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

Mayank Bajpai¹, Ranveer Kumar¹, Shishir Gaur¹, Anurag Ohri¹, and Hervé Piégay²

¹Indian Institute of Technology BHU Varanasi

²University of Lyon

January 18, 2023

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 utilization. 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, France.

1. Introduction

One of the most significant and essential natural resources is groundwater, 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). This finite freshwater resource is under tremendous pressure due to the rising demand for water to support human needs and development. The presence and spread of groundwater are influenced by various anthropogenic and natural variables. Therefore, studying potential groundwater 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 groundwater 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 groundwater 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 groundwater or surface water affects the other. Therefore, it is crucial to identify the best locations for groundwater development projects to lessen the strain on surface water resources and groundwater storage.

Groundwater potential zonation methodologies have come a long way with the use of physical methods (like borehole drilling and pumping tests), weighted overlay analysis (Pandey et al., 2014; Saraf & Choudhury, 1998; Sener et al., 2005; Srivastava & Bhattacharya, 2006) and machine learning methods (Bai et al., 2022; Corsini et al., 2009; Kordestani et al., 2019; Kumar et al., 2021; Naghibi et al., 2018; Rahmati et al., 2016; Rizeei et al., 2019; Zabihi et al., 2016). All procedures are based on the data collected by remote sensing,

with the exception of physical methods, which are expensive and time-consuming. The aquifer’s response to the anticipated groundwater development is not considered by these approaches (Weighted Overlay Analysis and Machine Learning). If the aquifer’s parameters are known, modelling the groundwater flow process can analyze the impact of pumping on the aquifer’s capacity to support the expected groundwater development.

The use of capture maps for decision-making in groundwater management has been gaining popularity over the few decades with the development of fast and easy-to-handle groundwater models with flopy (in python) and GMS (Groundwater Modeling System) along with the automation of these models with the help of MATLAB, python, pearl and FORTRAN. The term “capture map” was first articulated by Charles V. Theis in 1940 (Theis, 1940), where he mentioned that water withdrawal by well is balanced by a loss somewhere. He also mentioned that the cone of depression around a well can expand to areas of recharge and discharge, thus altering the inflow and outflow patterns of the system. 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 (S. A. Leake et al., 2010; Nadler et al., 2018), transient models based on previous data (Barlow & Dickerman, 2001; Cosgrove & Johnson, 2004; S. A. Leake et al., 2010; S. a Leake et al., 2008) and steady-state models (Cosgrove & Johnson, 2004; S. A. Leake et al., 2010) 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 (S. A. Leake et al., 2010). They also mentioned different types of capture maps to study the effects of groundwater pumping on surface water bodies. In Leake et. al. 2008, a 100 years model (from 1902 to 2003) was developed for northern Sonora, Mexico, and Upper San Pedro Basin in southeastern Arizona, USA, with a withdrawal rate of 100 and 1000 cubic meters per day (S. a Leake et al., 2008). The maps for 10 and 50 years of withdrawal depicted limitations of capture maps at few dry cells. In 2005, Cosgrove et. al. studied the relation between river seepage and groundwater pumping for various Snake River reaches in Idaho using MODRSP code to unit steady-state response functions for computing capture map (Cosgrove & Johnson, 2005).

The hydroclimatic time series, like river discharge, river stage, rainfall, etc., are frequently forecasted for designing, planning and monitoring water-based activities/systems. Machine learning has been intensively used in the last two decades for forecasting. These include, and are not limited to, the support vector regression (Belayneh et al., 2014; Parchami-Araghi et al., 2013; Seo et al., 2015), Artificial Neural Networks (Daliakopoulos et al., 2005; Napolitano et al., 2011; Parchami-Araghi et al., 2013; Sarangi et al., 2006; Seo et al., 2015), Feed Forward Neural Network (FFNN) with Genetic Algorithm (Bahrami et al., 2016; Danandeh Mehr & Kahya, 2017; Ketabchi & Ataie-Ashtiani, 2015) and LSTM (Long Short Term Memory) (Kratzert et al., 200 C.E.; Supreetha et al., 2020; Zhang et al., 2018). Recently, the use of deep neural networks, including LSTM, GRU (Gated Recurrent Unit), DGRU (Deep Gated Recurrent Unit), CNN-LSTM (Convolutional Neural Network-Long Short Term Memory) have shown significant performance improvement over the traditional forecasting techniques like ARIMA (Autoregressive Integrated Moving Average: Statistical forecasting technique), NAR (Non-linear Auto Regression: neural network based forecasting), neural networks and FFNN (Kratzert et al., 200 C.E.; Sagheer & Kotb, 2019; Supreetha et al., 2020; Zhang et al., 2018).

A novel methodology has been proposed to delineate suitable locations for pumping based on the river capture fraction and cost of pumping. A calibrated groundwater model was performed under future stress scenarios to calculate the river capture fraction and pumping cost. Additionally, using Weighted Overlay Analysis (WOA), the Groundwater Potential Index (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; S. A. Leake et al., 2010) or transient model with historical data (Barlow & Dickerman, 2001; Cosgrove & Johnson, 2004; S. A. Leake et al., 2010; S. a Leake et al., 2008). The river capture map has never been derived with the forecasted data nor been used to determine future groundwater development locations. In this study, the forecasted data has been used to generate a river capture map for the future scenario. Based on the river capture, the cost data and GWPI, the suitable locations for pumping have been delineated.

The objective of this research work is-

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

2. Methodology

The calibrated groundwater model has been developed and correspondingly calculated the river capture and cost maps for future stress periods. The use of forecasting techniques along with its hyper-parameter tuning for the generation of input time series for the groundwater model is discussed, which is further used for developing a capture map and cost map. The Saaty's Analytical hierarchical process (Saaty, 2014) has been used to calculate weights for WOA. In the end, river capture, cost of pumping and GWPI were used to detect suitable pumping locations.

< Insert Figure 1 around here >

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, river discharge, etc., is always a determining factor in the capture map's accuracy. In this study, we used LSTM and NAR while predicting stream discharge, rainfall, and stream water level. After comparing the prediction accuracy on test data, the most accurate method was utilized to forecast the time series data.

2.1.1. Long Short-Term Memory (LSTM)

LSTM is a special type of Recurrent Neural Networks (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 (Josefowicz et. al., 2015). Each memory unit in LSTM have different type of gates- 'forget gate', 'input gate', 'output gate' and 'modulation gate', to manage the memory.

< Insert Figure 2 around here >

Figure 2 provides the structure of the memory unit, which has input vector $x(t)$ and previous time step hidden state vector $h(t-1)$. The different type of layers like ‘LSTM layer’, ‘input layer’, ‘drop out layer’, ‘fully connected or dense layer’, and ‘output layer’ connects to form a neural network for forecasting. In this work, two network architectures are used (‘Single LSTM layer model’ and ‘Stacked LSTM layer model’) (**Figure 3**).

< Insert Figure 3 around here >

The hyperparameters are used to control the prediction accuracy as well as robustness of the model. The hyperparameters should be properly calibrated before the final predictions. Here, we used the Linear Bayesian Optimization to find the best set of hyperparameter values. Some of the important hyper parameters of LSTM network are:

Shape of input features: The shape of the input features is defined by the number of previous data sequences used in the prediction of the next time step.

Maximum epochs: The number of times a training vector is updated.

Initial learning rate: A high learning rate may cause the missing of the optimum solution, whereas a low learning rate causes slow convergence.

Mini-Batch Size: Instead of feeding the entire dataset into training, small batches of the data are formed and passed iteratively.

Regularisation weight: This term adds to the function to prevent the overfitting of the function over the dataset.

Number of hidden units in LSTM: Too many hidden units overfit the model, and its lower value underfit the model.

2.1.2. NAR (Non-linear Auto-Regressive neural networks)

NAR neural network applies a discrete non-linear autoregressive network model for forecasting data (Ruiz et al., 2016). The network consists of multiple hidden layers (n-layers) that perform the non-linear operations and an output layer that perform the regression operation. A lagged series of $p(x1)$, also known as feedback delay, is passed to the hidden layer. The number of hidden layers (n) and length of lagged series (p) are the parameters to be adjusted (Dufour & Piégay, 2010). NAR has an open network with lagged inputs and known outputs for calculating weights (W, b). After training, a close loop is formed with the hidden layer and output layer for the predictions (**Figure 4**).

< Insert Figure 4 around here >

2.1.3. Hyperparameter tuning with Linear Bayesian Optimization (LBO)

LBO is an efficient technique for the global optimization of functions with fully expensive evaluations (Snoek et al., 2012). MATLAB provides an inbuilt function (Bayesopt function) for implementing Linear Bayesian Optimization. Both the Models (LSTM and NAR) have been hyper-tuned with the Bayesopt function in MATLAB with Root Mean Squared Error as the objective function. A total of 6 and 2 parameters of LSTM and NAR have been tuned (**Table 1**).

2.1.4. Accuracy assessment of Forecasting Models

A standard metric for comparing values predicted by a model or estimator with the values observed is the root-mean-square error (RMSE). The RMSE is the square root of the mean of squared errors. Each error has an impact on RMSE that is proportional to the size of the squared error; as a result, more significant errors have an outsized influence on RMSE. The RMSE combines the sizes of prediction errors for different data points into a single indicator of prediction accuracy. The RMSE is calculated for the test data as

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Here, the error ($e_i = (x_i - y_i)$, $i = 1, 2, \dots, n$) has been calculated for the n sample of test data (x_i = Observed data, y_i = Predicted data)

2.2. Groundwater modelling and simulation

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

2.3. Capture Map & 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 non-linear 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 change in 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 -1000 m³/s was added for the cells in the study domain. Here, location is expressed as row (r) and column (c) because the MODFLOW divides the domain over space and time for simulation. By changing the location of the dummy well (r, c) and running the MODFLOW simulation, cost and capture data were calculated as:

$$\begin{aligned} & \text{Cost} = \sum_{j=1}^n \left(\frac{p}{d} \right) \cdot w(j) \\ & \text{Capture} = \sum_{j=1}^n \left(\frac{q(j)}{Q} \right) \cdot t(j) \end{aligned} \quad (2)$$

Where, r is the row number; c is the column number; i is the time step; $t(j)$ is the time interval length at j^{th} time step; d is the depth of groundwater from the surface; p is the price of pumping for 1000 m³/s of discharge for unit depth movement; $w(j)$ is the discharge from the feature; $q(j)$ is the discharge of the well. For this study, the river is the feature for the generation of the corresponding capture map. Also, we assumed the price of pumping (p) at a certain time step to be constant over space. The procedure can be

organized as the following algorithm (Leake et. al., 2010):

·	Read the .h5 input file & extract the well data from .h5 file	·	Set the discharge of dummy well a
---	--	---	--

2.4. Groundwater potential index (GWPI)

GWPI is a dimensionless quantity representing groundwater availability in an area. It has been determined using a weighted linear arrangement, as suggested by Venkatramanan et al., 2019.

$$\overline{\overline{(4)}}$$

Here, w_i is the Normalized weight of groundwater influencing factors calculated using the AHP method and w_j is the weight assigned to each subclass of the groundwater influencing factor. The thematic raster for different hydrological parameters has been prepared in the GIS environment, and the GWPI is calculated with the raster calculator. Initially, Ten groundwater influencing parameters (Elevation, Topographic wetness index, slope, Distance from river, drainage density, land use map, geomorphology, soil hydraulic conductivity, multi-resolution ridge top flatness and multi-resolution valley bottom flatness) have been selected for weighted overlay analysis. Expert analysis has 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 groundwater recharge and its movement, as the groundwater table frequently reflects the local terrain (Weiss, 2001). Groundwater potential is greatly influenced by soil hydraulic conductivity, which is a primary driver of groundwater recharge. The multi-resolution valley bottom flatness 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 (Gallant & Dowling, 2003). Since the active river aquifer exchanges can replenish the withdrawn groundwater, proximity to rivers and streams increases groundwater potential.

2.5. K-Means clustering

K-Means Clustering (MacQueen, 1967) is an Unsupervised Learning algorithm which groups the unlabeled dataset into different clusters, such that the sum of distances between the data point and their corresponding cluster centres is minimized. The three independent variables (dimension) viz: river capture, pumping cost and GWPI have been used to cluster the corresponding locations into 5 clusters, representing areas in different zones for groundwater development. The k-means clustering has been implemented with the sci-kit learn library (Pedregosa et al., 2011) in 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 (Dole-Olivier et al., 2019; Dufour & Piégay, 2010; Gaur, Chahar, et al., 2011; Gaur, Mimoun, et al., 2011; Graillot et al., 2014; Lalot et al., 2012; Lejot et al., 2007) have also contributed to the better understanding of the considered domain and its vicinity.

< Insert Figure 5 around here >

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. Hyper-parameters tuning with linear bayesian optimization was performed. The models were fine-tuned in the case of overfitting or underfitting. Among the three techniques- LSTM, stacked LSTM, and NAR, the one with the least (average) root mean squared error (RMSE) was selected for final forecasting. The recharge and evapotranspiration for the future time series have been taken as the ten-year moving average.

< Insert Figure 6 around here >

< Insert Figure 7 around here >

< Insert Table 1 around here >

3.3. Model Conceptualisation

The Ain River drains a basin that is around 3640 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 5). The average channel width and slope throughout this segment are around 60 metres and 1.3%, respectively. At Chazey-sur-Ain, the river's mean annual discharge was estimated to be 120 m³/s between 1959 and 2019. Low flow conditions characterize summer, and most floods occur between October and March. The 1 in 10 years flood recorded at Chazey-sur-Ain is 1200 m³/s. The river bed is made of gravel and cobbles with active meandering reaches in the lower valley (about 40 km from the confluence to upstream).

< Insert Figure 8 around here >

The model conceptualization has been done using available geological and hydrological data. The modelled aquifer thickness varies from 240m to 550m, as observed in the lithology data by BRGM (Bureau de Recherches Géologiques et Minières), where SRTM data was 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. The layer's average thickness was determined to be 25 m. For the older sediment, the initial estimates of hydraulic conductivity were taken to be 0.0018 m/s, and for the younger alluvium, 0.003 m/s. The aquifer hydraulic conductivity was estimated from the pumping test and the literature data obtained by BRGM. The collected data were utilized as starting values of parameters, which were then adjusted during the numerical model calibration to create the best match between the simulated and observed data. A specific yield (Sy) of 1 to 17% has been taken for the alluvial deposits.

3.3.1. Boundary Conditions:

The constant head and constant flow boundaries were used to define the model domain. Figure 8 demonstrates how the watershed division line and the no-flow barrier were used to determine the eastern and western boundaries.

3.3.2. River stage and observation wells data

The available observation well data was inconsistent in terms of temporal similarity. Out of the total available 20 observation wells, 8 wells had data from 2008 to 2010, while the 7 wells had from 2002 to 2015, and the rest have a varied temporal resolution. The water table variation is observed to be in the range of 1.2m to 3.7m, which shows a stable groundwater scenario. Further, the river stage data, measured at 5 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.

3.3.3. Evapotranspiration and Groundwater recharge:

Rainfall was solely taken into account as a way to recharge the groundwater 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. The recharge zones were defined using the land-use data. Six categories were taken for classification of Landsat 8 imagery to generate land-use data, such as 'water body', 'agriculture', 'fellow land', 'built-up', 'forest/vegetation', and 'sand'. The starting value of groundwater recharge was taken as 50%, 10%, 50%, 60%, and 80% of the rainfall for vegetation, built-up, agriculture, fellow land, and sandy area of total precipitation (1650 mm/year). With an extinction depth of 2 m, the projected annual evapotranspiration is 638 mm. The potential evapotranspiration was assumed to be evenly distributed. For the future forecast, the 10-year monthly rainfall average was taken, and corresponding recharge values were used.

3.3.4. Groundwater extraction:

Apart from irrigation, groundwater is utilized for domestic and industrial consumption. Water demand as per agricultural practices and domestic use was calculated based on field survey. Total agriculture water demand was also computed through the classification of satellite imagery data for three seasons. The groundwater abstraction varies seasonally as well as yearly due to variable demands for irrigation. Water demand data was collected from three sources (Association Syndicate of Ain Irrigation (ASIA), Directorate Department of Agriculture and Forestry (DDAF) and Rhône-Méditerranée-Corse Water Agency (AERMC)). Total Annual Withdrawal for Domestic, Industrial and Irrigation purposes are 8340700 m³, 5063204 m³ and 27000000 m³, respectively. Thus, the total catchments identified are 372, representing a total annual volume of 40,130,104 m³ which may additionally increase by 27 Mm³ in some conditions (ref: BRGM RP57945). The total water

requirement computed by this method is found to be 106.78 mcm which is 5% more than the value available by the administrative authority, i.e. 101.43 mcm.

3.4. Model Calibration

With stress periods spaced four months apart, a regional groundwater flow model was built and calibrated to the transient-state condition. For the calibration of the model, the variables hydraulic conductivity, recharge, and specific yield were used. The model was calibrated based on the observed heads of groundwater at 20 equally spaced observation wells. 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 was available for four observation wells, was used to validate the model.

< Insert Figure 9 around here >

3.5. River capture and pumping cost

The calibrated groundwater model was simulated for future stress periods by taking the forecasted time-series data as input. The model was then coupled with MODFLOW by modifying the hdf5 input files in MATLAB. The groundwater model has been used in identifying 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.

The river capture map was generated by considering a discharge well of -1000 m³/sec draft at each grid cell and calculating the river inflow and outflow change. The cost map was generated based on the mean groundwater head data (depth to the water table).

3.6. 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 assigned a weight based on the AHP method. Finally the normalized weights of all the parameters have been calculated with AHP, and GWPI has been calculated with the raster calculator.

< Insert Table 2 around here >

< Insert Figure 10 around here >

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. The forecasting by three hyper-tuned models over the testing

set for all three series (Pont-d'Ain ,Albarine_St Denis-en-Bugey and Chazey-sur-Ain) and the original series are shown in **Figure 6**.

It is observed from **Figure 6** and **Figure 7** that, unlike NAR, 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 have the least RMSE. For the hyper-tuned LSTM and NAR models, the optimum range of hyperparameters has been tabulated in **Table 1**. Most of the long-term forecast is highly uncertain, but the 4 months interval average RMSE was low. In addition, the granularity of 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 2060, groundwater flow was simulated with these data. Predicted heads through the model were found stable with a slight upward trend. Further, capture maps related to river-inflow and river-outflow were generated. Another type of cost 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. 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 1000 m³/s from the pumping well (**Figure 11**). Most of the points satisfying more than 85% 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 located 3 to 5 km away from the river can lower the water gained through the aquifer by 40% to 60%. This decrement in the water gain by aquifer can be due to the lower inflow into the river or greater outflow out of the river. Most of it is due to the lower inflow into the river.

< Insert Figure 11 around here >

The cost map represents the pumping cost which is assumed to be proportional to the depth of the groundwater. For the area close to the river within a 1 km buffer, the water table is below 10m of depth from the surface, mostly limited to around 1 meter. 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 17m. The GWPI values were extracted for the model domain for clustering along with river capture and cost.

K-Means clustering with three variables (river capture, cost and GWPI) has been used to classify the model domain in five zones of groundwater 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 groundwater inflow into the river or excessive water extraction through the river due to excessive pumping. The elbow method is utilized to determine the optimal number of five clusters. The clusters have been analyzed as

1.	<i>Low cost, high capture and high GW potential:</i>	These cells represent high river water capture, which has high g
2.	<i>Low cost, moderate capture and low GW potential</i>	These areas represent good pumping zones for short durations s
3.	<i>Low cost, high capture and low GW potential</i>	These areas have a low cost of pumping, but most of the water
4.	<i>Low cost, low capture and high GW potential</i>	These areas are best for new groundwater development projects
5.	<i>High cost, high capture and mixed GW potential</i>	These highly scattered zones are the worst locations for GW de

< Insert Figure 12 around here >

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 13**. 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. In this area, water depth is comparatively shallow, and groundwater flow is from Jura Mountain.

< Insert Figure 13 around here >

5. Discussion

This research work determines the suitable pumping locations for the future scenario based on the river capture, pumping cost and GWPI. The time series data have been forecasted for 40 years up to 2060 with LSTM and NAR. The forecasted data has been used for simulating the groundwater flow up to 2060 with MODFLOW. The capture and cost maps were generated for the simulation period. The GWPI has been calculated with AHP and WOA.

The prediction capability 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 of data, then NAR with the least RMSE. However, they both performed similarly for the 4-month average time series data. The groundwater model is an essential part of the whole research work, and the accuracy of the capture map depends on the groundwater 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. 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. A balanced abstraction of groundwater, considering the extent of stream capture, can prove to be a better way to manage our existing water resources. The accuracy of the capture fraction is highly dependent upon the complexity of the model. A calibrated physical based groundwater model with accurate aquifer properties can provide the realistic and accurate aquifer response to given pumping stress. The response of the aquifer to given stress is highly correlated to the capture fraction. Incorporating the capture map with groundwater potential mapping can reduce the uncertainty related to later methodology. As evident from figure 13, the proximity to river does not justify for the development groundwater extraction projects. Care should be taken for the intrusion of surface water in groundwater potential zonation methodologies.

The lower Ain River basin has a very stable groundwater 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 analyze 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 in the map satisfy nearly 65.54% of the total water demand from the river aquifer exchanges. R-A exchanges in the middle part of the study area, near Chazey-sur-Ain, were more affected by pumping, whereas the upper part near Pont-d’Ain was found to be less affected by pumping.

6. Conclusion

A novel methodology to determine suitable groundwater pumping locations based on the concept of capture fraction and groundwater 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 minimize the cost and the capture from the river in high groundwater potential zones. The LSTM and NAR have been utilized to predict time-series data to simulate the groundwater 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 groundwater and surface water resources. This research provides novel constrain for the future development of groundwater based on balanced groundwater and surface water resource management.

7. Declaration and Acknowledgements

The authors provide the following declaration:

Funding: No funding

Conflicts of Interests/Competing Interests: Not Applicable

Availability of data and material: Due to ongoing research, data and material pertaining to this work cannot be shared.

Code Availability: Due to ongoing research, codes pertaining to this work cannot be shared.

Authors’ Contributions: Gaur S., Ohri A., and Piegay H. drafted and devised the idea of the work and supervised its timely progress and guided in the technical and literature aspects. Gaur S. and Ohri A. created the groundwater models. Bajpai M. carried out the model simulations and forecasting. Kumar, R. drafted the results and formatted the literature accordingly. All authors discussed the results and contributed to the final manuscript.

Acknowledgement: The 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” program operated by the French National Research Agency (ANR) and through a joined 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).

8. 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*

Adv, 3(1), 51–57.

Bahrami, S., Doulati Ardejani, F., & Baafi, E. (2016). Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine. *Journal of Hydrology*, 536, 471–484. <https://doi.org/10.1016/j.jhydrol.2016.03.002>

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>

Barlow, P. M., & Dickerman, D. C. (2001). Numerical-simulation and conjunctive-management models of the Hunt - Annaquatucket - Pettaquamscutt stream-aquifer system, Rhode Island. In *US Geological Survey Professional Paper* (Issue 1636). <https://doi.org/10.3133/pp1636>

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models. *Journal of Hydrology*, 508, 418–429. <https://doi.org/10.1016/j.jhydrol.2013.10.052>

Corsini, A., Cervi, F., & Ronchetti, F. (2009). Weight of evidence and artificial neural networks for potential groundwater spring mapping: an application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2008.03.015>

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>

Cosgrove, D. M., & Johnson, G. S. (2005). Aquifer Management Zones Based on Simulated Surface-Water Response Functions. *Journal of Water Resources Planning and Management*, 131(2), 89–100. [https://doi.org/10.1061/\(asce\)0733-9496\(2005\)131:2\(89\)](https://doi.org/10.1061/(asce)0733-9496(2005)131:2(89))

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>

Daliakopoulos, I. N., Coulibaly, P., & Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309(1–4), 229–240. <https://doi.org/10.1016/j.jhydrol.2004.12.001>

Danandeh Mehr, A., & Kahya, E. (2017). A Pareto-optimal moving average multigene genetic programming model for daily streamflow prediction. *Journal of Hydrology*, 549, 603–615. <https://doi.org/10.1016/j.jhydrol.2017.04.045>

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>

Dufour, S., & Piégay, H. (2010). Channel vertical mobility, hydro-geomorphic disturbances and understory vegetation in floodplain forests of the Ain River (France). *Géomorphologie: Relief, Processus, Environnement*, 16(4), 371–386. <https://doi.org/10.4000/geomorphologie.8101>

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., & Gailliot, D. (2011). 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., & Graillet, D. (2011). 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>
- Graillet, 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*, 253.
- Ketabchi, H., & Ataie-Ashtiani, B. (2015). Evolutionary algorithms for the optimal management of coastal groundwater: A comparative study toward future challenges. *Journal of Hydrology*, 520, 193–213. <https://doi.org/10.1016/j.jhydrol.2014.11.043>
- Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., & Pradhan, B. (2019). Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeology Journal*, 27(1), 211–224. <https://doi.org/10.1007/s10040-018-1848-5>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (200 C.E.). *Rainfall – runoff modelling using Long Short-Term Memory (LSTM) networks*. 6005–6022.
- 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., Graillet, 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. *EGU General Assembly 2012*, 5352. <https://hal.archives-ouvertes.fr/hal-00772833>
- 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>
- 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/>
- 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. *5th Berkeley Symp. Math. Statist. Probability*, 281–297.
- 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>
- Naghibi, S. A., Vafakhah, M., Hashemi, H., Pradhan, B., & Alavi, S. J. (2018). Groundwater augmentation through the site selection of floodwater spreading using a data mining approach (case study: Mashhad Plain, Iran). *Water (Switzerland)*. <https://doi.org/10.3390/w10101405>
- 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.

- Napolitano, G., Serinaldi, F., & See, L. (2011). Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily stream flow series: An empirical examination. *Journal of Hydrology*, 406(3–4), 199–214. <https://doi.org/10.1016/j.jhydrol.2011.06.015>
- 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>
- Parchami-Araghi, F., Mirlatifi, S. M., Ghorbani Dashtaki, S., & Mahdian, M. H. (2013). Point estimation of soil water infiltration process using Artificial Neural Networks for some calcareous soils. *Journal of Hydrology*, 481, 35–47. <https://doi.org/10.1016/j.jhydrol.2012.12.007>
- 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. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena*. <https://doi.org/10.1016/j.catena.2015.10.010>
- Rizeei, H. M., Pradhan, B., Saharkhiz, M. A., & Lee, S. (2019). Groundwater aquifer potential modeling using an ensemble multi-adoptive boosting logistic regression technique. *Journal of Hydrology*, 579(July), 124172. <https://doi.org/10.1016/j.jhydrol.2019.124172>
- 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 Heirarchy Process. In *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.stat05310>
- Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>
- Saraf, A. K., & Choudhury, P. R. (1998). Integrated remote sensing and gis for groundwater exploration and identification of artificial recharge sites. *International Journal of Remote Sensing*, 19(10), 1825–1841. <https://doi.org/10.1080/014311698215018>
- Sarangi, A., Singh, M., Bhattacharya, A. K., & Singh, A. K. (2006). Subsurface drainage performance study using SALTMOD and ANN models. *Agricultural Water Management*, 84(3), 240–248. <https://doi.org/10.1016/j.agwat.2006.02.009>
- Sener, E., Davraz, A., & Ozcelik, M. (2005). An integration of GIS and remote sensing in groundwater investigations: A case study in Burdur, Turkey. *Hydrogeology Journal*. <https://doi.org/10.1007/s10040-004-0378-5>
- Seo, Y., Kim, S., Kisi, O., & Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, 224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Snoek, J., Rippel, O., & Adams, R. P. (2012). *Scalable Bayesian Optimization Using Deep Neural Networks arXiv:1502.05700v2*. <http://proceedings.mlr.press/v37/snoek15.pdf>
- 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>

- Supreetha, B. S., Shenoy, N., & Nayak, P. (2020). Lion Algorithm-Optimized Long Short-Term Memory Network for Groundwater Level Forecasting in Udupi District, India. *Applied Computational Intelligence and Soft Computing*, 2020. <https://doi.org/10.1155/2020/8685724>
- Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., & others. (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. In *GIS and Geostatistical Techniques for Groundwater Science*. <https://doi.org/10.1016/C2017-0-02667-8>
- Weiss, a. (2001). Topographic position and landforms analysis. *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., Ducharne, 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>
- Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S., & Behzadfar, M. (2016). GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environmental Earth Sciences*. <https://doi.org/10.1007/s12665-016-5424-9>
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, 561, 918–929. <https://doi.org/10.1016/j.jhydrol.2018.04.065>