Chapter 3: DEVELOPING AN INTEGRATED SIMULATION-OPTIMIZATION MODEL FOR MANAGING GROUNDWATER PUMPING-INDUCED R-A EXCHANGES

3.1 Introduction

Streamflow depletion due to groundwater pumping is influenced by several factors, including the geometry and hydraulic properties of aquifers and the distance between wells and nearby streams. Research suggests that the timing and magnitude of streamflow depletion are closely related to the rate of groundwater extraction, but it can continue to increase long after pumping has ceased. Understanding these dynamics is essential for effective water resource management, particularly in regions where groundwater is heavily relied upon for agricultural and municipal use. This chapter presents a detailed account of the materials, numerical models, and simulation optimization framework employed to evaluate different groundwater models and river-aquifer exchanges and their management. The study utilizes Finite Difference method-based 3-D numerical models, including MODFLOW for groundwater flow and MT3DMS for contaminant transport, coupled with metaheuristic optimization algorithms such as MOPSO, NSGA-II, MOEA/D, and Pareto Search. The GW modeling is done in two phases. The GW head model was conceptualized and calibrated in the first phase to account for accurate R-A exchange quantification. In the second phase, nitrate transport and nutrient loading were estimated from the agricultural areas to the Ain River reaches.

Metaheuristic algorithms were tested against the standard functions to check their performance in high-dimensional search spaces. The most effective ones are utilized for the lower Ain River basin (LARB) model. Bayesian optimization was used to tune the hyperparameters for all the algorithms. A graphical user interface platform is developed to facilitate the framework's application for policymakers and individuals who lack familiarity with programming.

3.2 Methods

3.2.1 Study area

The research focused on the LARB, comprising of Ain department with a population of 652,432. Ain River is a significant tributary of the Rhône River in France (Figure 3.1). In this segment, river Ain traverses diverse accumulations of coarse alluvium, with a mean annual discharge of 120 m^3/s, as measured at the Chazey-sur-Ain gauging station from 1959 to 2015. The discharge variability within the Ain River catchment is evident, particularly during the summer months when low flow conditions prevail, resulting in an average discharge of 54 m^3/s during July and August at Chazey-sur-Ain. The lower Ain River is relatively broad, ranging from 50 to 100 meters during low flow periods (Wawrzyniak et al., 2016), with an average wetted stream width of 51.5 meters (±15.9 meters) and an active channel width averaging 89.5 meters (±31.2 meters). The hydrology of this downstream section is heavily influenced by the Vouglans Dam, a significant hydroelectric power plant located approximately 60 km upstream of the study area. The influence of the Vouglans Dam and a series of four additional dams has resulted in a significant sediment deficit in the lower Ain River (Rollet et al., 2014). This sedimentary disruption has led to a downstream reduction in the D50 value, representing the sediment distribution's median particle diameter. Specifically, the D50 value decreases from 18 mm in the first 7 km of the river to 16 mm between 7 and 13 km and further to 14 mm in the final stretch from 13 km

to the confluence with the Rhône River. The decreased sediment size and the riverbed's resulting degradation have altered the riverbed materials' hydraulic conductivity, potentially affecting the exchange rates between the river and the underlying aquifers. These alterations in sediment dynamics have had a profound impact on the R-A exchanges. Further, the region is characterized by complex hydrogeological conditions, including highly permeable graveled riverbeds and significant anthropogenic influences. Understanding these dynamics is crucial for assessing the long-term sustainability of groundwater resources in the region, particularly considering the reduced sediment transport and altered hydrological conditions caused by river damming (Wawrzyniak et al., 2016).

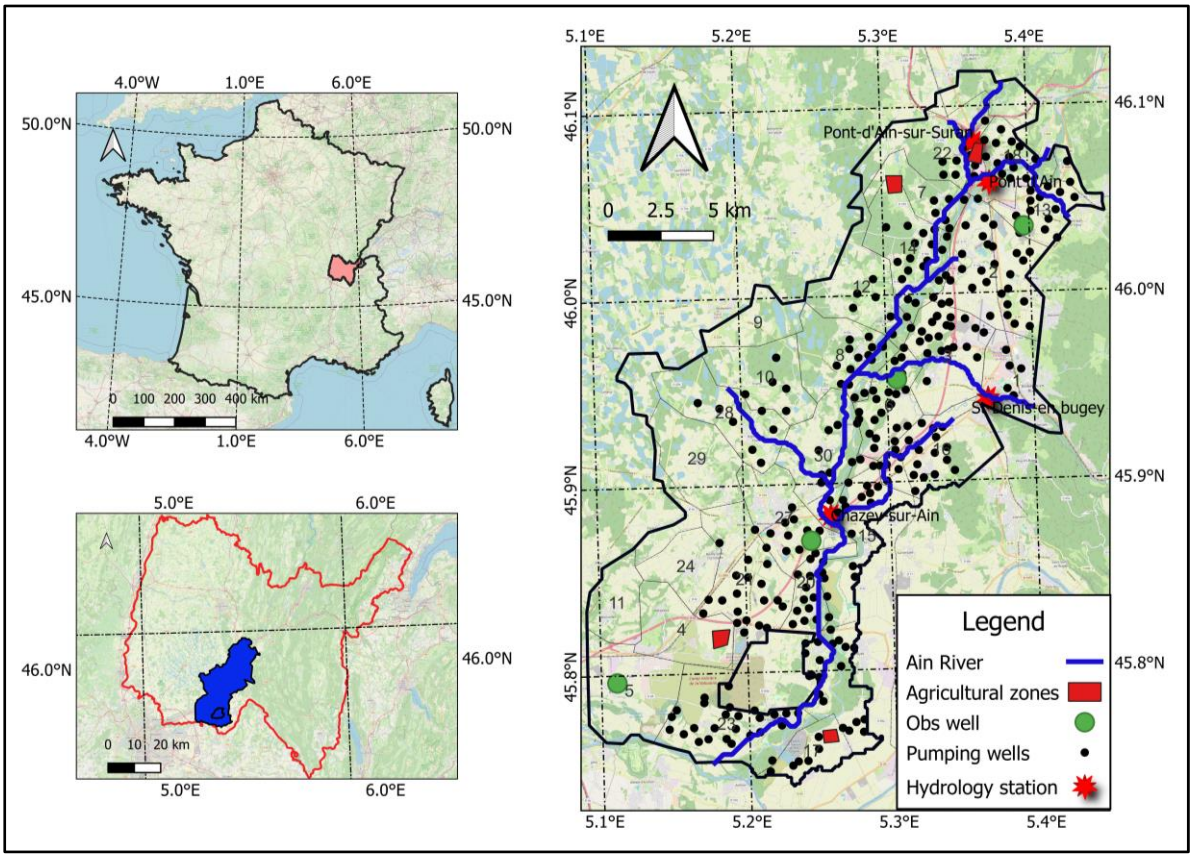

Figure 3.1 Lower Ain River basin with pumping wells, observation wells, and river stations.

3.2.1.1 Hydrogeological settings

The Ain River traverses an extensive alluvial plain with a gentle slope of approximately 1.3%. This plain consists of modern alluvial deposits that overlie fluvio-glacial sediments, which vary in thickness from 5 to 30 meters. Beneath these deposits lies the Miocene molassic substratum, which serves as the impermeable base of the geological structure. The alluvial plain is further characterized by two prominent morainic hills, oriented in a Northeast-Southwest direction, significantly shaping the river's downstream flow path. Since the 1950s, the riverbed of the Ain River has undergone significant incision, a process that the construction of upstream dams has exacerbated. This has led to a transition in the river's morphology from a braided to a more meandering fluvial form, particularly evident at the base of the morainic hills, such as in the large meander near Chazey-sur-Ain. The incision has resulted in the river becoming increasingly embedded within a narrow thalweg, also oriented Northeast-Southwest, surrounded by fluviatile alluvia deposits that range in thickness from 2 to 5 meters. The variability in the thickness of these deposits is considerable, with depths reaching up to twenty meters near Chazey-sur-Ain but thinning to complete erosion, exposing the molassic substratum, particularly near the river's confluence with the Rhône River.

The aquifer exhibits a gradient of approximately 2% within the study area, with hydraulic conductivity values ranging from 10^{-1} m/s in the fluviatile alluvia to 10^{-5} m/s in the molassic substratum. These conductivity variations contribute to a resultant groundwater flow velocity of about 15-35 m/day. The lateral groundwater flows in this region are predominantly oriented from North to South and are actively drained by the Ain River. The morainic hills influence groundwater flow in the valley's southern part, directing it into two distinct corridors. The western corridor channels groundwater towards the Rhône River at

the base of the Dombes plateau, where no active river channel is present at the surface. Different hydrogeological properties are shown in Figure 3.2.

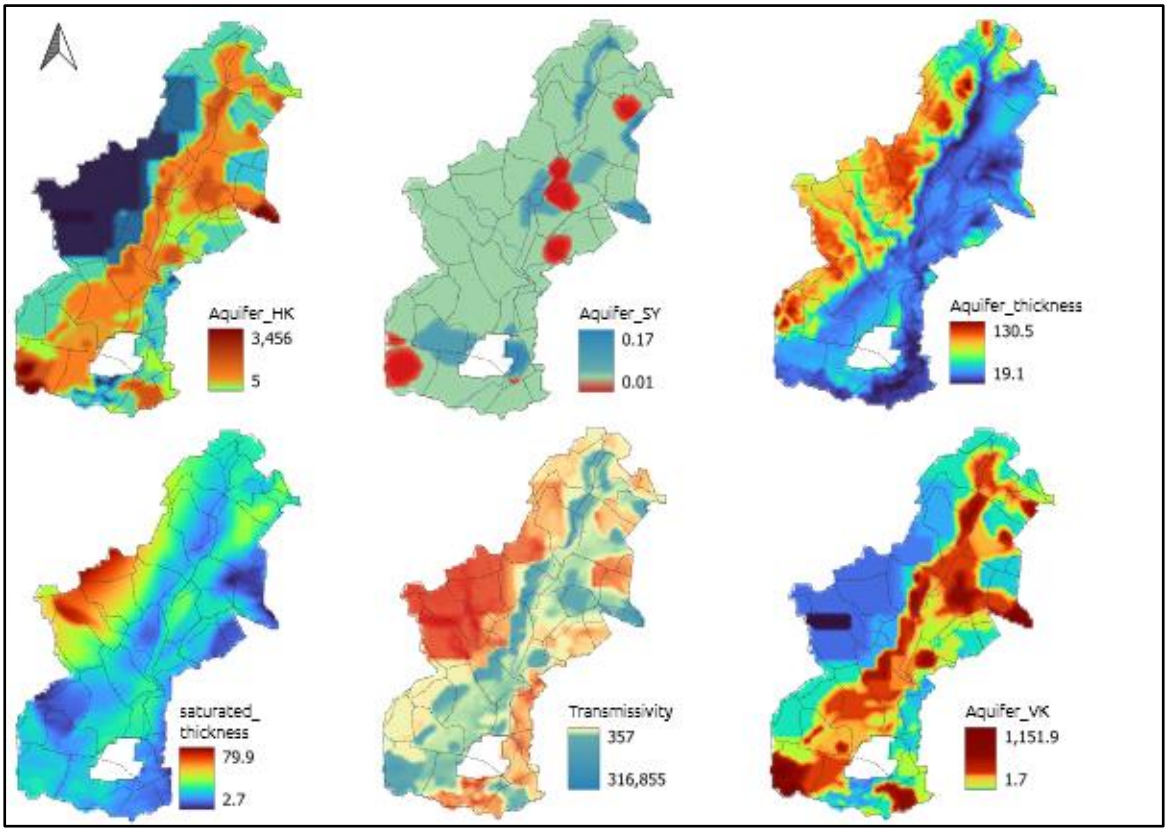

Figure 3.2 Hydrogeological properties of the aquifers in the study area: Hydraulic conductivity (HK), specific yield (SY), aquifer thickness (thickness), saturated thickness (saturated thickness), transmissivity, and vertical conductivity (VK).

Conversely, the eastern corridor directs groundwater between the two morainic hills underneath the current river channel, particularly near the Chazey-sur-Ain meander. The aquifer within this region is heavily exploited to meet irrigation, industrial, and domestic demands, with annual extraction volumes ranging from 35 to 62 million cubic meters.

Major crops in this region include cereals such as wheat and maize, oilseeds such as sunflower mustard, fruits, and vineyards. Extensive fertilizer use with consistently high nitrate levels has deteriorated the shallow aquifer. The increase in GW extraction within the last few years has aggravated the situation. This extensive groundwater withdrawal has the

potential to lower the piezometric levels, leading to a deficit in groundwater contributions to the river, which can significantly impact river-aquifer exchanges, particularly during periods of low flow and elevated temperatures (DIREN Rhône-Alpes, 1996; SBVA, 2014). Such conditions are critical, as they can alter the natural balance between the river and its underlying aquifer, potentially affecting the sustainability of the region's surface water and groundwater resources.

3.2.1.2 Groundwater and River stage-discharge

For the region, a total of 15 piezometers were available, where a few piezometers had data from 2008 to 2010, and the remaining piezometers covered the data from 2002 to 2015. Some piezometers were outside the boundary domain. At last, data thus collected showed that the piezometer does not show extremely high fluctuation in values. This variation ranged from 4 m to 2.2 m, showing that the region's groundwater table and the groundwater scenario are stable. Further, the river water stage data, measured at 5 locations in the study area, was obtained from Banque Hydro, France. The data from 2002 to 2015 was examined to understand the trend. This analysis showed a fluctuation range of 2.54 m at Chezy and 1.67m at Pont d'Ain on the river Ain. Data of other tributaries, such as Albarine also showed a maximum fluctuation of 1.36m. The boxplot in Figure 3.3 illustrates the monthly discharge variability for three river stations—Chazey-Sur-Ain, Pont d’ Ain, and Lagnieu—over the period 2002–2019. Among these, Lagnieu consistently records the highest discharge, highlighting its role as a major downstream station with substantial flow contributions to the river Rhone.

In contrast, Chazey-Sur-Ain and Pont d’ Ain exhibit lower discharge values and a relatively narrow interquartile range, indicating more stable flow conditions. All stations display clear seasonal variability, with higher median discharges observed during late spring and early summer (May to June), likely driven by snowmelt and increased seasonal precipitation.

Notably, Lagnieu demonstrates a wider spread and a higher number of outliers, reflecting more frequent extreme flow events, which can be attributed to its downstream position and the cumulative impact of upstream tributaries.

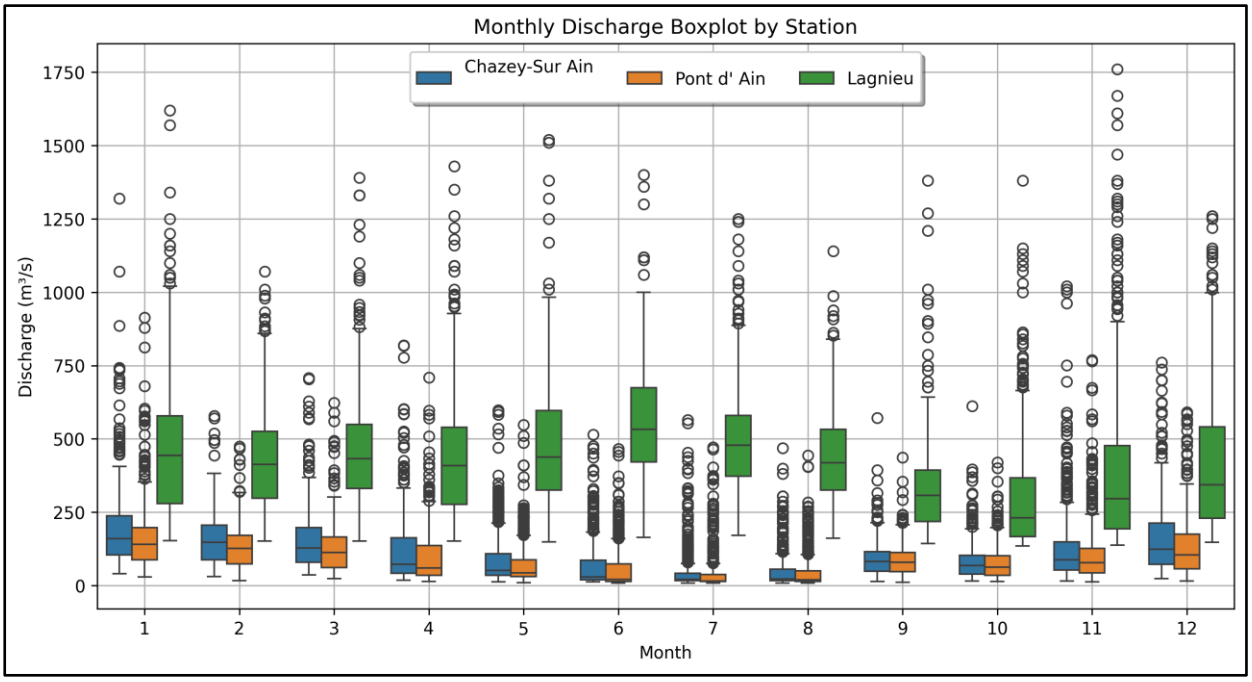

Figure 3.3 Monthly discharge variability at three river stations (Chazey-Sur-Ain, Point de Ain, and Lagnieu) from 2002 to 2019

3.2.1.3 Recharge and Evapotranspiration

Rainfall was considered as the source of groundwater recharge. Rainfall data from the Météo France database was used to calculate the recharge input values and applied uniformly over the polygons constructed in the model domain. The different recharge polygons were developed based on a land-use map created from satellite imageries. Six categories were taken for land use classification: water body, agriculture, fallow land, built-up, forest/vegetation, and sand. Supervised classification was performed to create the land-use map where the fusion of Sentinel 1 and Landsat 8 satellite imageries was used.

Random Forest model was used as a classifier, and training accuracy, test accuracy, and kappa coefficient were 99.98%, 98.45%, and 97.77%, respectively. The initial recharge value was 10%, 50%, 50%, 80%, and 60% for built-up, agriculture, vegetation, sand, and

fallow land of total precipitation, i.e., 1650 mm/year. The Evapotranspiration (EVT1) package in MODFLOW is critical in accurately simulating water loss from the water surface due to evapotranspiration. The basin's average annual sunshine duration of 2,006 hours is moderately high in France's diverse climatic regions. Depending on the area, it surpasses the national average, typically 1,500 to 2,000 hours. This level of sunshine contributes to a moderate evapotranspiration rate, facilitating agricultural productivity while influencing local hydrological dynamics. Estimated evapotranspiration was 638 mm/year with an extinction depth of 2 m. The potential evapotranspiration was considered uniformly distributed over the study area. For the future forecast, 10 years' monthly rainfall averages were taken, and corresponding recharge values were used.

3.2.1.4 Agriculture and Land Use

Regarding land use, the upper Ain River section is dominated by forestry and agricultural activities, mainly milk and cheese production pastures. Then, the middle Ain Valley presents 22% of urban and industrial areas, especially along the Bienne Rivers at Oyonnax, Nantua. This area is suspected to offer a high potential for agricultural pollution due to a large farming area. The Middle Ain River also hosts major dams, such as the Vouglans arch dam (height: 103 m), while run-of-the-river dams are established at Coiselet, Cize-Bolozon (CB) and Allement (ALL) on the Ain River itself. The lower Ain River Basin in France is characterized by diverse land use patterns, with agriculture being one of the dominant activities in the region. The fertile alluvial plains along the river support several types of agriculture, including cereal crops, vineyards, and pasturelands.

Irrigated agriculture is prevalent, particularly in areas closer to the river, where water availability has facilitated the cultivation of high-value crops. These agricultural activities contribute significantly to the local economy and exert considerable pressure on water resources, particularly during dry seasons when irrigation demand is high. The region's land

use is further shaped by mixed forest cover and urban areas, which influence the basin's hydrological dynamics. Based on data from the ADES portal, four primary agricultural zones were identified where nitrate levels in groundwater exceed permissible limits. Three crop types (wheat, maize, and oilseeds) were taken for study in each irrigation zone. Due to their high solubility and limited biodegradability, nitrates can persist in groundwater for extended periods, leading to significant environmental and health concerns.

3.2.1.5 Water demand and supply

Water consumption in the study area is mainly by agriculture, the rest for domestic purposes and some for industrial usage. The pumping rate was identified based on the area's agricultural, domestic, and industrial demands. In the context of agricultural irrigation, the simulation of groundwater demand is intricately linked to the specific irrigation requirements of various crops. The scheduling and allocation of irrigation periods are further influenced by the crops' growing habits and local agricultural practices, which are deeply rooted in the cultivation traditions of the region. Consequently, the return flow from agricultural irrigation, which represents the portion of applied water that percolates back into the groundwater system, is an essential factor incorporated into the groundwater model. This return flow not only replenishes the aquifer but also affects the overall water balance and the sustainability of groundwater resources in agricultural areas.

The groundwater abstraction varies seasonally and yearly due to variable demands for irrigation. The quantitative data, available from 2002, are obtained from three sources - the Rhône-Méditerranée-Corse Water Agency (AERMC), the Directorate Department of Agriculture and Forestry (DDAF), and the Association Syndicale of Ain Irrigation (ASIA). The groundwater abstraction is divided into three uses (Table 3.1). Thus, the total catchments identified are 372, representing a total annual volume of 40,130,104 m^3. which varies up to 27 Mm^3 in some conditions. 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.

Table 3.1 Different Crop Types and GW Demand in the Lower Ain River Basin

Crop type	Irrigation percentage
Meadows and dairy	46 %
Cereals (wheat, maize)	38%
Legumes and oilseeds	8%
Fodder	4%
Others	4%
Annual domestic groundwater abstraction	8,340,700 (m^3)
Annual industrial demand	5,063,204 (m^3)
Total irrigation demand	27,000,000 (m^3)

Given the dynamic nature of R-A exchanges, it is also essential to consider how land use and climatic patterns changes might affect nitrate transport and distribution. For instance, increased agricultural activity, urbanization, or shifts in precipitation patterns due to climate change can alter groundwater recharge rates, nitrate concentrations, and the timing and magnitude of nitrate fluxes to streams. Based on data from the ADES portal, four primary agricultural zones were identified where nitrate levels in groundwater exceed permissible limits. Due to their high solubility and limited biodegradability, nitrates can persist in groundwater for extended periods, leading to significant environmental and health concerns. Elevated nitrate concentrations pose a particular threat to human health. Additionally, nitrate pollution can have detrimental ecological effects on aquatic ecosystems, impacting the health of aquatic flora and fauna and degrading overall water quality.

To address these challenges, integrated hydrological models that simulate groundwater flow and nitrate transport are needed. Additionally, targeted monitoring and data collection efforts are necessary to validate model predictions and improve our understanding of nitrate

dynamics in coupled stream-aquifer systems. Thus, the conceptual models are developed for the lower part of the River Ain to identify the upwelling and downwelling exchanges between the stream and aquifer. The methodology of this thesis integrates simulation and optimization models to address groundwater management within the WEFE Nexus. The simulation model comprises two core components: the Nitrate Simulation Model (MT3DMS), which estimates nitrate loading to the stream reach, and the Groundwater Flow Simulation Model (MODFLOW), which simulates groundwater levels in wells. The optimization phase employs metaheuristic algorithms, including MOPSO, NSGA-II, MOEA/D, and Pareto search, to determine optimal groundwater discharge while maximizing river leakage. The general working flowchart for the algorithms is given in Figure 3.4. Finally, a decision-making process evaluates the trade-offs among competing solutions, enabling the formulation of optimal strategies for sustainable groundwater management and river-aquifer exchange. This comprehensive approach integrates physical simulation and advanced optimization techniques to improve water resource management in a complex, multi-objective framework. These models are integrated with optimization algorithms implemented in Python to determine optimal groundwater management and focus on improving R-A exchanges.

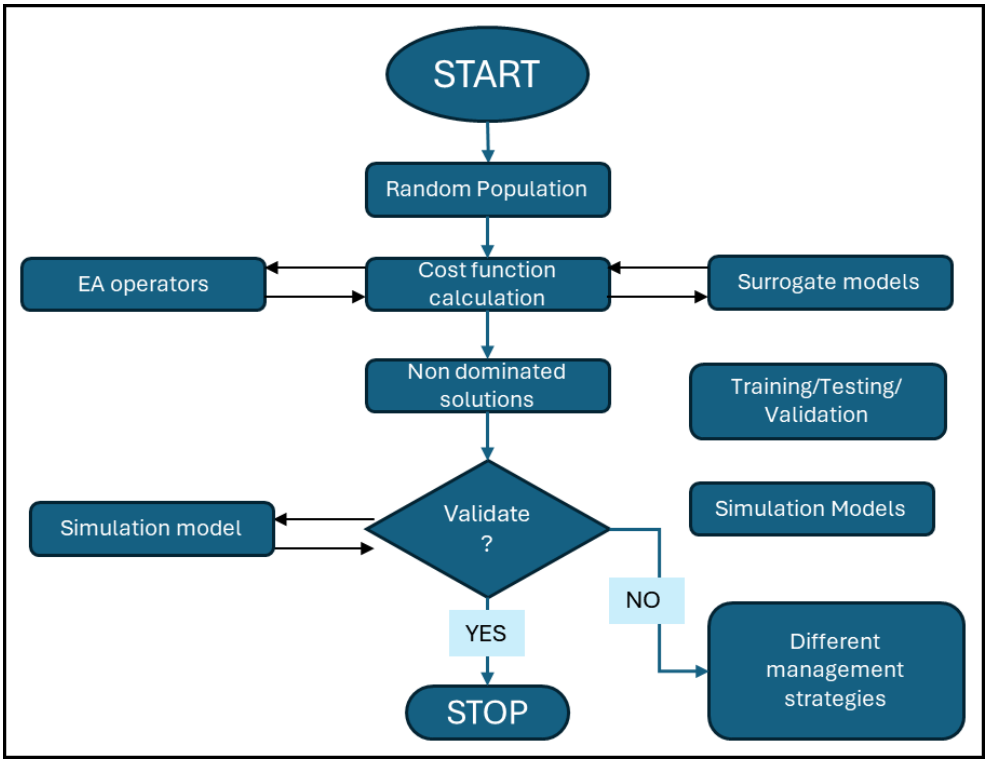

Figure 3.4 Workflow of evolutionary algorithms in groundwater S-O problems

3.2.2 Numerical modeling

In this study, numerical modeling is organized in two phases. MODFLOW first simulates transient grid-based groundwater flow to identify the region's hydrological flow patterns and upwelling/downwelling behaviors. The groundwater flow field is then coupled to a reactive transport model based on MT3DMS to simulate nitrate contamination and its transportation to the nearest river reach, reflecting the impacts of agricultural practices on alluvial groundwater and the stream. The general mathematical equation describing groundwater flow contamination fate, and transport is expressed as:

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) + W = S_y \frac{\partial h}{\partial t} \qquad (3.1)$$

$$\frac{\partial}{\partial x_i}\left(\theta D_{ij} \frac{\partial C^k}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(\theta v_i C^k) + q_s C_k^s = \frac{d(\theta C^k)}{dt}, k = 1,2,\dots,m \qquad (3.2)$$

$$\frac{dC_i^*}{dt} = \pm r_{im}, \quad im = 1,2, \ldots, (n-m) \tag{3.3}$$

Where:

K_x, K_y, K_z= Hydraulic conductivity in the (x), (y), and (z) directions respectively [LT^{-1}]; h is the hydraulic head (potential) [L], W is the source/sink term (e.g., recharge or pumping), i.e., a volumetric flow rate of sources and sinks per unit volume of aquifer [T^{-1}]; S_y is specific storage coefficient in a confined aquifer, and specific yield in unconfined aquifer, t is time [T^{-1}]; i represents coordinate axes aligned with the principal directions of permeability; h is the equivalent freshwater hydraulic head [L]; Z is the elevation [L]; θ is effective porosity; C is solute concentration [ML^{-3}]; D_{ij} is the hydrodynamic dispersion coefficient [L^2T^{-1}]; n is the total number of species, m is the total number of mobile species; C_k is the mobile concentration of species k [ML−3]; v_i is the seepage velocity [LT^{-1}]; C_s^k is the source or sink flux concentration for species k [ML−3].

3.2.3 MODFLOW model

Based on the conceptual model approach, the groundwater flow modeling was performed using MODFLOW 2005 (GMS) 10.7. It was adopted to create different geospatial data-based input layers for defining the area's surface recharge, boundary conditions, discharge wells, observation wells, and hydraulic conductivity. The Ain River, a significant surface water feature in the study area, is modeled to capture the dynamic exchanges between the river and the underlying aquifers using the stream (STR) package in MODFLOW 2005. The Stream package is utilized for modeling streams within a model. The flow in a stream is routed immediately to downstream streams. These interactions are governed by the hydraulic conductivity of the riverbed, which in this case is specified as the presence of an alpine river with a gravel stratum. The stream package allows for the simulation of both gaining and losing streams, where the direction and magnitude of water exchange depend

on the river's relative levels and the groundwater table. Within the STR package, the calculation of water flow from a river to an aquifer is performed using different techniques for hydraulically connected and disconnected systems. In MODFLOW terminology, hydraulic-connected connected streams occur when the water table elevation exceeds the streambed sediments' base. The interchange volumetric flow Q_{RAE} [L^3T^{-1}] between the river and the groundwater is determined in this scenario by employing the following calculation:

$$Q_{RAE} = \frac{K_c L w}{h_c}(h_{riv} - h) = C_{riv}(h_{riv} - h) \quad (3.4)$$

K_c is the hydraulic conductivity of the clogging layer, L is the length of the river within a cell, w is the width of the river, h_c is the thickness of the streambed layer, h_{riv} is the hydraulic head of the river, h is the groundwater head, and c_{riv} is the conductance of the streambed layer. Hydraulic conductance is a lumped measure that combines the river's geometry, the bed layer, and hydraulic conductivity. In MODFLOW, the stream-groundwater system is considered hydraulically separated when the water table h is below the elevation of the streambed bottom z_a. In this instance, the volumetric infiltration flux Q_{RAE} from the river to the aquifer is determined by:

$$Q_{MF} = c_{riv}(h_{riv} - z_a) \quad (3.5)$$

3.2.4 Model Conceptualization

The spatial discretization of groundwater models is often constrained by the need to balance computational efficiency and numerical accuracy. While finer spatial resolutions can enhance the representation of aquifer heterogeneities, they also increase computational demands. Conversely, coarser discretization reduces computational burdens but may oversimplify the modeled system. To balance this, a local grid refinement model was also constructed for a reach of the river Ain. In the model, a 250*250 grid size was chosen for the underlying sediments, where each layer was assumed to be horizontal and isotropic, as

shown in Figure 3.5. The topography, top, and bottom elevation of the study area were in the range of 240 m to 550 m where Shuttle Radar Topography Mission (SRTM) data was used to get the surface elevation of the area. The piezometric surface of groundwater was created using hydrograph data from 280 wells in the study area. The bottom surface was prepared with the help of well-log data obtained by BRGM. The mean thickness of the layer was 25 m.

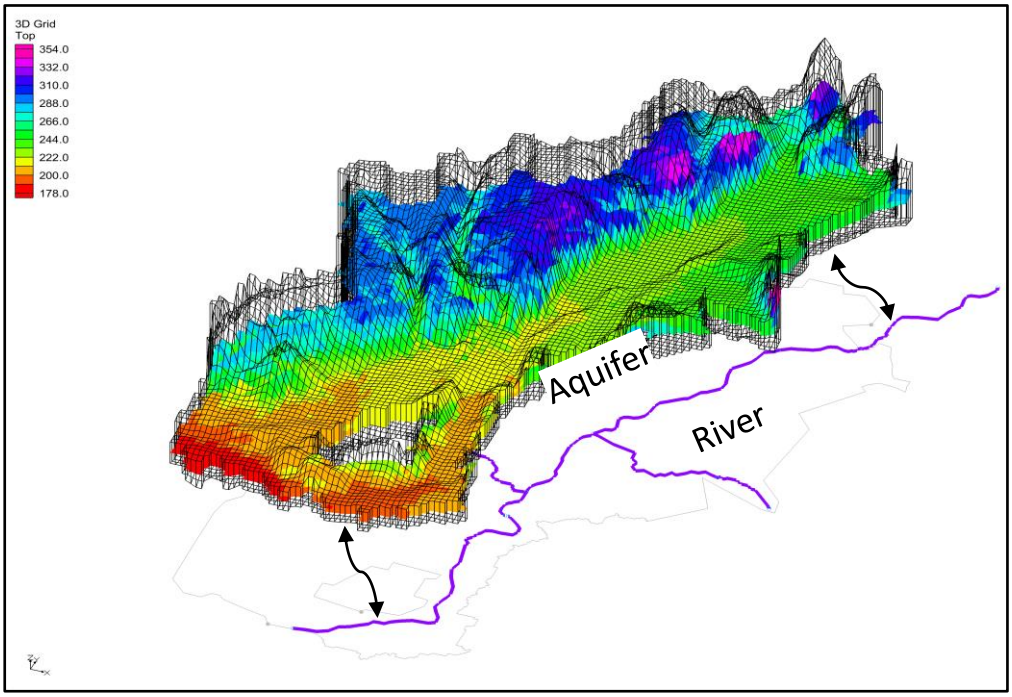

Figure 3.5 A 3-D conceptualization of the groundwater model

The initial hydraulic conductivity values varied from 0.0018 m/s for the older sediment and 0.003 m/s for the younger sediment. The horizontal hydraulic conductivity was estimated from the pumping well test data obtained from BRGM. Riverbed conductivity, which plays a crucial role in determining the R-A exchanges, was obtained from BRGM and Hydroportail, EAU France. Specific yield (*Ss*) values for the alluvial deposits ranged from 1 to 17%. Different other parameters are summarized in Table 3.2. Thus, the obtained data have been used as initial distributions that were subsequently modified during the Calibration of the numerical model to achieve the best fit between the simulated and the observed data.

Table 3.2 Different properties used for the simulation model

Property Name	Value
Aquifer thickness (m)	(18-180)
Hydraulic conductivity (m/d)	(5-3456)
Anisotropy ratio	1
Storage coefficient	9.9e-06
Grid size (m*m)	250*250
Stress periods	23
Time steps (d)	120
Porosity	30%
Specific yield	(0.005-0.173)
Recharge rate (m/d)	(0-0.0039)
Evapotranspiration rate (m/d)	(0.01-0.001)
Longitudinal dispersivity (m)	40

3.2.5 Boundary Conditions

In numerical models, three domains were considered to study the effect of changing boundary conditions (BC) on the river-aquifer exchanges. These domains characterize the hydraulic response of the boundaries to the groundwater condition and stream capture. In Domain-1, the model's central east and west sides were defined based on the watershed divide line, i.e., no-flow boundary. The bottom east and top west parts were considered specified flow from the outer system. In Domain-2, the alluvium plain of the Ain River was considered the eastern BC with specified head conditions, while the western sides of the model domain were defined based on specified flow boundaries. Domain-3 was chosen the same as Domain-1 with one modification in the lower eastern part. The Rhone River was introduced in the model through the specified head boundary. This modification helped to find the impact of incorporating the Rhone River in the modeling area. The model was developed by defining two types of external model boundaries, i.e., specified head and flow boundaries. Figure 3.6 shows the three types of domains considered to perform groundwater modelling.

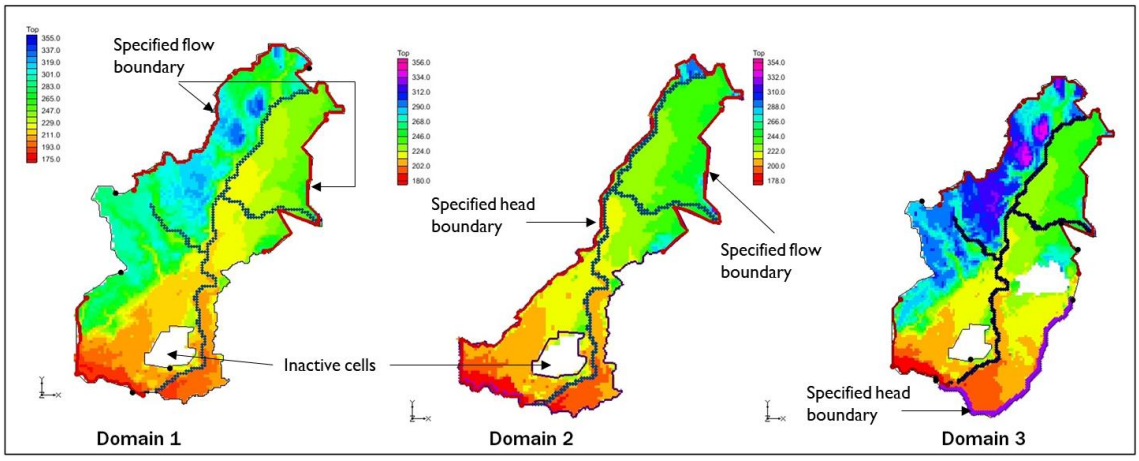

Figure 3.6 Different domains were conceptualized to study the effect of varying boundary conditions on GW flow budget and R-A exchanges.

3.2.6 Sensitivity analysis

Sensitivity analysis helps to find the parameters which have more influence on the model. Sensitivity analysis not only helps choose the most important parameters influencing the model outputs, but it also helps reduce the number of parameters for calibration. This reduces the computational time required by the optimization model in the identification of the optimal values of parameters. The Sobol method provides a quantitative measure of the influence of each input parameter on the model output, accounting for both individual effects and interactions between parameters. The first-order Sobol index (SI) represents the contribution of a single parameter, while the total-order index (ST) captures the total effect of a parameter, including interactions. These indices range between 0 and 1, with higher values indicating greater output sensitivity to that parameter. Using Sobol indices can help identify the most critical parameters controlling groundwater flow and solute transport and the strength of surface-subsurface coupling. Therefore, parameters like hydraulic conductivity specific storage, specific yield, and riverbed conductance were taken for sensitivity analysis, and their effect on the hydraulic head and discharge was studied. The most straightforward method for sensitivity analysis is local sensitivity analysis (LSA), in

which every input is varied one at a time, keeping the others unchanged. This approach does not account for input parameter uncertainty. Global sensitivity analysis (GSA) is performed to capture the effect of input parameter uncertainty. In this work, LSA is performed to get a broad picture, and Morris's index and Sobol index are calculated for GSA to obtain the effect of parameter uncertainty.

The Morris indices (u*) represent the average elementary effect of uncertain input parameters. To compute u*, random sampling in the parameter space P forms the 'r' trajectory of total size Nr * Np. N_r (250) is the number of random points sampled for trajectory, and N_p (5) is the number of input parameters. Then, the elementary effect on the deviation of parameter x_i from its tentative value along the j^{th} trajectory is given as:

$$EE_{x_i}(j) = \frac{f(x_1, \dots, x_i + \Delta, x_5) - f(\boldsymbol{x})}{\Delta} \tag{3.6}$$

From this elementary effect, the u* can be calculated as:

$$\mu^*_{x_i} = \frac{1}{r}\sum_{j=1}^{r}\left|EE_{x_i}(j)\right| \tag{3.7}$$

The number of simulations required to compute u* is r*(N_p + 1). To obtain a stable value, 'r' was set to 50; hence, three hundred simulations were performed. Though Morris's index overcomes this, the Sobol index is used.

(Zhou et al., 2019) has shown that if the model response f(x_i) square integrable is a function, then it can be written as:

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{N_p} f_{x_i}(x_i) + \sum_{1\le i\le j\le N_p} f_{x_i,x_j}(x_i, x_j) + \dots + f_{x_1,\dots x_{N_p}}\left(x_1, \dots, x_{N_p}\right) \tag{3.8}$$

where f_0 is constant, representing the expected value of f(x), the decomposed functions are orthogonal functions concerning a probability measure. The total variance, V(f(x)), of f(x) can be written as:

$$V\big(f(\boldsymbol{x})\big) = \sum_{i=1}^{N_p} V_{x_i} + \sum_{1\le i\le j\le N_p} V_{x_i,x_j} + \ldots + V_{x_1,\ldots x_{N_p}} \tag{3.9}$$

where V_{x_s} defines the combined effect of variables in $\boldsymbol{x_s}$ on model response and $\boldsymbol{x_s}$ is a subset of input parameters. The Sobol's indices, S_{x_s} are defined as:

$$S_{x_s} = \frac{V_{x_s}}{V\big(f(\boldsymbol{x})\big)} \tag{3.10}$$

Also, the total Sobol index for i^{th} input parameter is given as follows:

$$S_{x_i}^T = S_{x_i} + \sum_{j} S_{x_i,x_j} + \sum_{j,k} S_{x_i,x_j,x_k} + \cdots + S_{x_1,\ldots x_{N_p}} \tag{3.11}$$

Computing the Sobol index for time-intensive models can pose significant computational difficulties. To address this issue, a surrogate model is utilized of the study.

3.2.7 MT3DMS model

MT3DMS is a flexible three-dimensional transport model designed to simulate the processes of advection, dispersion, and source/sink mixing involving dissolved substances within groundwater systems. It relies on the hydraulic head, cell-by-cell flow distribution, and groundwater source and sink information (such as pumping, recharge, and surface water-groundwater exchange flows) provided by a MODFLOW model to establish the flow field within the aquifer system. This integration allows MT3DMS to simulate solute transport processes under varying flow regimes, enabling comprehensive analysis of groundwater contaminant dynamics. Lee et al. (2006) put forward a nitrogen transformation module for MT3DMS that accounts for ammonium oxidation, nitrification, denitrification,

organic carbon oxidation, and microbial reactions, which reveals the fate and transport of NH_4^+, NO_2^-, NO_3^-, N_2, DOC, and O_2. Eqs. (3.6)– (3.9) depict the primary reactions.

$$NH_4^+ + 1.5O_2(aq) \rightarrow NO_2^- + 2H^+ + H_2O \quad (3.12)$$

$$NO_2^- + 0.5O_2(aq) \rightarrow NO_3^- \quad (3.13)$$

$$CH_2O + 0.8NO_3^- \rightarrow 0.4N_2 + 0.8HCO_3^- + 0.2CO_2(aq) + 0.6H_2O \quad (3.14)$$

$$CH_2O + O_2(aq) \rightarrow CO_2(aq) + H_2O \quad (3.15)$$

The effects of nitrate leaching (NL) on hydrological processes are significant, particularly in regions experiencing rapid agricultural expansion (Hansen et al., 2017). Accurately quantifying nitrogen loss and its influence on hydrological components within the LARB is crucial for understanding the broader environmental impacts. In this study, nitrate leaching (expressed as kg N/ha/yr) was calculated using an empirical formula developed by Pratt (1984):

$$NL = 0.20 \times (N_i \times D)^{0.712} \quad (3.16)$$

Where N_i represents the nitrogen fertilizer input (kg/ha/yr), and D denotes the drainage (cm/yr), which is determined by the total infiltration recharge, including precipitation and irrigation return recharge (Yang et al., 2021a). This calculation highlights the importance of understanding nitrogen fluxes within agricultural landscapes and their downstream effects on groundwater and surface water systems. Nitrate leaching contributes to groundwater contamination and the nutrient loading of adjacent water bodies, potentially leading to eutrophication and other ecological disturbances.

Nitrate transport within the lower Ain basin was modeled using the MT3DMS, which specializes in simulating solute transport through groundwater systems. Main MT3DMS packages based on the governing principles were used in establishing the nitrate dispersion model. They are described below.

3.2.7.1 Advection

Advection is the process by which solutes are transported through groundwater due to the bulk movement of water. In nitrate transport, advection is critical as it dictates how quickly nitrate can move from one location to another within the aquifer system. The governing equation for advection can be expressed as:

$$\frac{\partial C}{\partial t} + v.\nabla C = 0 \tag{3.17}$$

Where C represents the concentration of nitrate (mg/L), t denotes time (days), v is the velocity vector of groundwater flow (m/day), which indicates the direction and speed at which water—and thus nitrate—is moving, and ∇C represents the spatial gradient of concentration, indicating how concentration changes in space.

In MT3DMS, this advection term is solved using various numerical methods, such as finite-difference or method-of-characteristics approaches. This thesis chose the third-order total variation diminishing (TVD) scheme as the solution scheme in the advection package.

3.2.7.2 Dispersion

Dispersion describes how solutes spread out from areas of high concentration to low concentration due to variations in flow velocity and molecular diffusion. This process is crucial for understanding how nitrate concentrations dilute over distance and time. The dispersion component is captured in the following equation:

$$\frac{\partial C}{\partial t} = D\nabla^2 C - v.\nabla C \tag{3.18}$$

Where D is the dispersion coefficient (m²/day), quantifying the degree of spreading, $\nabla^2 C$ represents the Laplacian operator, indicating how concentration varies spatially. The term $v.\nabla C$ accounts for advective transport as previously described. Dispersion in groundwater systems can be influenced by both molecular diffusion—resulting from the thermal motion of molecules—and mechanical dispersion, which occurs due to variations in flow speed

across different paths. In MT3DMS, dispersion coefficients are often estimated based on empirical relationships derived from laboratory tests or field data. For this study, the dispersion coefficients were obtained from a literature survey, and a value of 40 m was chosen dispersivity coefficient.

Incorporating the advection-dispersion equation in a GW flow system under generalized hydrogeologic conditions using:

$$\theta \frac{\partial C^k}{\partial t} + \rho_b \frac{\partial \bar{C}^k}{\partial t} = \frac{\partial}{\partial x_i}\left(\theta D_{ij} \frac{\partial C^k}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\, \theta v_i C^k + q_s C_s^k - q_s C^k \qquad (3.19)$$

Where

θ	is porosity or average volume of water content
C^k	is the dissolved concentration of species k in unit mass per volume
t	is time
ρ_b	is the bulk density of subsurface material
$\bar{C}^k$	is the concentration of species k, sorbed to the subsurface material
q_s	is the volumetric flow rate per unit volume representing sources or sinks
C_s^k	is the source or sink concentration of species k
∂	signifies the partial derivative of the variable that follows

This combined equation comprehensively explains how nitrate concentrations change over time due to advection, dispersion, and sink/ source mixing.

3.2.7.3 Streamflow depletion and nitrate loading

In many groundwater systems, particularly those affected by agricultural practices, streamflow depletion, and nitrate loading are interrelated processes. Streamflow depletion occurs when groundwater extraction reduces base flow contributions to nearby streams. As groundwater levels decline, historically gaining streams may shift to losing streams, causing surface water to seep into the aquifer through an unsaturated zone. This change can

significantly impact natural flow regimes, leading to decreased baseflow and more frequent low flow conditions due to increased surface water extraction for irrigation. This reduction can lead to increased concentrations of nitrates in groundwater as less water dilutes these contaminants. Studies have shown that these concentrations can fluctuate on various temporal scales—from daily to seasonal and annual cycles. Understanding these trends requires sophisticated statistical approaches to identify significant drivers while accounting for complex interactions and potential lags in response to environmental changes. Nitrate loading refers to introducing nitrates into groundwater systems through various means, such as fertilizer application or wastewater discharge. The interaction between streamflow depletion and nitrate loading can be modeled by incorporating boundary conditions that reflect these inputs and outputs within MT3DMS. For instance, when modeling a scenario where agricultural runoff contributes additional nitrates to a stream that feeds into an aquifer, one must adjust boundary conditions accordingly to reflect this dynamic relationship. Nitrate loading can be estimated using nitrogen budgeting techniques. For example, the amount of nitrate-N applied in one acre-foot of irrigation water can be calculated as:

$$Nitrate\ N\ (lbs) = Conc\ (ppm) \times 2.72 \times Total\ (acre) applied \quad (3.20)$$

This equation helps estimate how much nitrate-N contributes to groundwater contamination based on irrigation practices. The modeling framework allowed for a detailed representation of how these nitrates migrate toward nearby river systems, which is critical for understanding the impacts of agricultural runoff on water quality. MT3DMS employs a range of solution algorithms tailored to address the complexities of reactive transport, particularly for multi-species solute transport governed by kinetic rate equations.

The boundary conditions influence the flow and transport of nitrates in GW systems. This study took the Neumann condition boundary condition for the specified concentration

gradient modeling. The agriculture pattern varies in the study area within a year therefore, the solute concentrations change gradually. Transient concentration levels are defined based on the data available at the ADES portal (https://ades.eaufrance.fr/Recherche). The head-dependent boundary conditions were derived from the GW flow model itself. The model setup included defining boundary and initial conditions, hydraulic and transport parameters, and sources and sinks specific to the study area. The GW flow model derived initial conditions, aquifer parameters, and sources/sinks in the model. Critical parameters such as hydraulic conductivity, porosity, and dispersion coefficients are crucial for characterizing the porous media through which nitrate is transported. These parameters influence the speed and extent of nitrate movement and its interaction with the surrounding environment. The calibrated model showed an RMSE of 9.32 for nitrate concentrations, indicating good agreement between observed and simulated values (Fowler et al., 2008; Karterakis et al., 2007).

MODFLOW-MT3DMS coupling allows for an accurate representation of groundwater-surface water interactions, which are vital for understanding the overall hydrological cycle in the basin. The model's effectiveness was evaluated by comparing simulated nitrate concentrations with observed data, thus improving predictive capabilities for future scenarios regarding nitrate contamination and management strategies in the lower Ain basin.

3.2.8 Model Calibration

The Calibration of groundwater models is typically assessed using a combination of statistical metrics that quantify the goodness of fit between simulated and observed data (Schneider et al., 2022). RMSE measures the average magnitude of errors, giving more weight to large deviations, while MAE provides a more robust measure of the average absolute difference, less sensitive to outliers. RMSLE is helpful when dealing with right-

skewed target variables, as the log transform makes the distribution more normal. Median Absolute Error (MedAE) is a robust alternative to standard deviation, using the median of absolute differences instead of the mean. Finally, the coefficient of determination (R^2) indicates the proportion of variability in the target variable that can be explained by the model, with values closer to one indicating better fit. Several statistical metrics can be used to evaluate the performance of a groundwater (GW) calibration model. Below are the equations for Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Root Mean Square Logarithmic Error (RMSLE), Median Absolute Error (MedAE), and Coefficient of Determination (R^2).

RMSE

The RMSE measures the average magnitude of the errors between predicted and observed values, giving higher weight to larger errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - Y_i)^2} \tag{3.21}$$

where:

n = number of observations
y_i = observed value
Y_i = predicted value

MAE

MAE is the average of the absolute differences between predicted and observed values, providing a linear score that does not penalize large errors as heavily as RMSE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - Y_i| \tag{3.22}$$

RMSLE

RMSLE is useful when penalizing underestimations more than overestimations, particularly when the values vary widely.

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(y_i + 1) - \log(Y_i + 1))^2} \quad (3.23)$$

MedAE

MedAE provides the median of absolute errors, which is robust against outliers.

$$MedAE = median(|y_i - Y_i|) \quad (3.24)$$

R^2

R^2 quantifies how well the predicted values approximate the actual data points, with a value of 1 indicating perfect prediction.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{\sum_{i=1}^{n}(y_i - x_i)^2} \quad (3.25)$$

Where, x is the mean of observed values

These metrics were critical in assessing the performance of GW calibration models as they helped quantify how well model predictions aligned with observed groundwater data.

In the calibration processes, the groundwater head values computed by the model and observed at the observation wells were analyzed at a 95% confidence level at four separate locations. These metrics were calculated for four observation wells to quantify the discrepancy between the observed and simulated groundwater heads over the twenty-three stress periods. Furthermore, the differences between observed and computed values were found less in the middle part of the valley. In contrast, on the boundary of the model domain, they were not found within the 95% confidence level range due to inaccuracy in defining the head flux boundaries, as shown in Figure 3.7. Therefore, head flux boundary values were modified within the range of 10%, and the effect on the error was observed. The

constant flow boundary was initially calibrated in the steady state condition to incorporate the groundwater in-flow from the adjacent aquifer (Yeh & Mock 1996). The recharge rate was calibrated with the help of an inverse model called PEST (a popular parameter estimation program).

The mean RMSE of the models was 0.7928, indicating a reasonable fit between the observed and simulated data. Figure 3.8 shows that the RMSE and MAE are generally consistent, with minor variations reflecting localized model inaccuracies such as complex stratigraphy and irregular geometry. Such cells from the models were manually rectified while validating with official datasets. The R^2 values show more robust performance in Wells 1 and 2, while Wells 3 and 4 demonstrate the discrepancy due to complex anthropogenic stresses in these regions. Near well three, the land use typically comprises agricultural areas and households. The agrarian demand in near well four is higher than the other demands.

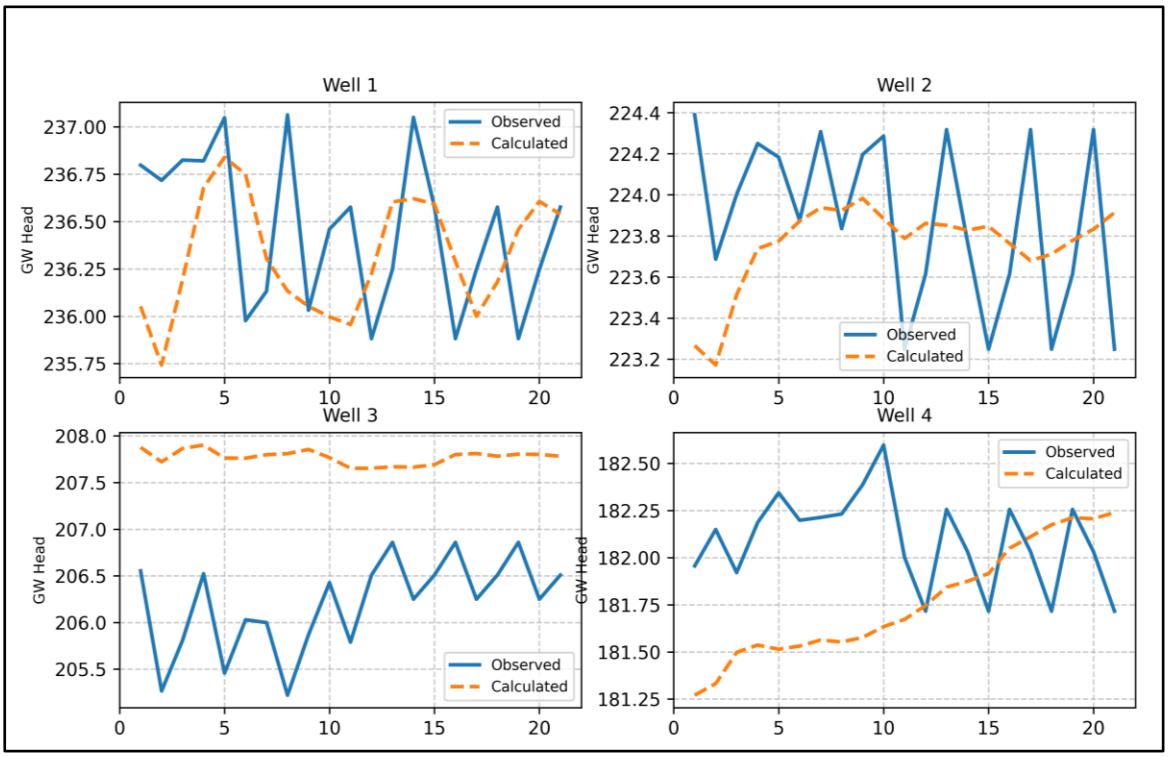

Figure 3.7 GW calibration at four observation wells present in the study area.

Figure 3.8 Different statistical measures were used in the comparative analysis of the error across four observation wells

The model's final outputs were GW flow contour maps and directions of groundwater movement, R-A exchange hotspots, and nitrate leaching into the river. The calibrated MODFLOW model simulates groundwater flow under future pumping and recharge scenarios.

3.2.9 Optimization Model

3.2.10 Multi-Objective Groundwater Simulation-Optimization

Integrating meta-heuristic evolutionary algorithms as an optimization technique has greatly expanded the capabilities of multi-objective groundwater simulation optimization (MO-GSOP). The main advantage of meta-heuristic EAs is their ability to optimize functions without relying on explicit knowledge of their characteristics. In contrast, traditional mathematical optimization methods are time-consuming and computationally demanding. A MO-GSOP can be mathematically represented as follows:

$$\min f_m(x), \quad m = 1,2,\dots,M \tag{3.26}$$

$$\text{Subject to } \boldsymbol{g_j(x) \geq 0}, \quad \boldsymbol{j = 1,2,\dots,J} \tag{3.27}$$

$$h_k(x) = 0, \quad k = 1,2,\dots,K \tag{3.28}$$

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \qquad i = 1,2,\dots,n \tag{3.29}$$

The objective function comprises M conflicting goals that need to be minimized. J inequality constraints and K equality constraints, if any, are introduced. There are upper and lower boundaries on the decision variables represented by $\boldsymbol{x_i}^{(L)}$ and $\boldsymbol{x_i}^{(U)}$ Respectively. The concept of dominance is crucial in selecting and identifying superior solutions. Below are some definitions that have context with multi-objective MO-GSOP:

Pareto Dominance: For any two solutions, a and b, in the obtained solution set, a dominates b if the solution f(a) is better than f(b) in at least one objective while not being worse than b in any objective.

Non-dominance: A non-dominated solution refers to a collection of solutions within a given set that any other member does not dominate. For any two solutions, a and b, a is considered non-dominated concerning b if neither a dominates b nor b dominates a.

Pareto front: The Pareto-optimal set represents the non-dominated set within the complete feasible decision space. The Pareto-optimal front is defined as the set of all points obtained by mapping the solutions from the Pareto-optimal set (i.e., PF = {f(x)|x ∈ PS}).

In multi-objective optimization problems, such as those encountered in groundwater simulation-optimization models, the objectives are often inherently conflicting, making it impossible to identify a single solution that optimizes all objectives simultaneously. Instead, the solution space comprises a Pareto optimal set—a collection of non-dominated solutions where any improvement in one objective can only be achieved at the expense of deteriorating at least one other objective. The primary aim of multi-objective optimization in groundwater management is to approximate the Pareto front—the boundary representing the optimal trade-offs between competing objectives. In this context, achieving a balance between convergence and diversity is critical. Good convergence implies that the solutions are as close as possible to the true Pareto front. This ensures the model captures realistic

trade-offs between objectives such as groundwater extraction, river-aquifer interactions, and water quality. Conversely, good diversity refers to the spread and distribution of the solutions across the objective space. This ensures that various alternative solutions are considered, representing different trade-offs across the objectives. This balance is crucial for effective groundwater management, as it provides decision-makers with a comprehensive set of strategies, ranging from maximizing groundwater extraction to preserving ecological functions and water quality, thereby supporting sustainable resource management.

3.2.11 Metaheuristic algorithms

While confronting complex, multi-objective challenges in GW management, traditional optimization algorithms frequently exhibit several limitations. Algorithms such as gradient descent and linear programming are fundamentally designed for single-objective, convex problems characterized by smooth and differentiable solution spaces. However, real-world water resource problems are often marked by non-linearity, discontinuity, and high dimensionality, which these traditional methods struggle to handle effectively. A key drawback of gradient-based algorithms is their propensity to become trapped in local optima, especially in non-convex solution spaces where the objective functions may involve complex interactions between variables such as recharge rates, extraction patterns, and aquifer responses. This issue is exacerbated because traditional methods typically assume a deterministic environment where all parameters are precisely known. Such an assumption is unrealistic in groundwater management, where significant uncertainty exists in parameters like future precipitation, aquifer characteristics, and demand projections.

Moreover, these algorithms often reduce multi-objective problems to a single composite objective through weighted sums, which can obscure the inherent trade-offs and fail to capture the complexity of decision-making in water resource management. Additionally,

the efficiency of traditional algorithms diminishes as the problem's dimensionality increases, which is a critical issue in groundwater optimization where numerous variables must be considered simultaneously across spatial and temporal scales. The high dimensionality, coupled with the non-convexity and uncertainty, often leads to computational inefficiency and suboptimal solutions. These limitations necessitate the use of more advanced optimization techniques. Metaheuristic algorithms navigate complex, high-dimensional spaces, avoid local optima, and handle multiple objectives and uncertainties, offering a more robust and adaptable approach to GW optimization and overcoming the significant shortcomings of traditional methods. This section explores several prominent metaheuristic algorithms used in this thesis—MOPSO, NSGA-II, MOEA/D, and Pareto Search—and discusses their application, hyperparameters, and effectiveness in optimizing water resource systems.

3.2.11.1 Multi-Objective Particle Swarm Optimization

MOPSO is an extension of the Particle Swarm Optimization (PSO) algorithm, tailored to handle multi-objective optimization problems (Coello Coello and Lechuga, 2002). PSO is inspired by the social behavior of birds flocking or fish schooling and is characterized by a population (swarm) of candidate solutions (particles) that move through the search space. The process begins with initializing a swarm of particles, where each particle represents a potential solution. Each particle has a position vector corresponding to the decision variables and a velocity vector determining its movement. Every particle keeps track of its personal best position (pBest). It is influenced by a global best (gBest) selected from an external archive of non-dominated solutions (Fallah-Mehdipour et al., 2011).

The fitness of each particle is evaluated using the defined objective functions. Non-dominated solutions from the swarm are stored in the archive, updated iteratively to ensure it contains only the best solutions. Each particle updates its velocity based on a combination

of its previous velocity, the distance to its pBest, and the distance to the gBest. Random factors are introduced in this update process to enhance exploration. The particle's position is then updated based on the new velocity, allowing the swarm to navigate towards promising regions of the search space. This iterative process continues until a termination criterion is met, such as achieving a predefined number of iterations or reaching convergence. The final output is a set of Pareto-optimal solutions representing trade-offs among the objectives. Integrating with simulation models such as MODFLOW, MOPSO can generate optimal strategies that balance socioeconomic and environmental needs, making it a powerful tool for managing groundwater resources (Rezaei and Safavi, 2020). MOPSO aims to find Pareto-optimal solutions, representing the trade-offs between conflicting objectives. Several hyperparameters are used in the algorithm, which are described below.

1. **Swarm size:** This parameter defines the number of particles in the swarm or solutions in each iteration. The swarm size directly influences the search space exploration's diversity and the algorithm's convergence behavior. A larger swarm size generally enhances exploration but may increase computational complexity.

2. Inertia weight: The inertia weight is a critical factor that governs the influence of the particle's previous velocity on its current velocity. By adjusting this parameter, one can control the balance between global exploration and local exploitation, affecting the algorithm's convergence rate and stability.

3. Cognitive coefficient (c1): This coefficient represents the weight assigned to the particle's personal best position. It reflects the particle's inclination towards its historical best solution, driving it to explore regions of the search space that have yielded favorable results.

4. Social coefficient (c2): The social coefficient indicates the weight assigned to the global best position found by the swarm. It represents the particle's tendency to move towards the best solution discovered by the entire swarm, promoting collaborative exploration of the search space.

5. Mutation rate: This parameter specifies the probability of introducing random alterations in the particles' positions or velocities. Mutation is employed to maintain diversity within the swarm and prevent premature convergence to suboptimal solutions, thereby enhancing the robustness and effectiveness of the optimization process.

6. Archive size: The total number of non-dominated solutions retained in the external archive throughout the optimization process. This archive is a repository for the Pareto-optimal solutions found during the optimization. The key functions of the archive are the facilitation of elitism, updating pareto solutions, and preserving diversity.

Together, these hyperparameters orchestrate the dynamic behavior of particles in the swarm, balancing exploration and exploitation to navigate the solution space and achieve optimal results effectively. MOPSO effectively explores large, complex search spaces and maintains diverse solutions. However, the algorithm may require careful tuning of hyperparameters to balance exploration and exploitation, and it can be computationally intensive, particularly as the number of objectives increases. The hyperparameters above were tuned based on the test functions of which exact solutions are known. The used parameters have been summarised in the table.

3.2.11.2 Non-Dominated Sorting Genetic Algorithm - II

NSGA-II is one of the most popular and widely used metaheuristic algorithms for solving multi-objective optimization problems (Deb et al., 2002). NSGA-II is an evolutionary algorithm that uses mechanisms inspired by natural selection to evolve a population of solutions towards the Pareto front. It improves upon its predecessor by introducing a fast,

non-dominated sorting approach, a crowding distance mechanism for maintaining diversity, and an elitism strategy to preserve the best solutions found during the search process. In NSGA-II, each solution is represented as a chromosome, and a population of these chromosomes undergoes selection, crossover, and mutation. The selection process involves sorting the population based on dominance relationships, where solutions are ranked according to the number of other solutions, they dominate (Kollat and Reed, 2005). The crowding distance is calculated to ensure the population remains diverse, avoiding premature convergence to a single solution.

In the NSGA-II, several hyperparameters are essential for guiding the evolutionary process and ensuring effective optimization:

1. Population size: This parameter specifies the number of individuals (or solutions) in the population at each generation. The population size influences the diversity of the solution set and the computational resources required. A larger population size can enhance the search capability but may also increase the computational burden.

2. Crossover probability: The crossover probability determines the likelihood that pairs of chromosomes (solutions) undergo crossover operations to exchange genetic material. This parameter influences the exploration of the search space by generating novel solutions by recombining existing ones. Higher crossover probabilities generally increase the potential for discovering novel solutions.

3. Mutation probability: The mutation probability indicates the likelihood of introducing random chromosome alterations. Mutation serves to maintain genetic diversity within the population and prevent premature convergence. A higher mutation probability increases the chances of exploring new regions of the search space but may also disrupt well-adapted solutions.

4. Number of generations: This parameter defines the number of iterations or evolutionary steps the algorithm will execute to evolve the population. The number of generations determines the duration of the optimization process and affects the convergence toward optimal or near-optimal solutions. More generations generally allow the algorithm to refine the solution set and approach the Pareto front more closely.

These hyperparameters collectively guide the evolution of the population in NSGA-II, balancing exploration and exploitation to solve multi-objective optimization problems effectively. NSGA-II is highly robust and capable of finding diverse Pareto-optimal solutions. Its elitism and crowding distance mechanisms ensure the algorithm converges to a high-quality solution without losing diversity. NSGA-II is less sensitive to the shape of the Pareto front, making it versatile across diverse types of multi-objective problems. However, the computational complexity of NSGA-II can be high, particularly with large populations or many objectives (Zheng et al., 2016).

Additionally, the algorithm's performance depends on the proper setting of genetic operators like crossover and mutation. In high-dimensional search spaces, NSGA-II struggles to find the optima. Nonetheless, it has been extensively used in water resources optimization, including designing water distribution and supply networks with GW use, optimization of multi-reservoir systems, and quality problems.

3.2.11.3 Multi-Objective Evolutionary Algorithm based on Decomposition

MOEA/D is a relatively recent addition to metaheuristic algorithms, offering a unique approach to multi-objective optimization (Zhang and Li, 2007). Instead of optimizing the objectives directly, MOEA/D decomposes a multi-objective problem into a set of scalar optimization subproblems. This decomposition is achieved using weight vectors representing different trade-offs between the objectives. Each subproblem optimizes a particular objective combination defined by its corresponding weight vector. By

decomposing the problem, MOEA/D effectively manages multi-objective optimization's complexity and computational burden. It allows for parallel search space exploration, enabling the algorithm to discover diverse solutions representing different trade-offs between the objectives.

One of the critical strengths of MOEA/D is its ability to share information among neighboring subproblems. This information exchange promotes both exploration and exploitation of the search space. The neighborhood structure in MOEA/D is defined based on the proximity of weight vectors. Subproblems with similar weight vectors are considered neighbors and are more likely to share information. Moreover, MOEA/D effectively balances global exploration and local search, which is crucial for finding diverse and high-quality solutions. Global exploration involves exploring different regions of the search space to discover new and promising solutions. Local search, on the other hand, focuses on refining existing solutions to improve their quality. Several hyperparameters are crucial for ensuring effective convergence and diversity of solutions in MOEA/D:

1. Number of subproblems: This parameter defines the number of scalar optimization problems into which the multi-objective problem is decomposed. Each subproblem focuses on a particular region of the objective space, and the number of subproblems influences the decomposition's granularity and the solution set's diversity. Many subproblems can lead to a more detailed exploration of the objective space but may increase computational complexity.

2. Crossover probability: The crossover probability specifies the likelihood that two solutions will be combined to produce offspring. This operator facilitates the exchange of information between solutions and promotes exploration of the search space. Higher crossover probabilities generally enhance the algorithm's ability to discover new and potentially better solutions.

3. Mutation probability: The mutation probability determines the likelihood of introducing random changes to a solution. Mutation maintains genetic diversity within the population and helps avoid local optima by exploring new regions of the search space. A higher mutation probability can lead to increased diversity but may also disrupt well-adapted solutions.

4. Neighbourhood size: This parameter defines the number of neighboring subproblems that share information with each subproblem. The neighborhood size affects the extent of cooperation and information exchange between subproblems, influencing the solution set's convergence behavior and diversity. A larger neighborhood allows for more extensive information sharing but may also increase computational demands.

MOEA/D is highly efficient and scalable, mainly when dealing with problems involving many objectives. Its decomposition strategy allows for a fine-grained search of the Pareto front, and the algorithm is well-suited to problems with many objectives where other algorithms might struggle (Chikumbo et al., 2012; Li et al., 2019). However, it requires careful setting of weight vectors, which can be challenging for problems with unknown or complex Pareto front shapes. The algorithm can also be sensitive to the neighborhood size parameter, influencing the balance between exploration and exploitation.

3.2.11.4 Pareto Search

Unlike other metaheuristic algorithms, Pareto Search prioritizes the exploration of trade-offs between objectives, focusing directly on locating the Pareto optimal set. The algorithm functions through an iterative process whereby non-dominated solutions are identified, and the Pareto front is updated as novel solutions are discovered (Yang et al., 2013). During the Pareto Search process, a population of candidate solutions is evaluated at each iteration. Solutions not dominated by any other within the population are incorporated into the Pareto front. This search continues until convergence, ensuring the final set of solutions provides

a comprehensive view of the trade-offs between competing objectives. The pseudo-code for the Pareto Search algorithm can be summarized as follows: initially, a population of random solutions is generated, and their objective values are evaluated. Non-dominated solutions are then identified to form the initial Pareto front. The algorithm proceeds iteratively, generating novel solutions through mutation or crossover, evaluating their objective values, and updating the Pareto front with newly identified non-dominated solutions. This process continues until a predefined termination criterion is met; at this point, the algorithm returns the Pareto front as the set of optimal solutions.

The hyperparameters for the Pareto Search algorithm include population size, which determines the number of solutions evaluated at each iteration; mutation rate, which represents the probability of introducing changes to solutions; crossover rate, which is the probability of combining solutions; and the convergence criterion, which defines the conditions under which the algorithm halts. The Pareto Search algorithm offers several advantages, including its direct focus on identifying the Pareto front, making it practical to visualize trade-offs in multi-objective problems. It is straightforward to implement and can be integrated with other algorithms to enhance search capabilities. However, its simplicity can also be a drawback, as it may not explore the search space as thoroughly as more complex algorithms. Moreover, it can be computationally expensive, particularly in high-dimensional or non-linear problem spaces. Recently, Pareto Search has seen applications in the water resources sector, where it has been utilized to balance water quality with supply and develop sustainable land-use strategies (Li et al., 2022). Its emphasis on generating clear trade-offs makes it particularly valuable in decision-making processes involving multiple stakeholders with diverse priorities.

3.2.11.5 Hyperparameter tuning

In this study, hyperparameter tuning was performed for the optimization algorithms using a Bayesian optimization framework to enhance their performance. The optimization process was applied to three key hyperparameters: population size (`pop`), the number of generations (`gen`), and crossover probability (`cross`). These parameters were represented as optimizable variables within defined ranges, with population size varying from 20 to 250 and the number of generations from 500 to 3000, both were logarithmically transformed to allow exploration across scales. The crossover probability was defined within a continuous range of 0.2 to 0.9. The objective function `main_ga, ` which encapsulates the performance metric of the NSGA-II, was used as the target for optimization. Bayesian optimization was executed for a four-hour runtime, incorporating stochasticity and running in a single-threaded mode. The best-performing configuration was identified by tracing the minimum value of the objective function across iterations, with the optimal parameters stored at the index corresponding to the minimum value. This approach ensured systematic and computationally efficient hyperparameter tuning, enhancing the robustness of the optimization algorithm for the problem at hand.

3.3 Linking groundwater model to optimization algorithms

3.3.1 PyGWMO module

Python-based Groundwater Many Objective Optimization (PyGWMO) is a Python-based high dimensional GW management module that handles the simulation data by MODFLOW through a .h5 file and optimizes for various management problems. Python was selected to develop this module due to its robust ecosystem of scientific libraries and its suitability for rapid prototyping. The module is designed to execute any MODFLOW model provided as input, optimizing the pumping rates of wells based on the specific optimization problem outlined. To ensure flexibility, the module allows users to choose

from various optimization algorithms and configure their parameters according to the needs of the problem. Additionally, the module incorporates different clustering methods for the strategic assignment of zones to the wells, enhancing the adaptability of the high dimensional optimization problems. A key objective in developing this tool was to make it accessible to users without Python programming expertise. Consequently, a user-friendly GUI was implemented, allowing non-programmers to utilize the full functionality of the module, thus bridging the gap between complex computational models and practical, user-driven GW management.

3.3.1.1 Structure of PyGWMO

The data management structure for the optimization process is organized into specific directories for efficient storage and retrieval of input and output files. Within the "*data*" folder, the MODFLOW model generated by the GMS software is stored. This directory also contains the outputs of the optimization algorithm, which are systematically placed in subfolders named "*optimizationx*," where "*x*" corresponds to the iteration number of the optimization process. Each optimization folder contains all pertinent input parameters in a **.json** file, while the output data is stored as **.csv** files, and the optimization history is recorded in a **.txt** file. These files are essential for further analysis and decision-making.

Additionally, an *"image"* folder is designated for storing visual outputs generated by the algorithm, including maps of the simulated domain, graphical representations of the different zones assigned to wells, and visualizations of the optimization process, such as cost evolution curves and Pareto fronts. The functional programming approach was chosen for the algorithm, as it generally offers greater efficiency than object-oriented programming. Functional programming emphasizes pure functions and avoids mutable states, leading to more predictable and often faster code execution. However, this approach is typically less intuitive and less easily human-readable than object-oriented programming,

which can pose code maintainability and understanding challenges. Despite this, the decision to use functional programming was driven by the need to optimize computational performance in handling the complex, multi-objective optimization tasks inherent in the algorithm. The structure of the software is given in Figure 3.9.

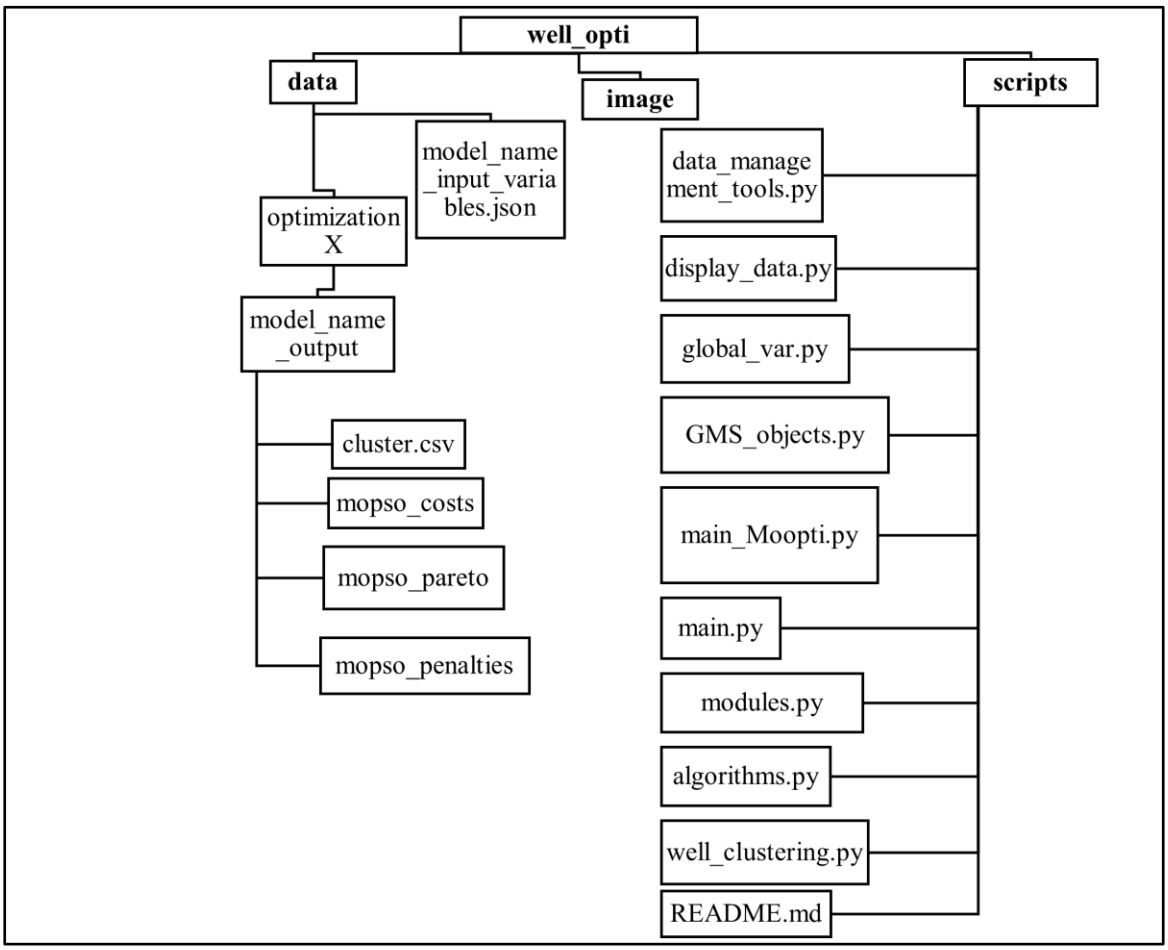

Figure 3.9 A schematic diagram of software structure and code dependency

The software structure is explained as follows:

```
├─── wells_opti          # Project folder
├─── data            # Folder containing results of optimizations
├─── image             # Image storing
├─── scripts           # Python scripts folder
```

The data folder is divided into the following subfolders:

```
├─── data                          # Data folder
├─── optimizationx                 # Folder containing results of optimization
├─── model_name_output       # Folder containing outputs of optimization
```

```
├──── model_name_input_variables.json  # Json file containing the input variables
```

The optimizationX folder contains the following information about the overall framework:

```
├──── optimizationX
  ├──── model_name_output
    ├──── cluster.csv    # Each row is a cluster, and it contains the cell ID of the Well
    ├──── mopso_costs-N_iter-N_pop_model_name.csv
    ├──── mopso_pareto_opt-N_iter-N_pop_model_name.csv
    ├──── mopso_penalties_opt-N_iter-N_pop_model_name.csv
    ├──── mopso_memory-N_iter-N_pop_model_name.txt
  ├──── model_name_input_variables.json
```

The scripts folder contains all the codes that are fundamental to the PyGWMO. It is structured as follows:

```
├──── scripts                    # Folder containing the Python scripts
  ├──── data_management_tools.py    # Reads the output data from GMS and computes the fitness function
  ├──── display_data.py      # Displays the output of the optimization and the map)
  ├──── global_var.py        # Sets up all the default variables (then modified by main.py)
  ├──── GMS_objects.py        # Contains the classes of objects from GMS
  ├──── main_MOopti.py         # Puts the input data in the optimization algorithms and runs
  ├──── main.py          # Display the API which will collect the input data
  ├──── modules.py          # Imports the packages
  ├──── moga_tools.py        # Multi-Objective Genetic Algorithm
  ├──── mopso_tools.py        # Multi-Objective Particle Swarm Optimization
  ├──── wells_clustering.py       # Creates the Area by performing a chosen clustering method
  ├──── README.md           # This file explains how to use the code
```

The *HDF5* file, named *`model_name.h5`,* is generated directly by the GMS and serves as a comprehensive repository for all simulation data. This file stores key response variables of interest, such as head values, well names, locations, and properties, as well as the geological attributes of each grid cell, which are encoded in either binary or floating-point format depending on the data type. MODFLOW interacts with this file by updating the head values

as simulations progress. Additionally, MODFLOW generates two essential files for the optimization process: the `*.ccf*` file, which contains aquifer leakage data, and the `*.drw*` file, which stores drawdown information. Both files are encoded in binary format, necessitating the development of custom functions to parse and interpret their contents accurately. These custom functions are designed to handle GMS-generated data, and MODFLOW is implemented in the `*wells_opti/scripts/data_management.py*` script. This section provides a detailed description of the various functions contained within the `*data_management.py*` script:

read_cells (): This function is responsible for reading data stored in an HDF5 file and generating GMS objects from the retrieved data. It serves as a crucial step in the initialization process by converting stored data into a format compatible with GMS.

read_wells (): Like `read_cells (), `this function specifically reads well data, such as discharge from the HDF5 file, and creates corresponding GMS objects. This function isolates the well data for further processing.

write_wells_property (array, areas): This function writes well properties into the HDF5 file. It assigns values from the input array to each well.

ccf_find_row_col(filecontent): This function identifies the occurrences of two known variables within a *.ccf* file. By locating these variables, the function deduces the structure of the remaining file content, enabling further interpretation and processing of the ccf data.

drw_find_values(filecontent): Analogous to `ccf_find_row_col(),` this function identifies the occurrences of drawdown of corresponding wells within a *.drw* file. From the positions of these variables, the encoding of the remainder of the file can be inferred.

read_modflow_array (), read_ccf(), and read_drw(): These functions are designed to decode byte data into floating-point numbers, facilitating the accurate interpretation of numerical data stored in various file formats used by MODFLOW.

Penalty (): This function computes the total penalty associated with the current model configuration, specifically evaluating the impact of the applied discharge rates on the wells. The penalty serves as a constraint measure within the optimization process.

mo_get_cost (particle, penalty, areas): This function calculates the cost of an individual particle based on the defined optimization problem. It incorporates the penalty and the spatial distribution of areas, running MODFLOW to evaluate the particle's performance within the optimization framework.

3.3.1.2 Software testing

A suite of numerical simulations was conducted to ensure the proper operation and performance evaluation of the PyGWMO software, as shown in Figure 3.10. These simulations employed well-established test functions commonly used in multi-objective

optimization to assess the efficiency and effectiveness metaheuristic algorithms.

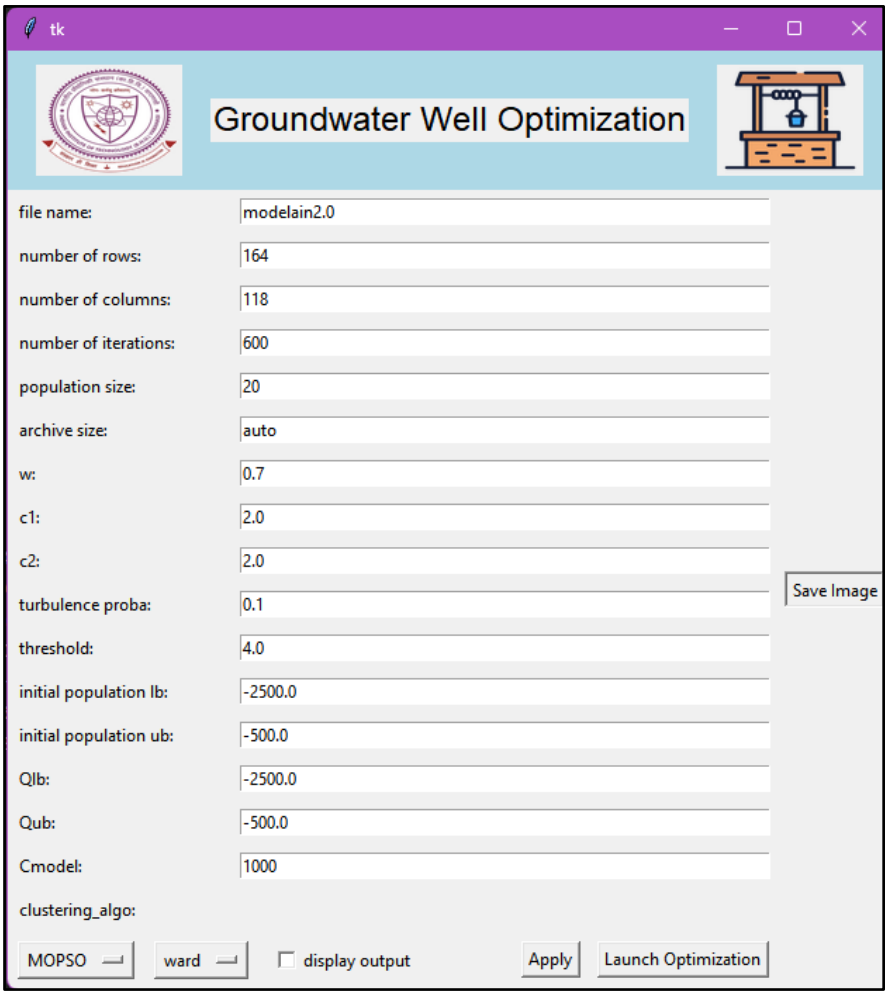

Figure 3.10 PyGWMO graphical user interface

The test functions utilized in this study included the ZDT-II and Ackley functions, among others. The functions were evaluated by changing the objective functions and decision variables. The robustness and accuracy of the implemented metaheuristics were thoroughly examined by evaluating the performance of the PyGWMO software on these standard test functions. All simulations were carried out on a high-performance computing workstation maintained by the Advanced Computing Laboratory, Department of Civil Engineering, IIT(BHU). The workstation was equipped with an AMD Ryzen Threadripper PRO 3975WX processor, featuring thirty-two cores with a clock speed of 3.50 GHz and 128 GB

of installed RAM. The GPU utilized in these experiments was the NVIDIA Quadro RTX 4000, a powerful graphics processing unit designed for professional applications.

3.4 Results

The GW head varies from 162.5 m to 273.3 m, with an average value of 240 m. The model domain's northern and western parts show relatively high GW heads, while the southern region substantially declines. The high GW head could be due to the high initial GW level in the area. However, the region has a significant drawdown with an average of 6 m within 7 years. This may be due to high extraction for agriculture and GW upwellings at several zones for maintaining river flow during summers. GW direction follows a natural path where the water from the higher head flows to the south of the region with the river as the major source/sink in the middle. Figure 3.11 illustrates the spatial distribution of the flow field and nitrate field at the end of the simulation period.

Regarding Nitrate advection and dispersion, the agricultural zones in the northern boundary show higher concentrations and disperse quickly towards the river where there are already hotspots in gaining river systems, where groundwater discharges into rivers, nitrate-laden aquifers can elevate nitrate concentrations in surface water, leading to eutrophication and associated ecological issues like algal blooms and oxygen depletion. Conversely, nitrate-rich river water infiltrates into aquifers in losing river systems, exacerbating groundwater quality degradation. In the middle of the simulation, after 3.5 years, the top agricultural zone nitrate leaches to the river cells with an average concentration of 55mg/l, while the second nitrate source disperses up to a few cells only. The third zone behaves similarly to the first one, reaching the southern boundary which touches river Rhone. The dispersion is larger than all the wells here because the travel path is bigger, contributing to larger areas of contamination. The last zone advects and disperses into river Ain with an average

concentration of 8 mg/l. At the end of the simulation all wells contribute to the nitrate load into the river cells.

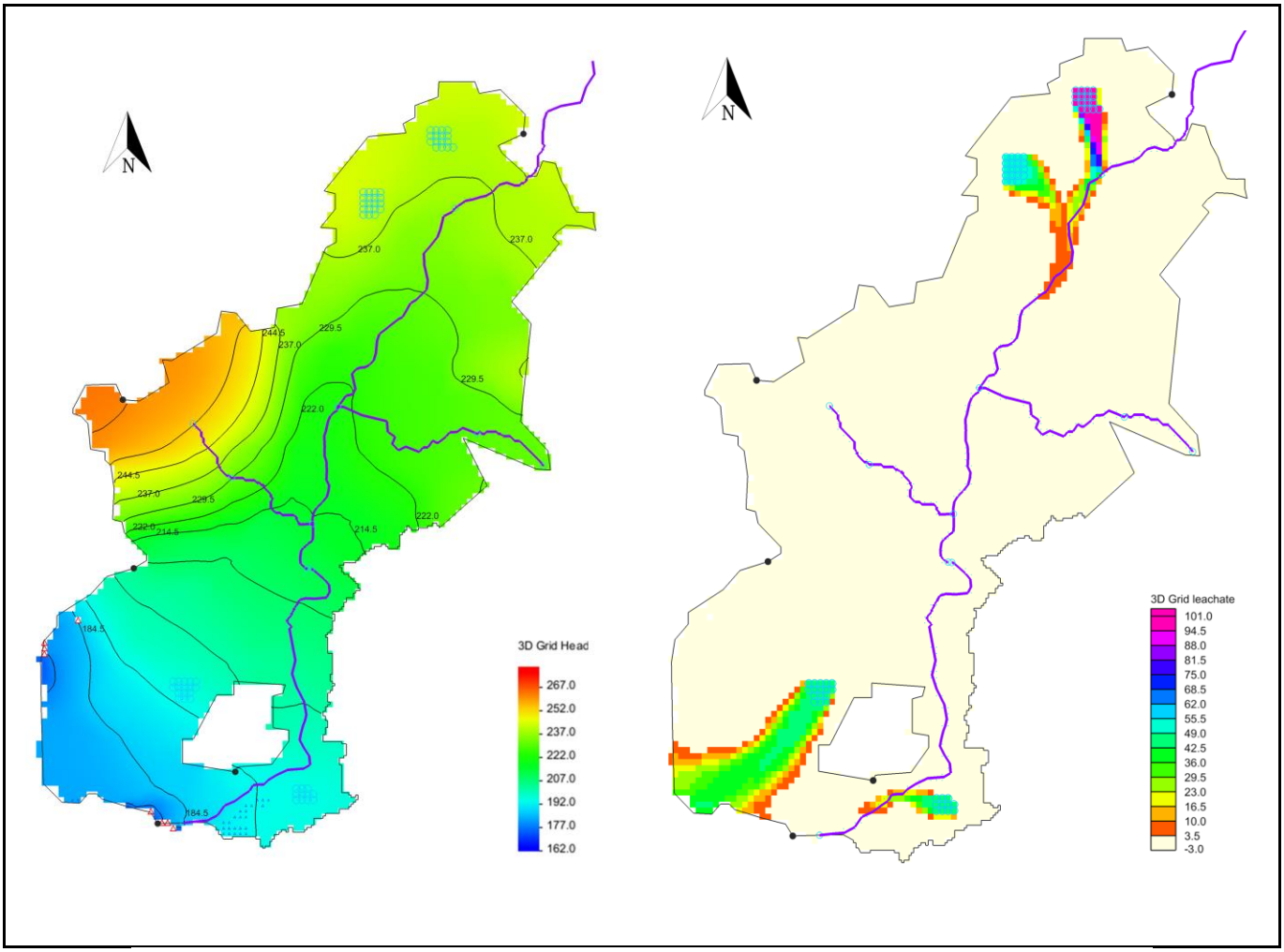

Figure 3.11 Groundwater head contours and Nitrate dispersion at the end of the simulation period

3.4.1 Flow budget and R-A Exchanges hotspots

Scattered sites of upwellings and downwelling were detected along the entire 40 km stretch of the Ain River, extending from Pont-d'Ain to Chazey-sur-Ain, revealing a nuanced spatial distribution characterized by intricate exfiltration and infiltration zones exhibiting significant temporal variability. These observations delineate areas of predominant river infiltration and groundwater exfiltration alongside 'mixed' zones where these processes

achieve approximate equilibrium, particularly during the summer. Notably, three distinct reaches were marked, each manifesting varying degrees of groundwater influx into the stream. The stream leakage varies from 7258.76 m^3 to -13,023 $m^{3,}$ indicating that the flow from the aquifer to stream is significantly higher than from stream to aquifer. Although aquifer recharge through stream leakage, it is very low compared to the other way around. The flow budget with a 90% interval plot in Figure 3.12 represents the variability and uncertainty associated with the simulated values over the considered time steps. Groundwater storage fluctuates significantly during the initial time steps but remains positive. However, a sharp decline in storage is observed after time step 20, indicating substantial groundwater depletion. The 90% confidence interval around the curve of the well demonstrates moderate variability across time steps, particularly between 5–10 and around 15, likely due to variations in pumping rates and their impacts on groundwater flow dynamics. River-aquifer exchange values exhibit relatively high fluctuations, with consistent negative flow values throughout, suggesting that groundwater predominantly discharges into the river system. The 90% confidence interval for stream leakage is wider at several time steps (e.g., around 10 and 20), highlighting variability in river-aquifer interactions, which may result from spatial differences in riverbed conductance or changes in groundwater gradients.

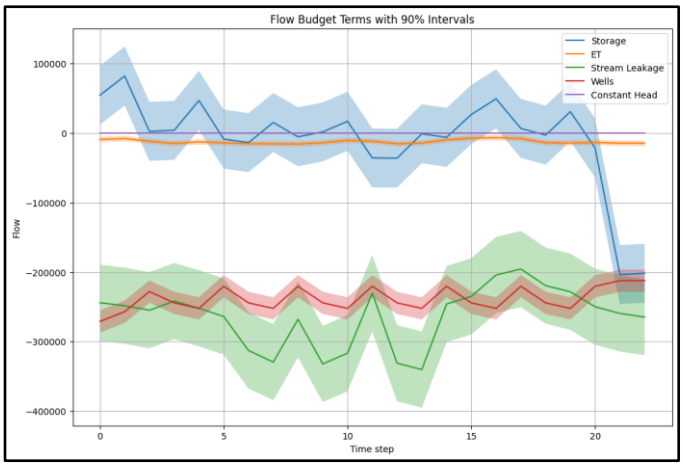

Figure 3.12 Flow Budget of lower Ain River basin. The colors represent different components of the flow budget, while the colored band shows the 90% confidence interval.

The spatiotemporal variation of R-A exchange is seen along the river in Figure 3.13. Three-time step values are shown in the plot at the start of the simulation model, at the middle, and at the end. The distribution of the three datasets is shown in Figure 3.14. Each distribution reveals a spread of values, with a central concentration near zero and extended tails indicating significant negative exchanges, suggesting groundwater outflow into the river system. The similarity in shapes across the three metrics highlights consistent patterns of R-A exchanges with minor variations in spread and distribution.

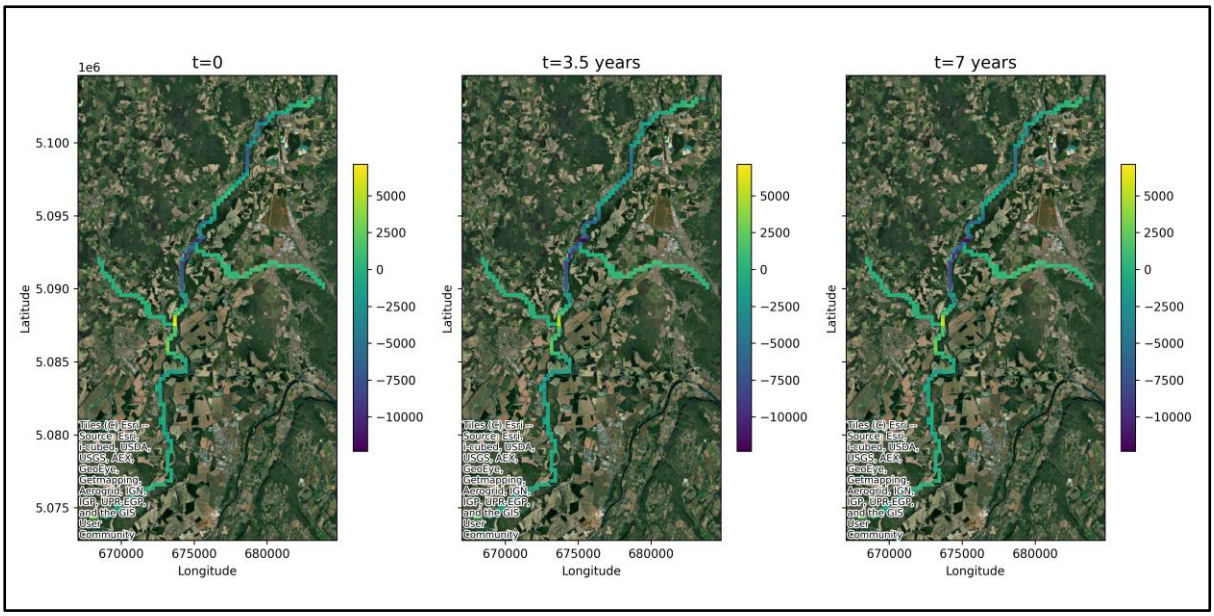

Figure 3.13 Variation of R-A exchange along the stretch of the river Ain at start, mid and end of simulation.

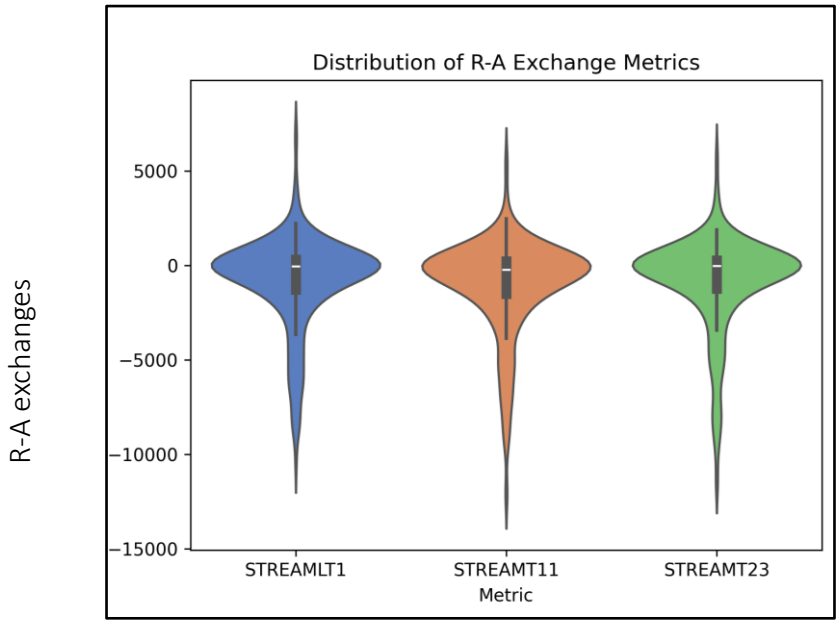

Figure 3.14 Distribution of R-A exchange metrics at the simulation's start, mid, and end.

The three zones of prominent exchanges are seen in Figure 3.15. The initial two reaches feature extensive groundwater upwellings, with peak values reaching up to 9767 m³, attributable partly to inflows from tributaries like the Albarine River. These areas are notably inhabited by thriving populations of brown trout and common grayling, sustained by contributions from numerous small, cold side channels facilitating lateral groundwater

accumulation. The first zone, located in the upper stretch of the river, serves as a hotspot for groundwater upwelling. GW input into the river in this region ranges between -2600 m^3 to -6500 m^3, indicating a steady contribution of GW to river flow. The uniform temporal variation suggests a clear seasonal pattern, where upwelling persists throughout different periods, supporting river flows during low-flow seasons. In the second zone, situated in the central portion of the river, groundwater upwelling is notably more pronounced, with inputs reaching up to -13,000 m^3. In the second reach, groundwater upwelling peaks approximately 1216 days into the simulation period, often associated with localized hyporheic upwellings downstream of gravel bars. However, due to grid size limitations, finer-scale patches within the river remain undistinguished. Some upwellings are also situated at the downstream terminus of the Ain River before its confluence with the Rhone River, exhibiting several discrete patches of upwelling activity, particularly notable during summer cycles.

Conversely, regarding downwelling or leakage from the river to the aquifer, only the third segment of the river demonstrates elevated values, reflective of significant groundwater losses. The third zone represents a downwelling region, where stream water leaks into the aquifer, facilitating groundwater recharge. This zone is characterized by significant stream leakage, with values reaching up to 7800 m^3 into the aquifer. The presence of this downwelling zone is critical for aquifer replenishment and highlights the river's role as a recharge source in this segment. Notably, the Albarine River behaves as a losing stream throughout most of its course, with exceptions noted where it intersects with the Ain River.

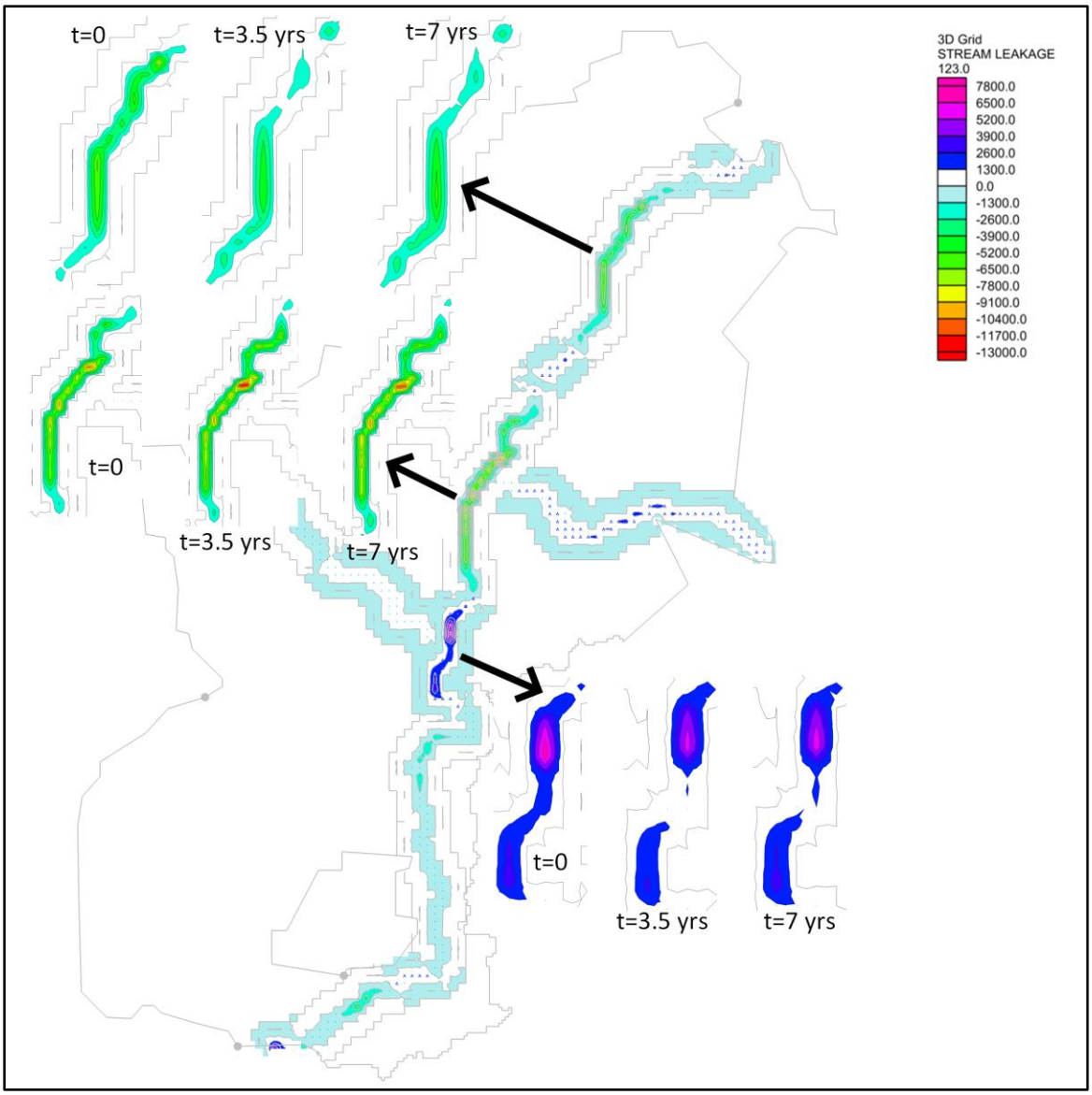

Figure 3.15 Three zones of R-A exchanges and their temporal variation at the simulation's start, mid, and end

3.4.2 Hydrogeological Factors Influencing Head and R-A Exchanges

The Sobol' I plot in Figure 3.16 and Figure 3.17 show first-order sensitivity indices for various input aquifer parameters obtained at three spatially distributed points (head1, head2, and head3) and R-A exchange parameters (LeakageIn and Leakage out) in all three model boundary conditions. Hydraulic conductivity across *head 1* to *head 3* shows high sensitivity across all models, with Model 1 and Model 2 showing similar high contributions, whereas Model 3 has slightly less influence. Specific yield shows varying sensitivity across different outputs. For instance, *Head 1* to *Head 3*: Model 1 shows a relatively higher

sensitivity, indicating that Specific Yield plays a significant role in these models. *LeakageIn* and *Out*: Sensitivity decreases, suggesting that Specific yield has a lesser impact on these outputs, especially in Model 3. Conductance exhibits moderate sensitivity in *Head* outputs, with Model 2 being more sensitive than others and very low sensitivity in *LeakageIn and Out,* which indicates conductance may not be a crucial factor affecting R-A in these models.

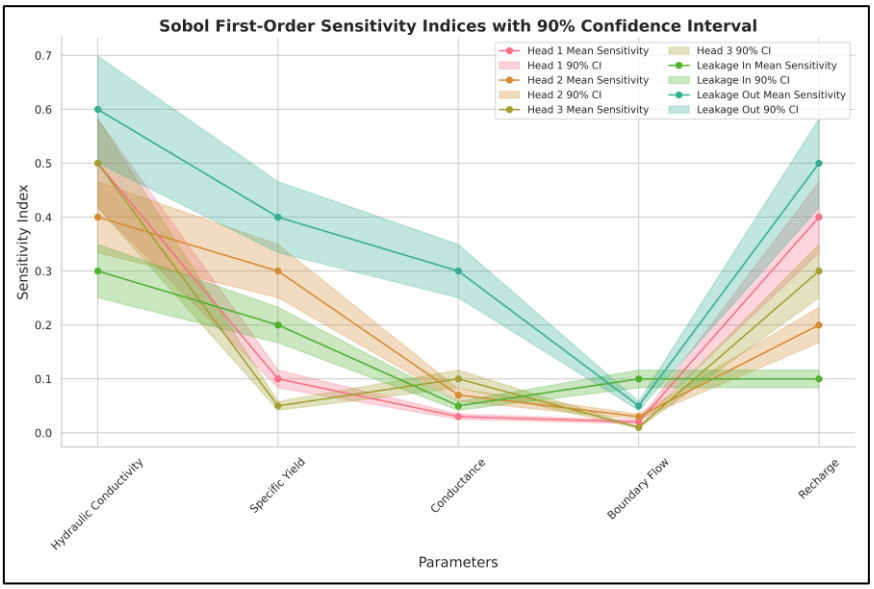

Figure 3.16 First-order sensitivity indices with 90% Confidence Interval

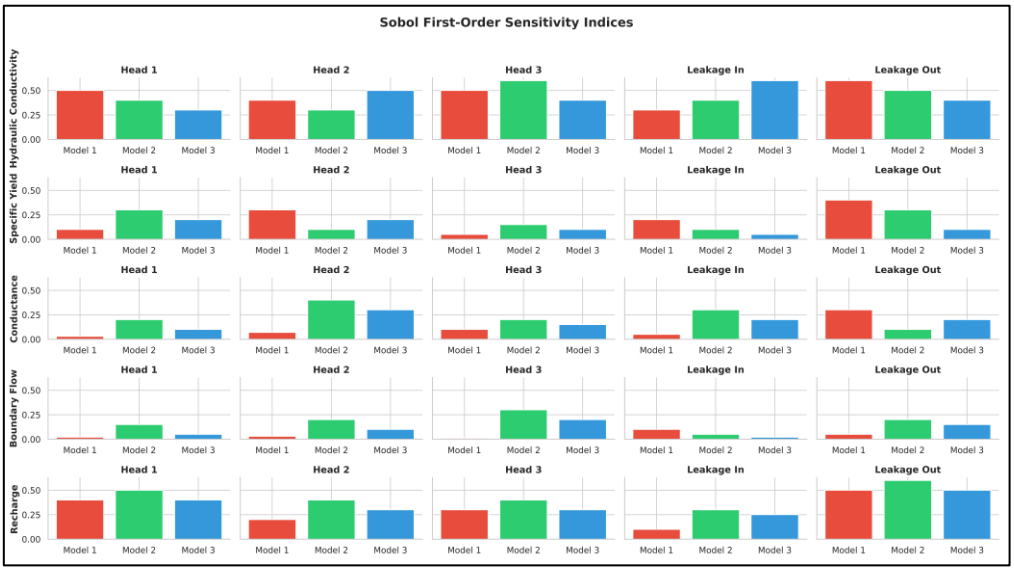

Figure 3.17 Sobol first-order sensitivity indices for various input aquifer parameters across three different model domains

In boundary flow parameter, the sensitivity is generally low, with Model 1 showing slightly higher sensitivity for *Head 1* and *LeakageIn*, suggesting a limited influence on these parameters. Recharge in the *Head 1* shows high sensitivity across all models, especially in Model 1. The sensitivity decreases for *Head 2* and *Head 3*, showing that recharge has a lesser impact on these outputs. Similarly, the recharge sensitivity is much lower for LeakageIn and Out, with minimal influence in all models, particularly Model 3.

Comparative Analysis Across Models shows that Model 1 (Red) generally shows higher sensitivity across most parameters, indicating that it might be more reactive to changes in input parameters compared to Model 2 (Green) and Model 3 (Blue). Models 2 and 3 show similar sensitivity patterns for most parameters but differ in magnitude. For instance, Model 2 exhibits more sensitivity to conductance than Model 3. This variability suggests that while all three models respond similarly to input changes, the magnitude of influence differs, possibly due to these models' structural or calibration differences. Hydraulic conductivity is a dominant parameter influencing most model outputs, highlighting its critical role in the system's hydrodynamics across all models. Specific yield and recharge are also influential but vary in degrees of impact depending on the model and output variable, indicating their secondary but significant roles. Whereas, boundary flow and conductance have lower sensitivity indices, suggesting that their influence on the model outputs is minimal, potentially allowing for less stringent calibration in those parameters without greatly affecting model performance.

The Sobol T-index plot presented in Figure 3.18 highlights the relative contribution of key parameters—hydraulic conductivity, specific yield, conductance, boundary flow, and recharge—across multiple groundwater models. Specifically, hydraulic conductivity and recharge exhibit significant sensitivity for Heads 1, 2, and 3, underscoring their dominant influence on groundwater levels. Similarly, LeakageIn and out show high sensitivity to

boundary flow and conductance, indicating the critical role of these parameters in controlling exchange fluxes.

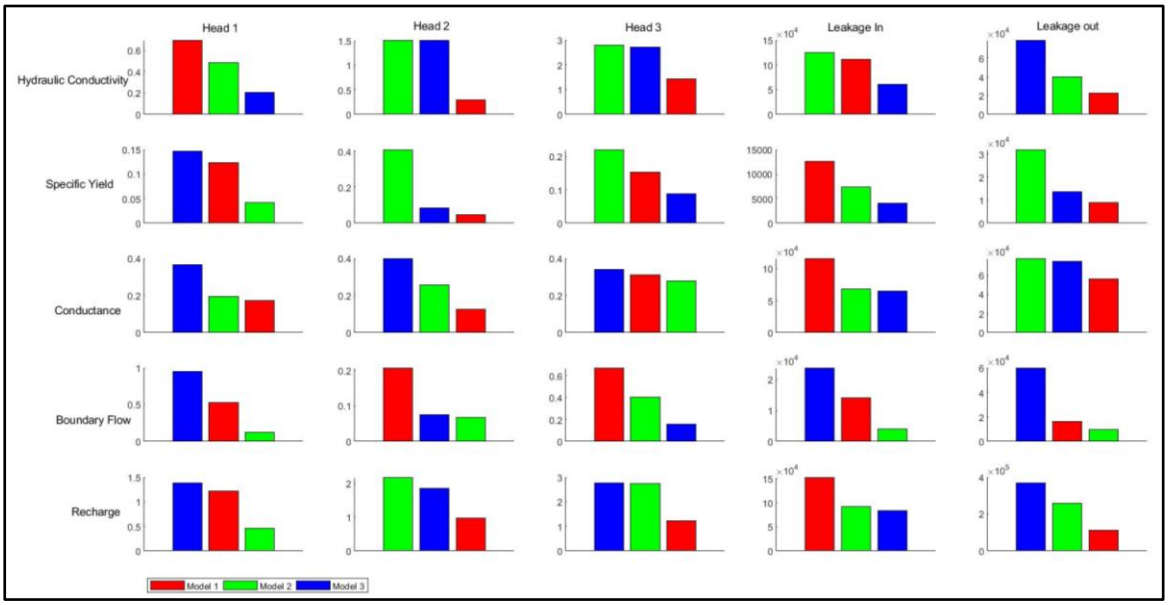

Figure 3.18 Sobol T indices for various input aquifer parameters across three model domains.

3.5 Summary

The chapter describes the study area, the LARB, covering its hydrogeological settings, including aquifer properties, river characteristics, and anthropogenic influences. The LARB is characterized by complex hydrogeological conditions, including highly permeable graveled riverbeds and significant human impacts. Further, this chapter comprehensively discussed the numerical models, and the S-O framework employed in this study. The numerical groundwater flow and nitrate transport models were rigorously calibrated and validated, ensuring their reliability for S-O integration. Integrating numerical models with optimization algorithms in the Python platform demonstrated the versatility and robustness of the S-O framework developed in this study. The chapter also emphasizes the importance of metaheuristic algorithms in addressing groundwater management challenges, particularly in handling multi-objective optimization problems with complex, non-linear relationships. This chapter's framework and methodology lay the foundation for

the subsequent analysis and discussions, offering contributions to groundwater management and river-aquifer exchanges.

The integration of numerical models, specifically for groundwater flow and nitrate transport, with advanced optimization algorithms within the Python platform demonstrated the effectiveness and adaptability of the S-O framework. This integration, facilitated by the development of the PyGWMO module, provides a powerful tool to evaluate various groundwater management strategies and understand the trade-offs between competing objectives. This module bridges the gap between complex computational models and practical groundwater management by providing a user-friendly interface and a flexible framework for optimization. The PyGWMO module integrates various optimization algorithms, including MOPSO, NSGA-II, MOEA/D, and Pareto Search, allowing users to select the most suitable algorithm for their specific optimization problem.