, the initial estimates of HC were taken to be 155.5 m/day, and for the younger alluvium, 259.2 m/s. The aquifer HC was estimated from the pumping test and the literature data obtained by BRGM. The area has been modelled with 250 m square grids and two layers. Two types of boundary conditions (no-flow boundary and specified flow boundary) were used to define the model domain. The inflow and outflow flux has been calculated by the observed GW heads using Darcy's flux (Darcy 1856). The Ain River is modelled as head-dependent boundary with stream package (STR1) in GMS. Figure 2(b) demonstrates how the watershed division line and the no-flow barrier were used to determine the eastern and western boundaries.

Further, the river stage data, measured at five locations in the study area, was obtained from Banque Hydro France. The river stage data from 2002 to 2015 was also examined to understand the data trend. This analysis showed a fluctuation of 2.54 m at Chazey-sur-Ain and 1.67 m at Pont-d'Ain, data from other tributaries, such as the Albarine River, also show a maximum fluctuation of 1.36 m. The water table variation is observed to be in the range of 1.2–3.7 m, which shows a stable GW scenario.

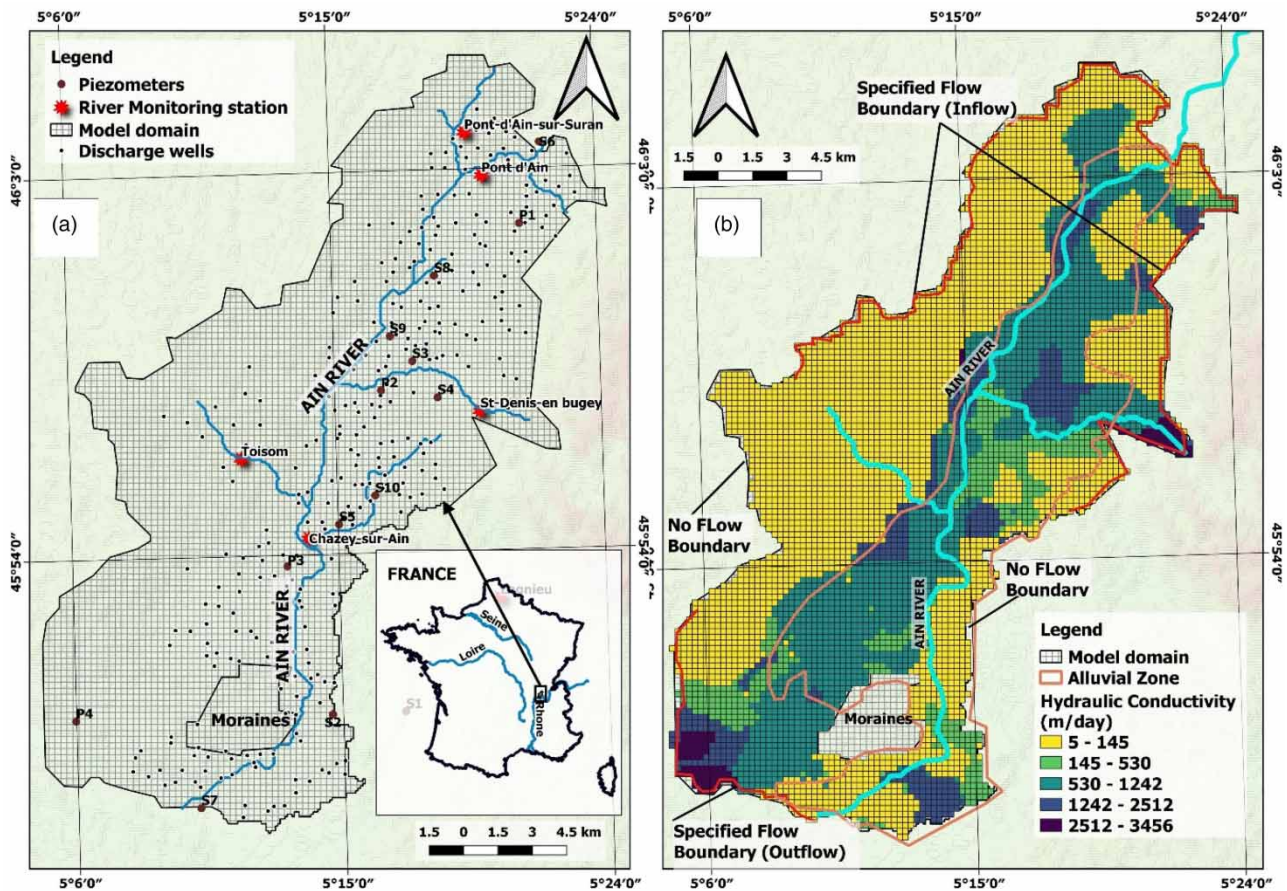


Figure 2 | (a) Study area situated at lower Ain River, France and (b) model conceptualisation with boundary conditions.

Rainfall was solely taken into account as a way to recharge the GW system in the area. The recharge input values were computed using rainfall data from the Météo France database and applied equally throughout the recharge polygons built in the model domain based on land use. The starting value of GW recharge was taken as 50, 10, 50, 60, and 80% of the rainfall (average 1,650 mm/year) for the six land use classes, namely vegetation, built-up, agriculture, fellow land, and sandy area respectively, which was further calibrated. With an extinction depth of 2 m, the projected annual evapotranspiration is 638 mm. The 10-year monthly rainfall average was taken for the future forecast, and corresponding recharge values were used. The model was calibrated from 2008 to 2010 using all observation wells, then again from 2010 to 2012 using the remaining wells, depending on the available data. Data from 2012 to 2015, which were available for four observation wells, were used to validate the model. The GW modelling of the area is further discussed in Bajpai *et al.* (2022).

3.4. GW capture and pumping cost

The calibrated GW model was run for future stress periods by taking the forecasted time-series data as input. The model for calculating the capture fraction was coupled with MODFLOW by modifying the hdf5 input files in MATLAB. The GW model has been used to identify the effects of long-run pumping (for the next 40 years) on water budgeting by considering the discharge well on the cell by cell basis. The cells lying in the alluvium zone around the river were only considered for the analysis.

3.5. GWPI

GWPI of the study area having an alluvium deposit has been calculated with WOA in the GIS environment. The five hydrological parameters considered have been reclassified into subclasses by the Natural Jenks method. These subclasses have been

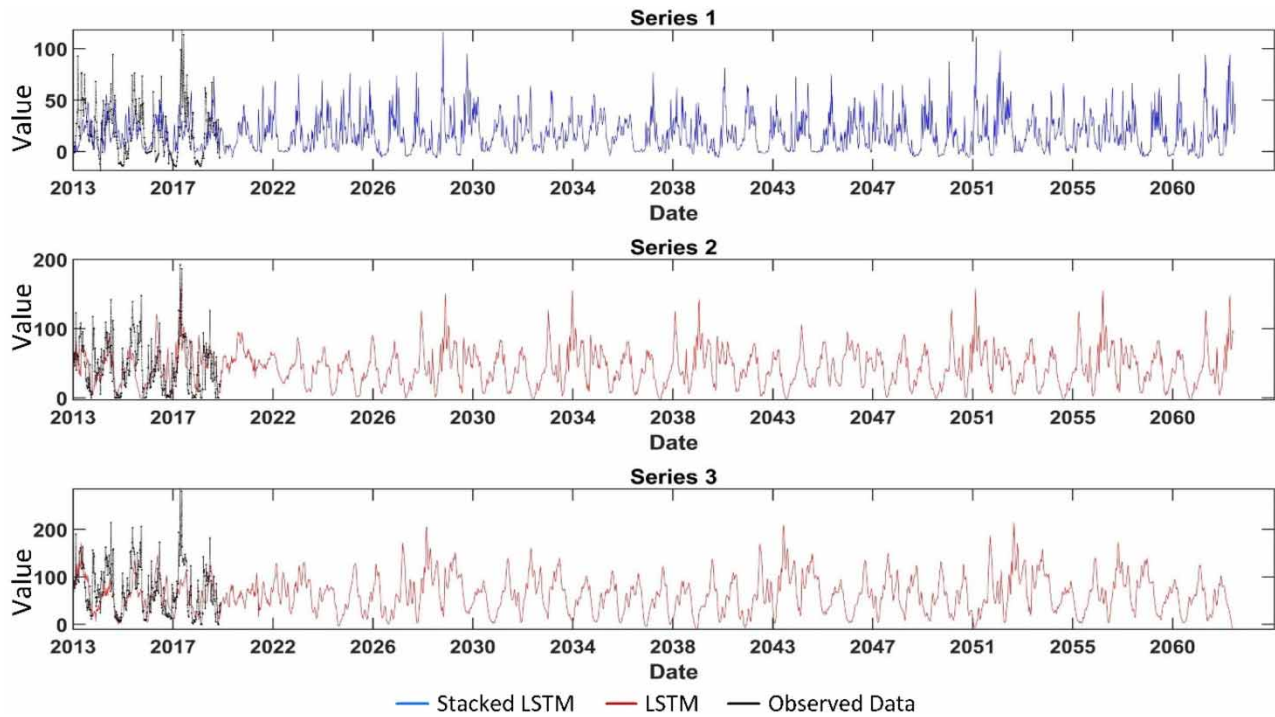


Figure 3 | Time series forecasting with LSTM and stacked LSTM (Series – 1: Albarine_St Denis-en-Bugey; 2: Pont-d'Ain; and 3: Chazey-sur-Ain).

Table 1 | Optimum range of hyperparameters for hyper-tuned LSTM and NAR model (Series – 1: Albarine_St Denis-en-Bugey, 2: Pont-d'Ain, and 3: Chazey-sur-Ain)

| | Hyperparameters | Optimum range | RMSE | | |
|---------------------|--------------------------------|------------------------|--------------|--------------|--------------|
| | | | Series 1 | Series 2 | Series 3 |
| LSTM | Shape of input features (lags) | 5–70 | 25.3 | 36.33 | 47.38 |
| | Maximum epochs | 60–230 | | | |
| | Initial learning rate | 0.001–0.05 | | | |
| | Mini-Batch Size | 11–60 | | | |
| | Regularisation weight | 10^{-7} to 10^{-5} | | | |
| | Number of hidden units | 5–70 | | | |
| Stacked LSTM | Same as above | Same as above | 24.94 | 37.17 | 55.7 |
| NAR | Shape of input features (lags) | 6–42 | 382.8 | 734.4 | 917.9 |
| | Number of hidden units | 5–28 | | | |

The selected algorithm for different time series data has been highlighted corresponding to the lowest RMSE value.

assigned a weight based on the AHP method (Table 2). Finally, all parameters' normalised weights have been calculated with AHP, and GWPI has been calculated (Figure 4).

4. RESULTS

The series with the least RMSE in testing was used to develop the MODFLOW model for water consumption-based capture maps and cost map generation. It is observed from Figure 3 that, unlike NAR, the LSTM-based model can capture complex features for a long-term forecast. However, in the case of predictions averaged over 16 weeks, both LSTM and NAR models provide similar results. Either single LSTM or stacked LSTM has the least RMSE. For the hyper-tuned LSTM and NAR models, the

Table 2 | Groundwater-influencing parameters and their weights

| S/N | Parameters | Normalised weights | Subclasses | Weights |
|-----|--------------------------|--------------------|-------------------|---------|
| 1 | DFR (m) | 0.092 | <1,230.42 | 0.559 |
| | | | 1,230.42–2,605.61 | 0.243 |
| | | | 2,605.61–4,089.36 | 0.111 |
| | | | 4,089.36–5,754.06 | 0.052 |
| | | | >5,754.06 | 0.035 |
| 2 | DD (km/km ²) | 0.062 | <0.67 | 0.049 |
| | | | 0.67–1.48 | 0.074 |
| | | | 1.48–2.38 | 0.143 |
| | | | 2.38–3.52 | 0.247 |
| | | | > 3.52 | 0.488 |
| 3 | MRVBF | 0.112 | < 1.32 | 0.037 |
| | | | 1.32–3.4 | 0.064 |
| | | | 3.4–5.21 | 0.129 |
| | | | 5.21–7.15 | 0.265 |
| | | | >7.15 | 0.506 |
| 4 | Elevation (m) | 0.204 | <211 | 0.589 |
| | | | 211–233.94 | 0.234 |
| | | | 233.94–273.36 | 0.093 |
| | | | 273.94–296.17 | 0.052 |
| | | | >296.17 | 0.033 |
| 5 | HC (m/day) | 0.53 | <18.53 | 0.032 |
| | | | 18.53–72.67 | 0.047 |
| | | | 72.67–126.8 | 0.135 |
| | | | 126.8–830.53 | 0.233 |
| | | | >830.53 | 0.553 |

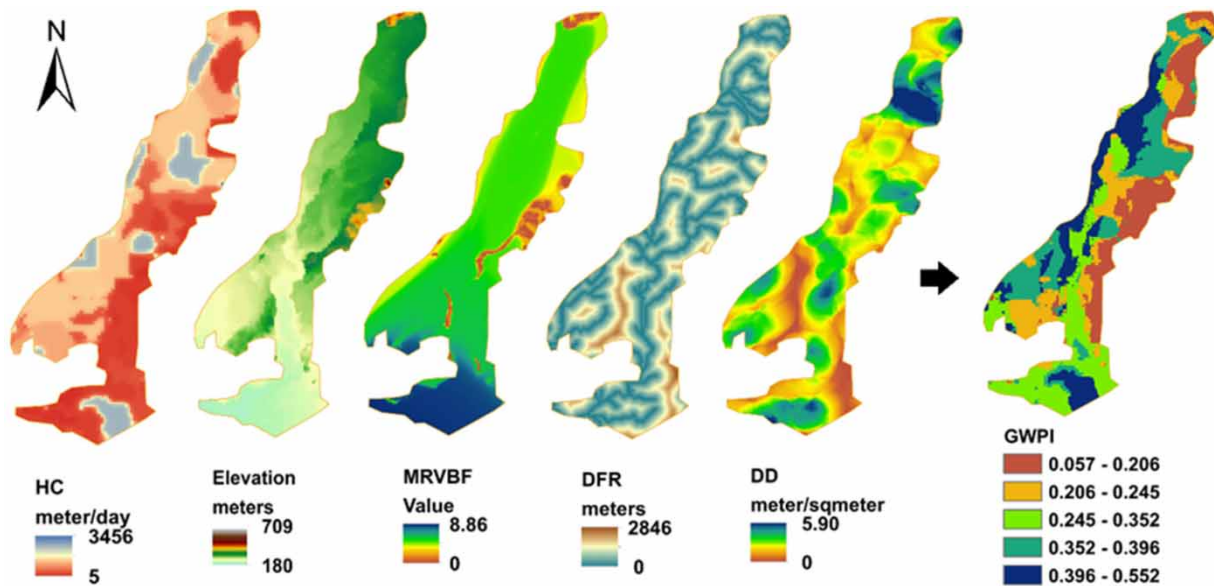


Figure 4 | WOA (GW-influencing parameters and GWPI).

optimum range of hyperparameters has been tabulated in Table 1, along with the RMSE values. Most of the long-term forecast is highly uncertain, but the 4-month interval average RMSE was low. In addition, the granularity of time series data for forecasting required for the MODFLOW model was high; hence these predictions were used to generate capture maps.

Once the forecasted data were created up to the year 2060, GW flow was simulated with these data. Predicted heads through the model were found stable (standard deviation – 0.375 m) with a slight upward trend (1.8 cm per year). Further, capture maps related to river-inflow and river-outflow were generated with the highest capture of 99.78% and the lowest of 36.39%. Another type of map was also developed by considering the cost of pumping, which is the function of water table depth.

The water consumption-based capture map for a particular source shows the fraction of water demand fulfilled by that source (here, the Ain River). The effect of pumping on river flow was studied. The water demand satisfied by the river varies from 36.39 to 99.78%, with an extraction rate of 1,000 m³/s from the pumping well (Figure 5; river capture). Most of the points satisfying more than 89% of the demand lie near the river. This fraction is more significant for the lower reach (Chazey-sur-Ain) than the upper reach (Pont-d'Ain). A pumping well 3–5 km away from the river can lower the water gained through the aquifer by 40–60%. This decrement in the water gained by the aquifer can be due to the lower inflow into the river or greater outflow out of the river.

The cost map represents the pumping cost which is assumed to be proportional to the depth of the GW. For the area close to the river within a 1 km buffer, the water table is below 10 m of depth from the surface, mostly limited to around 1 m. On the lower left side of the domain, some area in the vicinity of the river has large depth. This can be attributed to the hilly region and inaccurate modelling of the area. Nearly all the points on the right side of Ain River have a water table depth of less than 17 m. The GWPI values were extracted for the model domain for clustering along with river capture and cost.

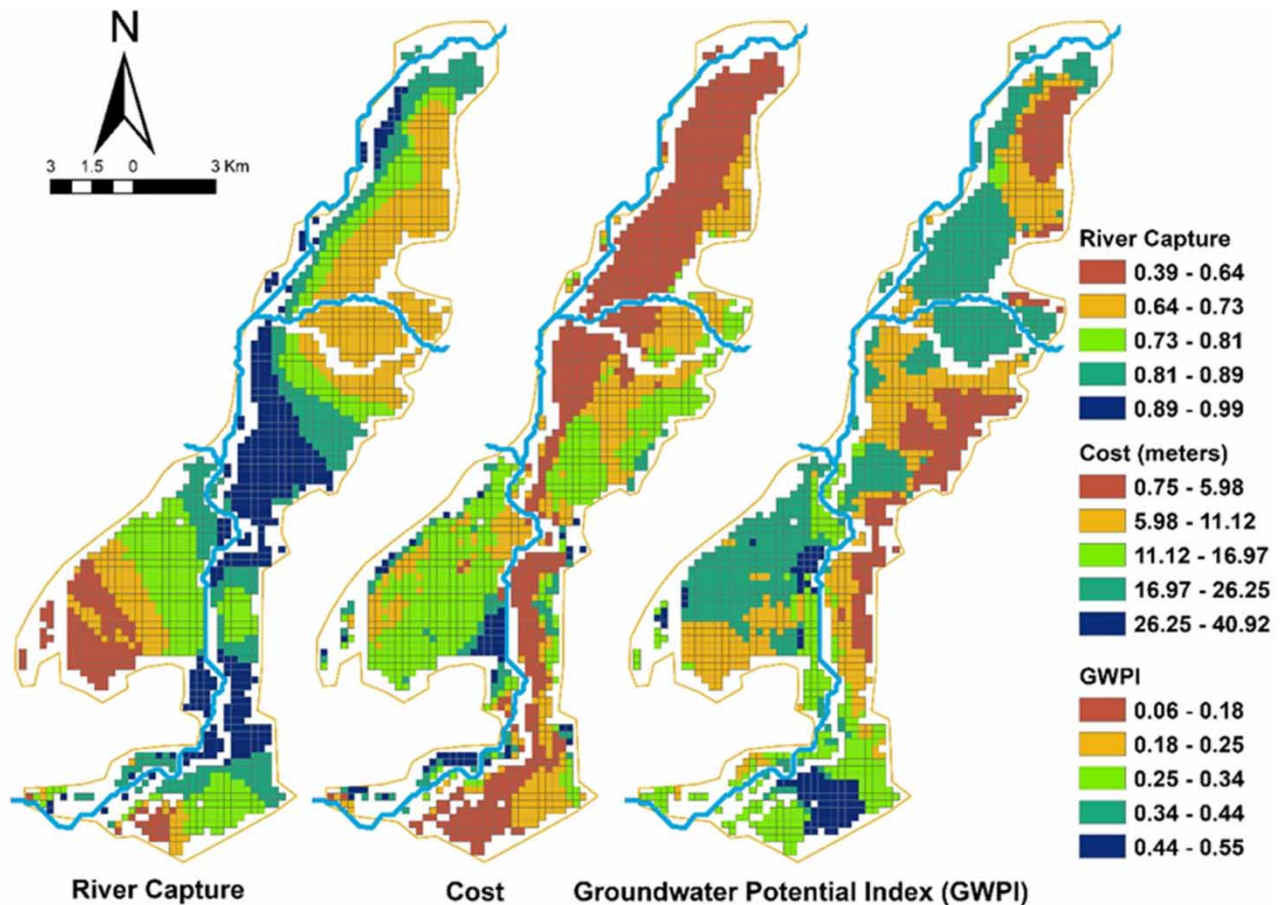


Figure 5 | River capture map, cost map, and GWPI for the model domain (alluvium zone).

K-Means clustering with three variables (river capture, cost, and GWPI) has been used to classify the model domain in five zones of GW development. Capture map of the river Ain helps to identify the location with less impact on the river flow. It is also affected by less GW inflow into the river or excessive water extraction through the river due to excessive pumping. The clusters have been analysed as

| | |
|------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. <i>Low cost, high capture, and high GW potential:</i> | These cells represent high river water capture, which has high GW potential. The pumping cost in these areas is low due to readily available river water. However, the dependency on the river in these areas can be detrimental to the river flow. |
| 2. <i>Low cost, moderate capture, and low GW potential</i> | These areas represent good pumping zones for short durations since the GW potential is low. |
| 3. <i>Low cost, high capture, and low GW potential</i> | These areas have a low cost of pumping, but most of the water is from the river only, which makes the area of low GW potential. |
| 4. <i>Low cost, low capture, and high GW potential</i> | These areas are best for new GW development projects, with low cost of pumping and low river capture. |
| 5. <i>High cost, high capture, and mixed GW potential</i> | These highly scattered zones are the worst locations for GW development works due to the high cost of pumping and immense strain on river water. |

Most of the worst locations are less frequent for both pumping costs and river water consumption. The spatial location of these points in the study area is presented in [Figure 6](#). The lower side of the domain contains both the best and the worst locations. The worst locations lie in the lower part of the area near the moraines, whereas the best pumping locations are situated in the upper part. The continuous supply of snow melt water from nearby streams originating from the Jura mountains results in shallow GW table in this upper part.

5. DISCUSSION

The prediction capabilities of LSTM and NAR have shown variation based on the temporal scale of the time series data. LSTM has proven superior in predicting temporal variation with daily time series data compared to NAR. However, they both performed more or less similarly for the 4-month average time series data. The GW model is an essential part of the whole research work, and the accuracy of the capture map depends on the GW model. Therefore, the model should be calibrated well before applying it for long-term prediction. In forecasting, volatility, a granularity of time series, and lead-time of prediction play a significant role in the selection of forecasting techniques. The study incorporates the prediction of time series through the forecasting method. The NAR neural network is computationally inexpensive compared to LSTM and is suitable for non-volatile time series with low lead time. Although LSTM-based networks are computationally expensive, they can predict a much more volatile time series with a higher lead time. Stochastic forecasting techniques can be used instead of deep learning-based for more accurate long-term predictions with wide uncertainty intervals.

The concept of capture has proven very useful in understanding and managing the water resource potential of an area at the reach scale as it represents the aquifer response to the pumping. A balanced abstraction of GW, considering the extent of stream capture, can prove to be a better way to manage our existing water resources. A calibrated physical-based GW model with accurate aquifer properties can provide a realistic and accurate aquifer response to a given pumping stress. Incorporating the capture map with GW potential mapping can reduce the uncertainty in the delineation of suitable pumping locations. As evident from [Figure 6](#), the proximity to the river does not justify the development of GW extraction projects (as in the lower part of the river, there are many cells which are not suitable for pumping).

The lower Ain River basin has a very stable GW table. Increased industrial and agricultural activities in the lower Ain River basin have, in turn, increased the stresses on the area's water resources in the future. In the present study, an effort was made to analyse the long-run impact of increasing water pumping on river water budgeting by developing a capture map and cost map for the lower Ain Valley. Developed maps are applied to identify the potential of the area. The capture map suggested that nearly 90 percentiles of the considered area satisfies nearly 65.54% of the total water demand from the river. The river capture in the middle part of the study area, near Chazey-sur-Ain, was more affected by pumping, whereas the upper part near Pont-d'Ain was found to be less affected by pumping.

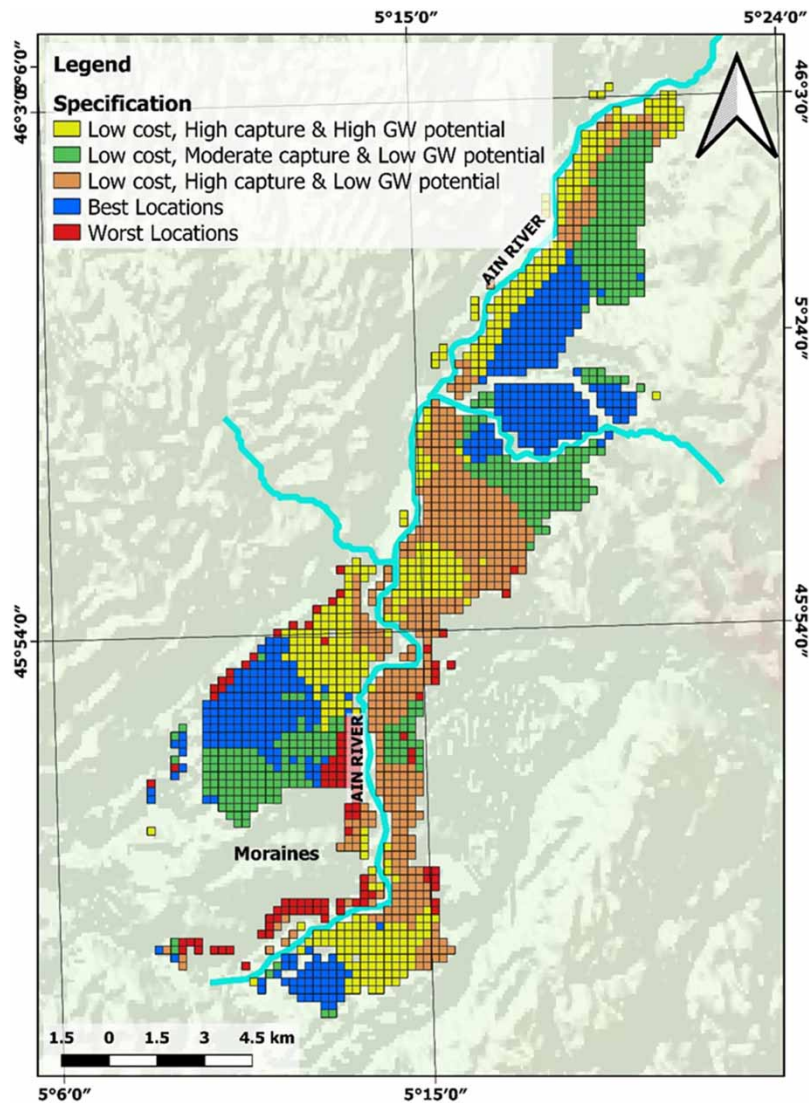


Figure 6 | Final suitable sites for GW development in the study area.

6. CONCLUSION

A novel methodology to determine suitable GW pumping locations based on the concept of capture fraction and GW potential has been proposed in this research work. The pumping cost was also considered in the unsupervised clustering (K-Means) to find the best locations to minimise the cost and the capture from the river in high GW potential zones. The LSTM and NAR have been utilised to predict time-series data to simulate the GW model up to 2060. This research is expected to help the decision makers, where the models are used for evaluating and testing the management outcomes related to GW and surface water resources. This research provides a novel constraint for the future development of GW based on balanced GW and surface water resource management. The water quality is also a constraint that should be included in holistic GW resource management. Although the best pumping locations have been derived by minimising the river capture, the aquifer's vulnerability to contamination due to pumping can be integrated with this methodology in future research work.

ACKNOWLEDGEMENT

This work was performed within the framework of the EUR H2O'Lyon (ANR-17-EURE-0018) of Université de Lyon (UdL) through the 'Investissements d'Avenir' programme operated by the French National Research Agency (ANR) and through a

joint Indian-French International Research Project (IRP-CNRS) on ‘Effects of River-Aquifer Exchanges on Riverine Ecosystem Resilience to Global Change Comparative Approach of the Ganga and Rhône River Basin Networks’ (2021–2026).

FUNDING

No funding received.

AUTHORS CONTRIBUTIONS

S.G., A.O., and P.H. drafted and devised the idea of the work and supervised its timely progress and guided in the technical and literature aspects. G.S. and O.A. created the groundwater models. M.B. carried out the model simulations and forecasting. R.K. drafted the results and formatted the literature accordingly. All authors discussed the results and contributed to the final manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Ayazi, M. H., Pirasteh, S., Arvin, A. K. P., Pradhan, B., Nikouravan, B. & Mansor, S. 2010 Disasters and risk reduction in groundwater: Zagros mountain southwest Iran using geoinformatics techniques. *Disaster Advances* **3** (1), 51–57.
- Bai, Z., Liu, Q. & Liu, Y. 2022 Groundwater potential mapping in Hubei region of China using machine learning, ensemble learning, deep learning and AutoML methods. *Natural Resources Research* **31** (5), 2549–2569. <https://doi.org/10.1007/s11053-022-10100-4>.
- Bajpai, M., Mishra, S., Gaur, S., Ohri, A., Piégay, H. & Graillot, D. 2022 Optimization of groundwater pumping and river-aquifer exchanges for management of water resources. *Water Resources Management* **36** (6), 1863–1878. <https://doi.org/10.1007/s11269-022-03112-5>.
- Barlow, P. M. & Dickerman, D. C. 2001 Numerical-simulation and conjunctive-management models of the Hunt – Annaquatucket – Pettaquamscutt stream-aquifer system, Rhode Island. *US Geological Survey Professional Paper*, <https://doi.org/10.3133/pp1636>.
- Cosgrove, D. M. & Johnson, G. S. 2004 Transient response functions for conjunctive water management in the Snake River Plain, Idaho. *Journal of the American Water Resources Association* **40** (6), 1469–1482. <https://doi.org/10.1111/j.1752-1688.2004.tb01599.x>.
- Crosbie, R. S., Scanlon, B. R., Mpelasoka, F. S., Reedy, R. C., Gates, J. B. & Zhang, L. 2013 Potential climate change effects on groundwater recharge in the High Plains Aquifer, USA. *Water Resources Research* **49** (7), 3936–3951. <https://doi.org/10.1002/wrcr.20292>.
- Darcy, H. 1856 *Les Fontaines publiques de la ville de Dijon*. V. Dalmont. Available from: <https://books.google.co.in/books?id=9nhKBu2-GJEC>.
- Doherty, J. 2015 *Calibration and Uncertainty Analysis for Complex Environmental Models*. Watermark Numerical Computing, Brisbane, Australia.
- Dole-Olivier, M. J., Wawzyniak, V., Creuzé des Châtelliers, M. & Marmonier, P. 2019 Do thermal infrared (TIR) remote sensing and direct hyporheic measurements (DHM) similarly detect river-groundwater exchanges? study along a 40 km-section of the Ain River (France). *Science of the Total Environment* **646**, 1097–1110. <https://doi.org/10.1016/j.scitotenv.2018.07.294>.
- Gallant, J. C. & Dowling, T. I. 2003 A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*. <https://doi.org/10.1029/2002WR001426>.
- Gaur, S., Chahar, B. R. & Graillot, D. 2011a Analytic elements method and particle swarm optimization based simulation-optimization model for groundwater management. *Journal of Hydrology* **402** (3–4), 217–227. <https://doi.org/10.1016/j.jhydrol.2011.03.016>.
- Gaur, S., Mimoun, D. & Graillot, D. 2011b Advantages of the analytic element method for the solution of groundwater management problems. *Hydrological Processes* **25** (22), 3426–3436. <https://doi.org/10.1002/hyp.8071>.
- Glover, R. E. & Balmer, G. G. 1954 River depletion resulting from pumping a well near a river. *Transactions, American Geophysical Union* **35** (3), 468. <https://doi.org/10.1029/TR035i003p00468>.
- Graillot, D., Paran, F., Bornette, G., Marmonier, P., Piscart, C. & Cadilhac, L. 2014 Coupling groundwater modeling and biological indicators for identifying river/aquifer exchanges. *SpringerPlus* **3** (1), 1–14. <https://doi.org/10.1186/2193-1801-3-68>.
- Harbaugh, A. W. 2005 The U.S. Geological Survey Modular Ground-Water Model – the Ground-Water Flow Process. *U.S. Geological Survey Techniques and Methods 6-A16*, p.253.
- Jozefowicz, R., Zaremba, W. & Sutskever, I. 2015 An empirical exploration of Recurrent Network architectures. *32nd International Conference on Machine Learning, ICML 2015* **3**, 23322340.

- Kumar, R., Dwivedi, S. B. & Gaur, S. 2021 A comparative study of machine learning and Fuzzy-AHP technique to groundwater potential mapping in the data-scarce region. *Computers and Geosciences* **155** (September 2020), 104855. <https://doi.org/10.1016/j.cageo.2021.104855>.
- Lalot, E., Paran, F., Graillot, D. & Batton-Hubert, M. 2012 Spatio-temporal approach and modeling for the identification and quantification of water exchange between a river and its alluvial aquifer – application to the Rhône River. In: *EGU General Assembly 2012*. Vienne, Austria. p. 5352. Available from: <https://hal.archives-ouvertes.fr/hal-00772833>.
- Leake, S. A., Pool, D. R. & Leenhouts, J. M. 2008 Simulated effects of ground-water withdrawals and artificial recharge on discharge to streams, springs, and riparian vegetation in the Sierra Vista subwatershed of the Upper San Pedro Basin, Southeastern Arizona. *U.S. Geological Survey Scientific Investigations Report 2008-5207*, April, **14**. <http://pubs.usgs.gov/sir/2008/5207/>.
- Leake, S. A., Reeves, H. W. & Dickinson, J. E. 2010 A new capture fraction method to map how pumpage affects surface water flow. *Ground Water* **48** (5), 690–700. <https://doi.org/10.1111/j.1745-6584.2010.00701.x>.
- Lejot, J., Delacourt, C., Piégay, H., Fournier, T., Trémélo, M.-L. & Allemand, P. 2007 Very high spatial resolution imagery for channel bathymetry and topography from an unmanned mapping controlled platform. *Earth Surface Processes and Landforms* **32** (11), 1705–1725. <https://doi.org/10.1002/esp.1595>.
- MacQueen, J. 1967 Classification and analysis of multivariate observations. In: *5th Berkeley Symp. Math. Statist. Probability* (Le Cam, L.M. & Neyman, J., eds.). University of California, Los Angeles, CA. pp. 281–297.
- Moore, I. D., Grayson, R. B. & Ladson, A. R. 1991 Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes* **5** (1), 3–30. <https://doi.org/10.1002/hyp.3360050103>.
- Nadler, C., Allander, K., Pohll, G., Morway, E., Naranjo, R. & Huntington, J. 2018 Evaluation of bias associated with capture maps derived from nonlinear groundwater flow models. *Groundwater* **56** (3), 458–469. <https://doi.org/10.1111/gwat.12597>.
- Nampak, H., Pradhan, B. & Abd Manap, M. 2014 Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology* **513**, 283–300.
- Pandey, N. K., Shukla, A. K., Shukla, S. & Pandey, M. 2014 Assessment of underground water potential zones using modern geomatics technologies in Jhansi district, Uttar Pradesh, India. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives XL-8* (1), 377–381. <https://doi.org/10.5194/isprsarchives-XL-8-377-2014>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. 2011 Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* **12** (85), 2825–2830. Available from: <http://jmlr.org/papers/v12/pedregosa11a.html>.
- Ruiz, L., Cuéllar, M., Calvo-Flores, M. & Jiménez, M. 2016 An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies* **9** (9), 684. <https://doi.org/10.3390/en9090684>.
- Saaty, T. L. 2014 Analytic hierarchy process. In: *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.stat05310>.
- Srivastava, P. K. & Bhattacharya, A. K. 2006 Groundwater assessment through an integrated approach using remote sensing, GIS and resistivity techniques: a case study from a hard rock terrain. *International Journal of Remote Sensing* **27** (20), 4599–4620. <https://doi.org/10.1080/01431160600554983>.
- Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., R. V., Beek, R. V., Y., Wada, Y., Longuevergne, L., M., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen J., M., Taniguchi M., M. F. P., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., I., Holman, I. & Treidel, H. 2013 Ground water and climate change. *Nature Climate Change* **3** (4), 322–329.
- Theis, C. V. 1940 The source of water derived from wells: essential factors controlling the response of an aquifer to development. *Civil Engineering* **10**, 277–280.
- Venkatramanan, S., Prasanna, M. V. & Chung, S. Y. 2019 *GIS and Geostatistical Techniques for Groundwater Science*. <https://doi.org/10.1016/C2017-0-02667-8>.
- Weiss, a. 2001 Topographic position and landforms analysis. In *Poster Presentation, ESRI User Conference*, San Diego, CA. https://doi.org/http://www.jennessent.com/downloads/TPI-poster-TNC_18x22.pdf.
- Wu, W. Y., Lo, M. H., Wada, Y., Famiglietti, J. S., Reager, J. T., Yeh, P. J. F., Ducharme, A. & Yang, Z. L. 2020 Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. *Nature Communications* **11** (1), 1–9. <https://doi.org/10.1038/s41467-020-17581-y>.

First received 19 January 2023; accepted in revised form 2 June 2023. Available online 13 June 2023