

Article

Optimal Identification of Concealed Copper Mineral Deposit Locations Utilising Isotope Signatures in Groundwater Systems

Ronald Maharaj¹, Ioan Sanislav²  and Bithin Datta^{1,2,*} ¹ Discipline of Civil Engineering, College of Science & Engineering, James Cook University, Townsville, QLD 4811, Australia² Economic Geology Research Centre (EGRU), Earth and Environmental Science, James Cook University, Townsville, QLD 4811, Australia

* Correspondence: bithin.datta@jcu.edu.au

Abstract

This study presents a synthetic proof-of-concept surrogate-assisted simulation–optimisation framework for identifying potentially mineralised locations using groundwater copper-isotope observations. The methodology adapts established groundwater contaminant-source characterisation principles, in which unknown source release histories are reconstructed from down gradient concentration measurements, to the investigation of concealed copper mineralisation. An isotope-informed MODFLOW–MT3DMS model was used to generate synthetic groundwater concentrations of Cu_{Tot} (total copper), ^{63}Cu , and ^{65}Cu for hypothetical mineralisation scenarios. Model responses were evaluated at 20 observation wells over five stress periods. These synthetic data were used to train, validate, and test a Gaussian Process Regression (GPR) surrogate model. The GPR model replicates the numerical simulator responses during testing. The test results show R^2 values exceeding 0.9998 and corresponding RMSE and MAE values of 0.03022 and 0.00571, respectively, indicating high predictive accuracy within the simulated parameter space. The trained surrogate was subsequently coupled with an Adaptive Simulated Annealing (ASA) algorithm to solve the inverse source-characterisation problem. The optimisation formulation contained 20 decision variables, representing time-varying source fluxes at four predefined candidate locations over five stress periods. Three candidates represented active synthetic mineral sources, whereas the fourth was intentionally assigned as an inactive dummy location to evaluate whether the framework could distinguish active from non-mineralised candidates. The optimisation converged in approximately 1080 s under the reported computational configuration. The reconstructed flux histories identified the three active candidates while assigning a near-zero flux to the dummy candidate. These results demonstrate that, under controlled synthetic conditions, isotope-informed inverse modelling can discriminate between predefined active and inactive candidate locations and reconstruct their temporal source strengths. However, the results do not constitute validation of a field-ready mineral-exploration method, because both the training and evaluation data were generated using the same underlying numerical framework. Subject to validation with independent field observations and more realistic geological, hydrogeochemical, and uncertainty representations, the proposed methodology could provide a computational screening and hypothesis-testing tool for prioritising prospective mineralised zones for further investigation.



Academic Editor: Micòl Mastrocicco

Received: 19 August 2026

Revised: 30 September 2026

Accepted: 1 October 2026

Published: 4 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)[Attribution \(CC BY\) license](#).

Keywords: inverse modelling; copper deposit identification; contaminant source detection in groundwater; copper isotopes; surrogate modelling; machine learning; groundwater systems; linked simulation optimisation; decision model

1. Introduction

The exploration of concealed or undercover mineral deposits remains challenging because mineralised bodies may have limited or no direct surface expression. Conventional geological, geophysical, and geochemical exploration methods can therefore encounter difficulties in identifying prospective mineralised zones beneath transported cover or complex geological formations. Groundwater offers a potentially useful exploration medium because its interaction with mineralised rocks can mobilise dissolved metals and transport their geochemical signatures beyond the immediate mineralised zone. Consequently, spatial and temporal variations in groundwater chemistry may provide indirect evidence for locating concealed mineralisation [1,2]. Copper occurs naturally as two stable isotopes, ^{63}Cu and ^{65}Cu . Variations in their concentrations and relative abundance can develop through mineral dissolution, sorption, precipitation, and other water–rock interaction processes. Copper concentrations and isotope ratios may therefore provide useful indicators of copper sources and mineral–water interactions. In the present study, the candidate source zones represent concealed Cu-bearing mineralised bodies capable of releasing dissolved copper through groundwater–mineral interactions. The simulated concentrations of ^{63}Cu and ^{65}Cu represent the isotope-specific copper mass entering the aquifer from these mineralised zones. The isotope representation adopted in this proof-of-concept study is deliberately simplified. The two copper isotopes are represented as separate transported species and assigned different adsorption parameters under nonlinear Langmuir sorption. This representation produced differential retardation and corresponding spatial and temporal variations in the $^{65}\text{Cu}/^{63}\text{Cu}$ concentration ratio. The resulting isotope-specific responses at the monitoring wells were evaluated as potential indicators of the contributions from different candidate source locations. This numerical representation provided a controlled mechanism for investigating whether isotope-dependent transport responses contain useful information for inverse source discrimination. It is not intended to reproduce the full range of natural copper-isotope fractionation mechanisms operating in field hydrogeochemical systems.

The isotope fractionation was assumed to have already occurred at the potential deposit locations and the transport in space and time was modelled from the deposit location. The isotope signatures were specified as absent initially, outside the potential deposit locations. At the deposit locations, the isotopes were assumed to be already fractionated, and the transport occurred from the source location. The original mass of the total copper in the deposit locations was specified and it served as the boundary condition for the further transport process.

The proposed inverse problem was conceptually adapted from groundwater contaminant-source characterisation. In conventional source-identification studies, unknown source locations, release magnitudes, and release histories were estimated from concentration measurements collected at groundwater monitoring wells [3–6]. Such problems are inherently ill-posed because monitoring observations are limited, hydrogeological properties are uncertain, and different source configurations may produce similar concentration responses [4,6,7]. In the present application, this inverse-modelling principle was transferred to mineral exploration; groundwater copper-isotope responses were used to evaluate the problem, which predefined candidate locations with the simulated observations to recon-

struct their time-varying source contributions. The intention was not to classify mineralised zones as contaminant sources but to adapt the mathematical source-characterisation framework to a different hydrogeological identification problem. Source-flux reconstruction and deposit-location identification were related, but distinct components of the proposed framework were evaluated. Source-flux reconstruction estimates the magnitude and temporal variation of copper released from each predefined candidate location, which can be expressed as a vector.

Deposit-location identification is subsequently inferred from the reconstructed fluxes. A candidate location that consistently requires a non-zero flux to reproduce the monitoring-well observations was interpreted as potentially active, whereas a candidate assigned a negligible or near-zero flux was classified as inactive under the assumptions of the model. Therefore, the present framework discriminates among predefined candidate locations; it does not directly estimate previously unknown spatial coordinates. Simulation–optimisation approaches have been extensively applied to groundwater source-characterisation problems by coupling numerical groundwater flow and transport models with optimisation algorithms [3,5,8]. Although numerical simulators provide physically based representations of groundwater systems, an optimisation procedure may require thousands of forward-model evaluations. Repeated execution of a detailed groundwater simulator can consequently impose substantial computational demands, particularly for transient, multilayer, and multi-species problems with numerous decision variables [9,10].

Surrogate models address this computational limitation by approximating the relationship between numerical-model inputs and simulated responses. They are trained using datasets produced by the original numerical simulator and subsequently provide rapid predictions during the optimisation process [11]. Gaussian Process Regression (GPR) is a flexible surrogate-modelling metamodel capable of representing nonlinear input–output relationships while providing high predictive accuracy within the sampled parameter space [12,13]. Trained surrogate models can therefore replace repeated numerical-model evaluations during optimisation and enable more computationally efficient reconstruction of unknown source histories [14].

Previous groundwater studies have coupled surrogate models with optimisation techniques such as Generalised Simulated Annealing (SA), Particle Swarm Optimisation (PSO), Genetic Algorithms (GAs), and other evolutionary algorithms to identify unknown contaminant sources in complex aquifer systems [15–17]. These studies have demonstrated that trained surrogate models can facilitate the identification of source locations and release histories while substantially reducing the computational burden associated with repeated numerical simulations.

However, the existing applications have primarily focused on groundwater contamination. Their adaptation to concealed mineral exploration using the separate groundwater transport responses of stable copper isotopes remains insufficiently investigated. In parallel with advances in inverse and surrogate modelling, isotopic and geochemical tracers have increasingly been used to interpret mineral–water interactions, transport processes, and possible source contributions within groundwater systems [1,16]. Isotope-specific observations may provide information beyond that obtained from total concentration measurements alone, particularly when the isotopes undergo different transport or retention behaviour. Nevertheless, limited research has examined whether such isotope-dependent responses can be incorporated into a simulation–optimisation framework to discriminate among candidate concealed mineralised locations.

The principal research gap addressed in this study is therefore the absence of an integrated inverse framework that evaluates whether isotope-specific groundwater transport responses can help distinguish between candidate concealed mineralised locations.

The novelty does not arise from the independent use of GPR or Adaptive Simulated Annealing (ASA). Rather, it arises from their integration within a framework that (i) adapts groundwater source-characterisation principles to concealed mineral exploration, (ii) represents ^{63}Cu and ^{65}Cu as separate transported species, (iii) incorporates simplified isotope-dependent sorption and retardation, (iv) integrates isotope-informed MODFLOW–MT3DMS simulation, GPR surrogate modelling and ASA-based inverse optimisation, and (v) evaluates whether reconstructed source fluxes can discriminate active mineralised sources from an intentionally inactive candidate location. Accordingly, the objective of this study is to develop and evaluate a synthetic proof-of-concept with a linked surrogate and the optimiser (GPR–ASA) framework for reconstructing time-varying copper source fluxes at predefined candidate mineralised locations.

First, a numerical groundwater flow and isotope-transport model is used to generate synthetic forward-simulation data. Second, a Gaussian Process Regression (GPR) surrogate model is trained, validated, and tested using the simulated responses. Third, the trained surrogate is coupled with the ASA algorithm to estimate the source fluxes over multiple stress periods. Finally, the reconstructed flux histories are used to discriminate between three active synthetic mineralised locations and one inactive dummy candidate.

No field measurements were used in this proof-of-concept study. In a future field application, real data would include site-specific geological and hydrogeological parameters used to construct and calibrate the numerical model, together with copper-isotope concentrations measured at observation wells. Any such validation of the proof of concept for any new methodology needs synthetic data to replicate different scenarios, especially when groundwater is the medium of transport. Different scenarios will need years and decades after implementation to obtain any useful field data. Therefore, the new methodology is evaluated as a proof of concept using synthetic data replicating real-life situations. These synthetic data are produced by solving mathematical simulation models for specified input data and scenarios. For the purpose of methodology evaluation, the term “scenario” in the context of the simulation model refers to the plausible deposit locations that may represent active, inactive or a combination of both. These scenarios produce the simulated isotope signatures. In the context of the optimisation-based decision model, the scenarios are the very large number of candidate solutions in the iterative search for the optimal solution. These scenarios are utilised in the simulation model to iteratively compute resulting isotope signatures through the linked simulation model in order to identify the optimal solution for the defined objective function.

Simulated data represents the solution results obtained from the simulation models for specified scenarios, which include parameter values, boundary conditions, source and sink terms, if any, etc. The simulation generates the isotope signatures at different locations and times based on assumed plausible deposit locations. The iterations of the optimisation model generate the plausible candidate source locations to be evaluated based on the objective function for minimising the difference between the simulated and the measured isotope signatures. Therefore, the best match is accepted as the optimal solution representing the source deposits. Also, fractal anomalies for efficient delineation of contaminant plumes in groundwater systems were proposed by [18,19]. A few other researchers have also attempted to utilise data mining algorithms, including machine learning (ML) and deep learning (DL), to better estimate the boundaries for mineral prospecting [20,21]. Fractal anomaly techniques were also utilised for the delineation of the plausible copper deposits periphery by type of research, particularly in the areas which are mineral exploration; however, the limitation of this research was the non-use of the isotope signature concentrations [22]. The study evaluates the framework under controlled synthetic conditions and provides a methodological foundation for future investigation using in-

dependent field observations, more comprehensive isotope-fractionation processes and realistic representations of geological, hydrogeological, and measurement uncertainty.

2. Problem Formulation

The groundwater-based concealed mineral identification problem was formulated as a simulation–optimisation (SO) inverse modelling framework in which the objective was to reconstruct source histories responsible for the observed groundwater concentration responses. Several studies with the same conceptualisation of linked simulation–optimisation have been conducted, particularly in areas of intrusions and contaminant tracking [5,18,23–28]. An illustrative groundwater system was adopted to provide a controlled environment for evaluating the proposed methodology for generating synthetic data. The current study focuses on the inverse modelling formulation development based on the output of the established numerical model. Numerical modelling codes such as MODFLOW for groundwater flow simulation and MT3DMS for contaminant transport simulation were used to construct the numerical model as constraint bindings of all the parameters [29]. Spatial discretisation of the study area was performed using the finite difference method (FDM), allowing groundwater flow and contaminant transport processes to be simulated over the entire modelling domain. The numerical model was developed using the USGS platform in which the MODFLOW and MT3DMS codes were embedded. A four-layered grid was built to represent the depth of the numerical model. These layers also represented layered geological heterogeneity in the modelled study area. The study area was discretised into smaller dimensions mathematically; the overall dimension of the model was 5000 m × 5000 m with a depth of 200 m.

Three species of copper Cu_{Tot} , ^{63}Cu and ^{65}Cu were structured inside the simulator and were controlled by the ISoMEC (Isotope Macro Engine Controller), an external coupling engine for the simulator. The ISoMEC is an external system engine built and coupled with the numerical simulator to enable automation and facilitate simulation by disallowing heavy file manipulation and back writing of previous numerical values due to poor synchronization. This computational simulation flaw arises due to the complex study area considered, including multiple layers and a large number of discretised cells. The original idea of automation was adopted from [27], in which the GWSID (Groundwater Source Identification) is an in-house programme which was built to couple the simulator and optimiser together. The ISoMEC builds around the same logic; however, the architecture of the system design facilitates the automation of the process to maintain the separation of the isotopes for computational efficiency. ISoMEC was developed to enable the automated execution of MODFLOW and MT3DMS without requiring the graphical user interface to remain open. This was achieved by externalising the simulation through preparing and formatting model input files. This architecture enabled execution directly through their file paths and executables located within the main operating directories as an external run. Macro commands integrated with the MATLAB toolbox environment with the version R2020b was used to control the execution sequence, synchronise model inputs and manage the simulation workflow, thereby providing a stable, reproducible and fully automated simulation platform.

The current study involves establishing the preliminary proof of concept for using the isotopic transport-based approach to develop a methodology for detecting unknown mineral deposits. At this proof-of-concept stage, the coupled reactive transport model has not been considered. However, the fully coupled reactive transport model has been developed to simulate isotope fractionation from geochemical reactions. The methodology applies the Adaptive Simulated Annealing (ASA) algorithm, which has the capability to search for a solution set based on its parameters [30,31]. In contrast, the present study

adopts a simplified transport-based framework in which a prescribed isotopic source signature is propagated through groundwater transport using isotope-specific adsorption parameters. This approach provides a head start for the methodology development in which the isotope problem is given a separate design. It assumes that fractionation has occurred at the deposit locations, which implies that the ratio at the point of origin or the deposit location has been specified as per the defined source isotope ratio.

These individual copper species were subsequently transported independently through the groundwater system using the implemented transport model, thereby generating concentration responses for each isotope signature at the specified observation points. The inverse problem was therefore formulated as the estimation of unknown, time-varying source fluxes that minimises the discrepancy between observed and simulated isotope signatures in terms of milligrams per litre by considering Cu_{Tot} , ^{63}Cu , and ^{65}Cu responses. The resulting forward simulation data subsequently formed the basis for surrogate model development and optimisation, allowing rapid evaluation of candidate source configurations without repeatedly executing the computationally intensive numerical groundwater model.

The identification of unknown contaminant source characteristics from groundwater observations represents a classical inverse modelling problem. In such problems, groundwater concentration measurements are available at monitoring locations, while the source release history responsible for generating the observed responses remains unknown. The objective is therefore to estimate the source characteristics that best reproduce the observed groundwater concentrations or isotope signatures. In the present study, the inverse problem was formulated as a source flux reconstruction problem for candidate concealed copper mineralisation zones. Figure 1 illustrates the potential mineral deposit locations considered within the study area through the model discretisation. These plausible mineral deposit locations were identified based on prior geological and hydrogeological interpretation and were treated as potential Cu isotope release zones. The above diagram is a plan view of the specified study area for the implemented numerical model, which has been solved as a four-layered three-dimensional system. This figure also shows the general groundwater flow direction from NW to SE, as indicated by the arrow.

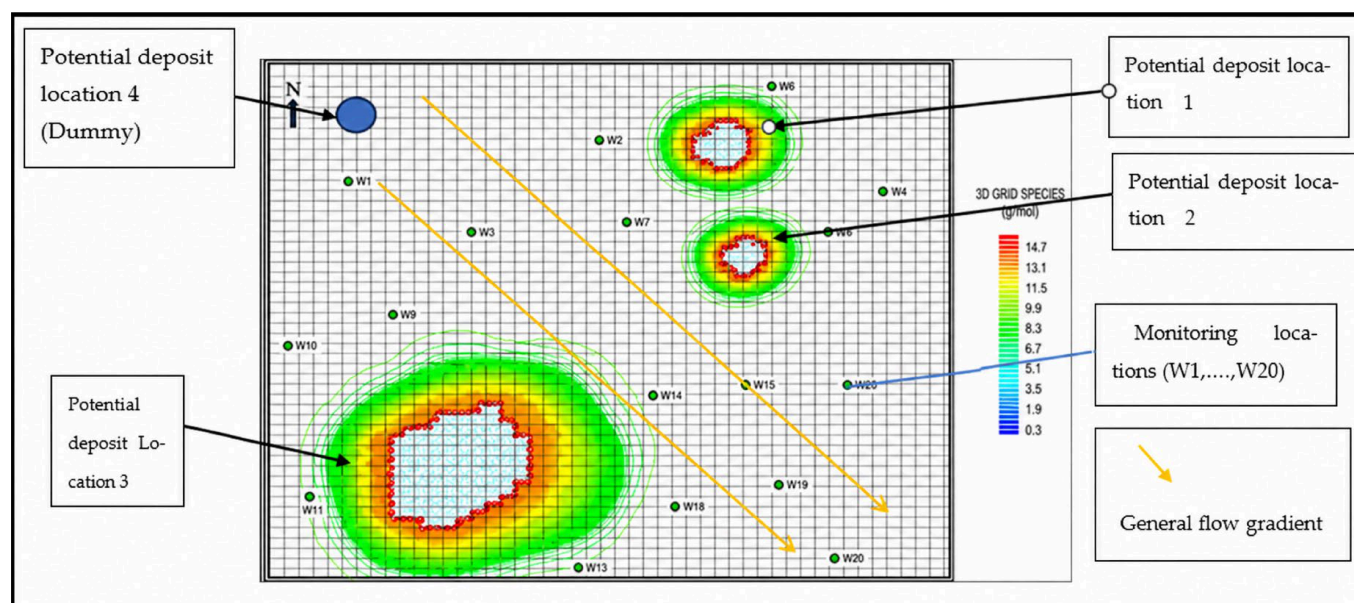


Figure 1. Schematic representation of the study area and the numerical simulation model.

The mineral deposit locations were assumed to be unknown to the inverse decision model for characterisation of the copper deposit locations. To represent temporal variations in source behaviour in terms of isotope release, the release history at each potential location was divided into five stress periods during which these isotopes were released. The use of stress periods was consistent with transient groundwater flow and transport modelling, where source conditions and hydraulic properties varied over time [29,32]. Each stress period was assigned independent flux values, allowing the optimisation framework to reconstruct the isotope release signatures throughout the incorporated modelling period. Continuous and persistent near-zero isotope fluxes from any of the potential locations imply dummy or non-existent deposit locations. Consequently, the optimisation problem consisted of twenty unknown decision variables corresponding to four candidate deposit locations (three active and one dummy deposit location) across five stress periods, as shown in Table 1. Each table entry, denoted by $[i,k]$, corresponding to a single decision variable representing the estimate for the actual source flux assigned to the i^{th} candidate deposit location during the k^{th} stress period. Consequently, the optimisation problem consisted of a total of 20 decision variables (4 candidate locations $\times$ 5 stress periods). This matrix therefore defines the optimisation search space, comprising 20 independent decision variables whose optimal values are estimated simultaneously to reconstruct the temporal source flux histories and identify the active deposit locations. This formulation enables both the spatial distribution and temporal evolution of the isotope signatures to be estimated simultaneously.

Table 1. Indexed representation of potential deposit location isotope fluxes over different time periods.

S [i,k]	Time Period 1 (320 Days)	Time Period 2 (640 Days)	Time Period 3 (960 Days)	Time Period 4 (1280 Days)	Time Period 5 (1600 Days)
Deposit Location 1	[1,1]	[1,2]	[1,3]	[1,4]	[1,5]
Deposit Location 2	[2,1]	[2,2]	[2,3]	[2,4]	[2,5]
Deposit Location 3	[3,1]	[3,2]	[3,3]	[3,4]	[3,5]
Dummy Deposit Location 4	[4,1]	[4,2]	[4,3]	[4,4]	[4,5]

The optimisation process sought to identify the set of flux values for each deposit location that produced isotopic responses most consistent with the observed data. ^{63}Cu , ^{65}Cu , Cu_{Tot} concentrations were utilised as the primary response variables for evaluating candidate solutions. The complexity of this illustrative inverse problem arises from the highly nonlinear relationship between these deposit locations' fluxes/release histories, groundwater flow and resulting isotopic signature responses. The non-linearity in the hydrogeological responses also arises due to the non-linear adsorption isotherms incorporated and due to the heterogeneous system resulting from the multiple geological strata. Optimisation using the numerical simulator directly requires thousands of computationally expensive model evaluations [33]. Therefore, a surrogate-assisted optimisation approach was adopted in which an MSVR (Multi-Surrogate Vector Regression) metamodel was developed to approximate simulator responses for the optimisation process [16,34]. This substantially reduced computational requirements while maintaining the ability to reconstruct source release histories with a high degree of accuracy. The resulting optimisation problem was subsequently solved using the Adaptive Simulated Annealing (ASA) algorithm, which systematically explored the decision-variable space and identified the source flux configuration that minimised the discrepancy between observed and predicted groundwater responses.

Model Construction and Domain

A three-dimensional model was built using a grid approach. The model contained the specified flow and transport parameters with a discretised grid frame. The main coverages such as wells, concealed mineral potential locations, hydraulic zones, aquifers and rivers were also specified. Three potential deposit locations were chosen as shown in Figure 1; that is, Locations 1, 2, and 3 were identified as three actual concealed mineral deposit locations, together with the fourth location chosen as a dummy or falsely identified location. The model domain was discretised into four hydrogeological layers comprising more than 40,000 computational cells. Constant-head boundaries were assigned along the model margins using prescribed groundwater-head values of 250 m, thereby establishing the regional hydraulic gradient and principal groundwater-flow direction from [NW] to [SE], as mentioned earlier and seen in Figure 1. Hydraulic conductivity was spatially distributed to represent geological layers and zones. Layers were assigned average values within the range shown in Table 1. Layer 1 was assigned an average value of 1.1 m/d, Layer 2 as 0.023 m/d, Layer 3 as 0.004 m/d, and Layer 4 as 0.0006 m/d. Where distinct geological zones occurred within a layer, separate conductivity values were assigned to represent differences in material permeability. A 30-year period numerical simulation was executed even though the actual transport times depend on groundwater velocities, hydraulic gradients, pore connectivity, retardation, reaction kinetics, weathering and source-release mechanisms. Thus, Table 2 summarises the parameter ranges, while Figure 1 illustrates their spatial distribution throughout the model domain. Groundwater flow within the model domain was established by assigning constant-head boundary conditions along the boundaries of the model domain to reproduce the regional hydraulic gradient across the aquifer. The physical flow boundary for the aquifer study area is actually away from the potential deposit locations; thus, these boundary conditions do not have sufficient effect on the flow in the vicinity of the potential deposit locations. Figure 1 shows only a part of the actual study area modelled. Therefore, the positions of the monitoring wells and the potential deposit sites are away from the physical boundary to reduce the boundary effects on the flow and transport processes in the vicinity of the potential deposit locations. The flow boundary conditions in general are constant flow boundaries specified at selected boundary cells. Hydrogeologic parameters such as porosity, specific yields, rate of recharge, water level heads, etc., were specified. Since the well heads were not steady with time, this assumes that the subsurface flow conditions were transient.

The nonlinear Langmuir adsorption formulation was selected because sorption capacity in natural porous media is finite and adsorption does not necessarily increase linearly with dissolved copper concentration. Unlike a linear isotherm with a constant distribution coefficient, the Langmuir formulation represents the progressive occupation of available sorption sites and the approach towards saturation at higher concentrations. Applying isotope-specific Langmuir parameters to ^{63}Cu and ^{65}Cu provided a controlled mechanism for producing different retardation and plume responses for the two isotopes. This enabled the evaluation of whether isotope-dependent transport responses provide additional information for identifying concealed copper sources. However, this simplified formulation does not represent all-natural copper-isotope fractionation processes, which may also involve mineral dissolution, precipitation, complexation, redox reactions, kinetic effects, and heterogeneous sorption. These additional processes should be incorporated into future site-specific reactive-transport investigations.

The specified constant-head boundaries that permitted groundwater inflow and outflow throughout the simulation period are presented in Table 2 below. The hydrogeological parameters used to construct the model had layer-based recharge uniformly distributed over the model surface area to represent natural groundwater replenishment. The differen-

tial values of the adsorption coefficients for ^{63}Cu , ^{65}Cu , and Cu_{Tot} are given in Table 2. No decay term is applied because both are stable isotopes.

The hydrogeological boundary conditions were treated as no-flow boundaries throughout the periphery of the specified study area, thereby restricting groundwater movement across those boundaries. This combination of boundary conditions enabled the development of a stable regional flow field suitable for contaminant transport simulations. Three potential copper mineral deposit locations were incorporated into the groundwater flow and simulation model. Rather than prescribing constant source releases, source fluxes were allowed to vary over five sequential stress periods, each of which was 320 days, enabling the simulation of different release histories. Groundwater isotopic signature's responses were available at twenty observation wells distributed throughout the model domain to capture the spatial and temporal evolution.

Table 2. Hydrogeological parameters used in the numerical model.

Aquifer Parameters	Values
Hydraulic Conductivities	m/d
Layer 1	Avg (0.9–3.2)
Layer 2	(0.01–0.51)
Layer 3	(0.003–0.05)
Layer 4	(0.0005–0.00001)
Adsorption Coefficient (Kd) for ^{65}Cu	1.4
Adsorption Coefficient (Kd) for ^{63}Cu	1.2
Adsorption Coefficient (Kd) for Cu_{Tot}	1.6
Observation Wells	20
Specific Storage	3.00×10^{-7}
Effective Porosity	
Layer 1	0.25
Layer 2	0.25
Layer 3	0.25
Layer 4	0.25
Specific Yield	
Layer 1	0.009
Layers 3–4	1.00×10^{-5}
Longitudinal Dispersivity	5 m
Transverse Dispersivity	0.01 m
Vertical Dispersivity	0.001 m
Average Vertical Recharge	1.00×10^{-8} (m/d)
Length	5000 m
Depth	200 m
Width	5000 m

3. Development of the Gaussian Process Regression Surrogate Model

The optimal estimation of concealed mineral deposit location characteristics required a large number of numerical model evaluations. Although the MODFLOW–MT3DMS sim-

ulation framework provided a physically realistic representation of groundwater flow and contaminant transport processes, repeated execution of the numerical simulator during optimisation can result in substantial computational demands. To overcome this limitation, a surrogate modelling approach was adopted to approximate the behaviour of the numerical simulator while significantly reducing computational requirements [9,25,35]. In the present study, a Gaussian Process Regression (GPR) metamodel was developed using synthetic datasets that were generated from the isotope-informed groundwater transport simulator. The surrogate model was designed to learn the complex nonlinear relationships between distinct deposit location characteristics and resulting groundwater concentration responses. Once trained, the GPR model was capable of rapidly mimicking groundwater responses without requiring repeated execution of the numerical simulator. Therefore, the machine learning (ML) pattern-based surrogate model could execute the function of the numerical simulator as a direct approximator. For the successful superseding of the numerical model by the metamodel, (GPR) was developed in three stages of training, validation and testing to an appropriate level of statistical satisfaction. Satisfactory performance results of the trained surrogate model needed to be achieved to verify its accuracy and efficiency. The training dataset was generated through a series of automations using the ISoMEC, which was linked to the numerical simulator files for feeding and updating the data. The resulting groundwater responses at observation wells were extracted and compiled into a database for surrogate model development. Latin Hypercube Sampling (LHS) was used to generate 2000 independent parameter combinations for the forward simulations. These runs were the true representation of the numerical model behaviour dictated by the flow and transport codes. The source-release fluxes were sampled from uniform distributions within the prescribed bounds of 0–100 mg/day, providing systematic coverage of the multidimensional input space. The selected number of simulations represented a practical balance between the computational cost of repeatedly executing MODFLOW–MT3DMS and the amount of data required to train the GPR surrogate. The fluxes of Source 1, Source 2, and Source 3 were independently sampled from uniform distributions bounded between 0 and 100 mg/day. The source starting time and release duration were defined according to the five stress periods of 320 days each within the 1600-day cumulative reconstruction period. ISoMEC automatically transferred each sampled parameter combination into the numerical model, executed the corresponding forward simulation, and extracted the concentration responses used to develop the GPR surrogate. Future studies should examine progressive training-set sizes, adaptive sampling, and active-learning approaches to determine the minimum simulation requirement for a specified surrogate accuracy.

The input variables used for surrogate model development included observation wells, observation time, source deposit locations and flux information. The output variables consisted of groundwater concentration responses for three different species of Cu_{Tot} , ^{63}Cu , and ^{65}Cu , which represented isotope signatures. Gaussian Process Regression (GPR) metamodels were selected due to their ability to accurately represent highly nonlinear systems while providing robust predictive capabilities [36,37]. Unlike conventional regression approaches that assume a predefined functional relationship between inputs and outputs, GPR constructs predictions using covariance structures learned directly from the training data. This characteristic makes GPR particularly suitable for groundwater transport problems where complex interactions exist between hydrogeological parameters, deposit location conditions and contaminant migration processes [38]. Model training involved training the surrogate with the numerical simulation data and subsequently evaluating its predictive performance using independent validation and testing datasets. The validation process was undertaken to assess the ability of the surrogate model to generalise beyond the training data and accurately reproduce unseen simulator responses. The resulting sta-

tistical performance metrics demonstrated excellent predictive capability across all model outputs [39,40]. High coefficients of determination (R^2) and low error statistics indicated strong agreement between numerical simulator outputs and surrogate predictions. These results confirmed that the trained GPR metamodel was capable of accurately reproducing the functional behaviour of the numerical simulator and could therefore be used as a computationally efficient substitute during the optimisation stage. The validated GPR surrogate models subsequently formed the core predictive engine of the inverse modelling framework. This significantly improved computational efficiency and enabled practical implementation of the Adaptive Simulated Annealing (ASA) optimisation framework for the isotope signature source reconstruction.

3.1. Training, Validation and Testing Dataset Development

The Gaussian Process Regression (GPR) surrogate model was developed using a comprehensive dataset generated from repeated forward simulations by the numerical groundwater flow and transport model as discussed in Section 3. Rather than relying on analytical solutions, the GPR metamodel was trained entirely from numerical realisations produced by the externalised MODFLOW–MT3DMS simulation framework developed in this study. The simulation framework automatically modified deposit location parameters executed by the numerical model and extracted the resulting groundwater concentration responses at the pre-defined observation wells using the Latin hypercube sampling (LHS) method. The simulated dataset generated consisted of multiple realisations representing different deposit location release scenarios. Each realisation was characterised by variations in magnitude, source activation period, observation well location and observation time. For every simulation, the numerical model produced groundwater concentrations for the three transported species Cu_{Tot} , ^{63}Cu , and ^{65}Cu , together with the corresponding isotopic composition $\delta^{65}\text{Cu}$. These numerical outputs formed the target variables for surrogate model development. The generalisation performance of the GPR surrogate was evaluated using an independent testing dataset containing previously unseen LHS-generated simulation realisations. Data partitioning was performed according to simulation identifiers rather than individual observation rows. Therefore, all records associated with a given realisation were assigned exclusively to the training, validation, or testing dataset. No overlap of simulation realisations occurred among the three subsets.

The predictor variables supplied to the GPR model consisted of seven input parameters:

- Observation well identifier.
- Simulation time (days).
- Duration of the travel.
- Deposit Location 1: isotope release activity as source fluxes.
- Deposit Location 2: isotope release activity as source fluxes.
- Deposit Location 3: isotope release activity as source fluxes.
- Deposit Location 4: isotope release activity as source fluxes.

The response variables comprised four outputs:

- Cu_{Tot} : total concentration.
- ^{65}Cu concentration.
- ^{63}Cu concentration.
- $\delta^{65}\text{Cu}$.

Prior to the stage of model training, the complete numerical simulation dataset of input and output was compiled together for developing the surrogate. From the entire runs of input and corresponding output, data was randomly divided into three independent subsets to ensure unbiased model evaluation, as presented in the earlier research conducted by [12,28] for the saltwater intrusion. From the techniques of using the LHS, it was taken

care that no two runs had duplicates or similar values thus distinguishing each one from one another. The uniqueness of each run with different source inputs generated a set of non-identical outputs, which were then used for categorisation into three sets of data, particularly training, validation and testing. It has been proven that such categorisation allows data to provide statistical confidence to check the performance of the metamodel in terms of efficiency and associated uncertainties. A total of 70% of the data were assigned to the training dataset, which was used to estimate the hyperparameters of the GPR. A total of 20% of the remaining data were used for the validation dataset, which was employed during model development to monitor prediction accuracy and reduce the possibility of overfitting. Finally, 10% was used as an independent testing dataset, which remained unseen during model construction. This three-stage data partitioning ensured that the developed surrogate was capable of accurately reproducing groundwater transport behaviour over a wide range of deposit locations and release conditions while maintaining strong predictive performance on previously unseen scenarios [12]. From the entire data generation and training process, the workflow was automated through the coupling of ISoMEC and MODFLOW-MT3DMS. The execution was through the command-based window hosted inside the MATLAB Version R2020b toolbox. The automated coupling of ISoMEC and Simulator generated 2000 runs of the dataset, distinct from each one. 1400 runs were used for the training of the GPR metamodel, 400 runs for validation and 200 for testing.

3.2. Surrogate Model Performance Assessment

The predictive capability of the developed surrogate, which is the GPR metamodel, was evaluated using independent validation and testing datasets. Model performance was quantified using four commonly adopted statistical indicators, namely the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Normalised Root Mean Square Error (NRMSE). These statistical indices are widely employed in hydrological modelling, machine learning and environmental surrogate modelling because they provide complementary measures of prediction accuracy [41–44]. The coefficient of determination measures the proportion of variance explained by the surrogate model and is expressed as Equation (1) below:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

where y_i denotes the simulated numerical-model value, $\hat{y}_i$ is the surrogate prediction, and $\bar{y}$ is the mean of the observed values.

The Root Mean Square Error (RMSE) is calculated as shown by the formula in (2)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

while the Mean Absolute Error (MAE) is shown in (3) below

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

To facilitate comparisons among variables with different numerical scales, the RMSE was normalised as (4)

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \quad (4)$$

The coefficient of determination (R^2) measures the proportion of variance in the numerical model outputs that is explained by the surrogate model [38]. A R^2 value of 1 indicates perfect agreement between predicted and simulated values, whereas values approaching zero indicate poor predictive capability. In general, R^2 values greater than 0.90 are considered indicative of good predictive performance for environmental modelling applications [41]. In the present study, the surrogate model (GPR) achieved testing R^2 values exceeding 0.9998, with several outputs producing values effectively equal to unity, demonstrating that the developed GPR metamodel successfully reproduced the numerical simulator responses with almost no loss of information [42]. The Root Mean Square Error (RMSE) quantifies the average magnitude of prediction errors while assigning greater weight to larger deviations due to the squared residual term. Consequently, RMSE is particularly sensitive to large prediction errors and is commonly regarded as one of the most informative indicators of overall model performance [38]. The low RMSE values obtained for all outputs indicate that prediction deviations between the surrogate and the numerical simulator remained consistently small throughout the dataset. Unlike RMSE, the Mean Absolute Error (MAE) represented the average absolute prediction error and was less influenced by isolated large deviations. MAE therefore provided a direct measure of the average prediction accuracy and is often recommended alongside RMSE to provide a balanced assessment of surrogate model performance [39]. The consistently low MAE values obtained for both validation and testing datasets further demonstrate the robustness and stability of the developed GPR metamodels. Since the simulated outputs span different numerical ranges, the Normalised Root Mean Square Error (NRMSE) was also computed to facilitate comparison among variables. NRMSE expresses the prediction error relative to the range of observed values, allowing model performance to be compared across outputs with different units and magnitudes. Lower NRMSE values indicate greater predictive accuracy, with values below approximately 10% generally considered indicative of highly accurate predictive models in environmental applications [39].

In the present study, the NRMSE values were substantially below 1% for all predicted variables. This reveals that most outputs exhibited values below 0.3%, confirming that prediction errors were negligible relative to the magnitude of the numerical simulator responses. Overall, the combined evaluation of R^2 , RMSE, MAE, and NRMSE demonstrates that the Gaussian Process Regression (GPR) surrogate provided an exceptionally accurate representation of the MODFLOW–MT3DMS numerical simulator. The close agreement between the validation and testing statistics further indicated excellent generalisation capability and suggested that the developed surrogate model exhibits minimal overfitting. Such predictive performance provided confidence that the surrogate model could reliably replace repeated numerical simulations for computationally intensive optimisation and inverse deposit location [40,43,44]. The exceptionally high predictive performance of the GPR can be partly attributed to the structured relationships among the three simulated outputs Cu_{Tot} , ^{63}Cu , and ^{65}Cu . The isotope concentrations were generated through mass partitioning of the total copper concentration according to the prescribed source isotope ratio and were subsequently governed by the same groundwater-flow and transport system. Consequently, the outputs were strongly correlated and varied smoothly across the sampled parameter space, reducing the complexity of the input–output relationships learned by the GPR. Isotope-dependent transport and sorption could modify the original proportions; therefore, the isotope concentrations were not treated as fixed fractions of Cu_{Tot} . Nevertheless, their physically constrained relationships contribute to the high GPR performance. This interpretation is supported by predicted-versus-simulated plots and error metrics obtained from independent testing realisations.

The results demonstrated in Table 3 show excellent predictive capability of the proposed surrogate model across all output variables. The testing R^2 values exceed 0.9998 for every response, with several outputs achieving values effectively equal to unity, indicating that the GPR successfully reproduced the numerical simulator responses with negligible loss of accuracy. Similarly, the RMSE and MAE values remained consistently low for both the validation and testing datasets, while the NRMSE values were below 0.23% for all outputs, confirming that prediction errors were extremely small relative to the range of the simulated responses. A close agreement between the validation and testing statistics further indicates that the surrogate models generalise well to unseen simulation scenarios and exhibit no evidence of significant overfitting. This level of predictive accuracy demonstrates that the GPR model would provide a computationally efficient yet highly realistic approximation of the original MODFLOW–MT3DMS numerical simulator [45]. This representation of the numerical model makes the approach well suited for integration within the simulation–optimisation (SO) framework developed in this study. The resulting surrogate substantially reduces computational cost while preserving the accuracy required for inverse groundwater deposit location identification and flux estimation.

Table 3. Summarises the predictive performance evaluation for Gaussian Polynomial Regression training process surrogate models.

Output	Validation RMSE	Validation MAE	Validation R^2	Testing RMSE	Testing MAE	Testing R^2	Testing NRMSE
Cu_{tot}	0.04689	0.01610	0.99995	0.04741	0.01619	0.99995	0.00160
^{65}Cu	0.03016	0.00571	0.99980	0.03012	0.00571	0.99980	0.00229
^{63}Cu	0.01227	0.00453	0.999995	0.01227	0.00453	0.999995	0.000495

4. Adaptive Simulated Annealing Optimisation Framework

Following the successful development and validation of the Gaussian Process Regression (GPR) surrogate model, an inverse optimisation framework was established to reconstruct the unknown deposit location release histories associated with the candidate concealed copper mineralisation zones. The objective of the optimisation process was to determine the temporal deposit location flux distribution that minimised the discrepancy between observed concentrations as specified and those predicted by the optimisation model. The Adaptive Simulated Annealing (ASA) was selected as the primary optimisation algorithm for the proposed framework. ASA demonstrated its proven capability for solving highly nonlinear inverse problems characterised by multiple local optima, discontinuous objective surfaces and high-dimensional parameter spaces [42,46,47]. Inverse groundwater source identification is recognised as a computationally challenging optimisation problem because numerous combinations of mineral deposit release histories can generate similar transport geochemical responses [6,48]. This non-uniqueness results in an objective function containing multiple local minima, making conventional gradient-based optimisation techniques susceptible to premature convergence. Other algorithms such as Particle Swarm Optimisation (PSO) and Genetic Programming (GP) would also have been effective in searching for the optimal solution [16]. Adaptive Simulated Annealing (ASA) overcomes these limitations by incorporating probabilistic acceptance criteria, allowing temporary acceptance of inferior solutions during the early stages of optimisation. Such behaviour enhances exploration of the global search space while progressively reducing randomness through an adaptive cooling schedule, thereby improving convergence towards the global optimum [49,50]. Unlike classical simulated annealing, ASA automatically adjusts the search distribution according to the sensitivity of each decision variable. The adaptive

annealing schedule accelerates convergence without sacrificing global search capability, making the algorithm particularly suitable for inverse contaminant deposit locations characterisation problems where objective function landscapes are highly irregular [51]. Within the present study, ASA was directly coupled with the trained GPR metamodel. Instead of repeatedly executing the computationally demanding MODFLOW–MT3DMS numerical simulator during every optimisation iteration, groundwater concentrations were predicted instantaneously using the trained GPR metamodel. Consequently, thousands of candidate deposit location configurations could be evaluated within a practical computational time while preserving the predictive behaviour of the original numerical simulator.

4.1. Decision Variables

The optimisation problem consisted of twenty unknown decision variables representing deposit locations and Isotope fluxes distributed across three of these potential deposit locations and one dummy potential deposit location incorporating five sequential stress periods. The decision vector is expressed as

$$\mathbf{x} = [S_{1,1}, S_{1,2}, S_{1,3}, S_{1,4}, S_{1,5}, S_{2,1}, S_{2,2}, S_{2,3}, S_{2,4}, S_{2,5}, S_{3,1}, S_{3,2}, S_{3,3}, S_{3,4}, S_{3,5}, S_{4,1}, S_{4,2}, S_{4,3}, S_{4,4}, S_{4,5}]$$

where

- $S_{(i,k)}$ denotes the contaminant deposit location flux (mg/day);
- i represents the candidate deposit location ($i = 1, 2, 3, 4$);
- k denotes the stress period of 320 days ($k = 1, 2, 3, 4, 5$).

Each stress period represents a fixed release interval within the simulation period, allowing the optimisation algorithm to reconstruct the temporal evolution of contaminant release from each deposit location.

4.2. Candidate Solution Representation

Within the Adaptive Simulated Annealing (ASA) optimisation process, a candidate solution represents one possible mineral deposit location release scenario proposed by the optimisation algorithm. Rather than representing isotope signatures directly, each candidate solution consists of a complete set of flux values assigned to the predefined deposit points over the specified stress periods. For every candidate solution generated by ASA, the mineral deposit flux values are supplied to the trained (GPR) metamodel, together with the observation well locations and simulation times. The metamodel predicts the corresponding species concentrations, which are then compared with the observed monitoring data through the objective function stated in Equation (5). Throughout the optimisation process, ASA continuously generated thousands of such candidate solutions. Poor-performing solutions were gradually discarded, while promising solutions were retained and refined through the adaptive annealing process. As the search progressed, the candidate solutions progressively converged towards a deposit location release history that minimised the discrepancy between observed and predicted species concentrations. The final optimal candidate solution therefore represented the most probable deposit location distribution capable of reproducing the measured groundwater responses within the study area.

5. Objective Function Formulation

The success of the inverse source identification process depends on the ability of the optimisation algorithm to identify flux configurations for a particular concealed mineral deposit location which accurately reproduces the observed groundwater responses. To achieve this objective, an objective function was formulated to quantify the discrepancy

between observed groundwater concentrations and the concentrations predicted by the GPR surrogate model.

The optimisation objective was formulated as the minimisation of the sum of squared residuals between the observed groundwater concentrations and the concentrations estimated by the GPR surrogate model. The objective function minimised the combined absolute concentration mismatch across all species, observation wells, and sampling times. All three outputs were expressed in mg/L, and their residuals were aggregated directly. Nevertheless, their numerical concentration ranges differed; therefore, the unnormalised Square Error formulation gave greater influence to variables with larger concentration magnitudes. In the present results, Cu_{Tot} , ^{63}Cu , and ^{65}Cu contributed approximately 48.90%, 38.12%, and 12.99% of the total objective-function error, respectively. Thus, ^{65}Cu retained an influence on the inverse solution but contributed less because of its smaller concentration magnitude. Future applications could employ species-specific normalisation, uncertainty-based weighting, or inverse-variance weighting to provide a more balanced contribution from each observation type.

During each iteration, the Adaptive Simulated Annealing (ASA) algorithm generated a candidate source flux configuration, which was supplied to the trained GPR surrogate model to estimate the corresponding Cu_{Tot} , ^{63}Cu , and ^{65}Cu concentrations at all monitoring wells and observation times. The optimal solution was identified as the source flux configuration that minimised the cumulative estimation error while satisfying the prescribed source flux bounds.

The linked optimisation simulation model for unknown source identification can be represented by Equations (5) and (6), which:

1. Minimised the summation of the difference between the estimated and the observed isotope signatures and the summation of the difference between the total copper concentrations and estimated copper concentrations, at each monitoring location at different times.
2. The linked simulation–optimisation constraints represented the simulation model as a set of binding constraints.

The optimisation simulation model was mathematically represented as follows:

$$\text{Minimise } F(X) = \sum_{k=1}^{n_k} \sum_{i_{ob}=1}^{n_{i_{ob}}} \left(\left(C_{\text{Cu}(\text{Tot})\text{obs}, i_{ob}}^k - C_{\text{Cu}(\text{Tot})\text{est}, i_{ob}}^k \right)^2 + \left(C_{(^{63}\text{Cu})\text{obs}, i_{ob}}^k - C_{(^{63}\text{Cu})\text{est}, i_{ob}}^k \right)^2 + \left(C_{(^{65}\text{Cu})\text{obs}, i_{ob}}^k - C_{(^{65}\text{Cu})\text{est}, i_{ob}}^k \right)^2 \right). \quad (5)$$

Subject to:

$$C_{\text{est}, i_{ob}}^k = f(q_s, C_s), \text{ Where } f(q_s, C_s, C_{\text{est}, i_{ob}}^k) \forall i_{ob}, k. \quad (6)$$

The set of constraints in Equation (6) essentially represents the linking of the optimisation algorithm with the numerical groundwater flow and transport simulation model through the decision variable where:

C_s is the Isotope concentration at the source location (ML^{-3}).

q_s is the volumetric flux of water per unit volume of aquifer (T^{-1}).

$q_s C_s$ is the Isotope flux ($\text{ML}^{-3}T^{-1}$).

$C_{\text{est}, i_{ob}}^k$ is the concentration estimated by the transport simulation model at observation monitoring location i_{ob} and at the end of time period k (ML^{-3}).

$S_{(i,k)}$ is the isotope flux at deposit locations (mg/day), with:

i = the location;

k = the stress period.

- $F(X)$ is the objective function to be minimised.
- X is the vector of unknown optimisation decision variables representing the source fluxes.
- k denotes the observation time index.
- n_k is the total number of observation times.
- i_{ob} denotes the groundwater observation well index.
- n_{ob} is the total number of groundwater observation wells.
- $C_{Cu(Tot),obs,i_{ob}}^k$ is the observed (synthetically generated) total copper concentration at observation well i_{ob} and time k .
- $C_{Cu(Tot),est,i_{ob}}^k$ is the surrogate-estimated total copper concentration corresponding to the trial solution.
- $C_{(^{65}Cu),obs,i_{ob}}^k$ and $C_{(^{65}Cu),est,i_{ob}}^k$ are the observed and surrogate-estimated concentrations of the ^{65}Cu isotope, respectively.
- $C_{(^{63}Cu),obs,i_{ob}}^k$ and $C_{(^{63}Cu),est,i_{ob}}^k$ are the observed and surrogate-estimated concentrations of the ^{63}Cu isotope, respectively.

$f(x, y, z, C_{est})$ represents the concentration simulation results at location coordinates defined by x, y, z obtained from the transport simulation model.

The objective function was evaluated using observations collected from multiple monitoring wells, distributed throughout the study area. For each candidate solution generated by the Adaptive Simulated Annealing (ASA) algorithm, the corresponding groundwater responses were predicted using the trained GPR surrogate model. The residual differences between observed and predicted concentrations were subsequently calculated and aggregated to produce a single objective function value. The decision variable vector consisted of twenty source fluxes corresponding to four candidate deposit locations with five stress periods, as depicted in Table 1, where $S[i,k]$ denotes the source flux associated with source location (i) during stress period (k). The objective function was designed to simultaneously preserve information associated with total isotope-specific transport behaviour. Incorporating Cu_{Tot} , ^{63}Cu and ^{65}Cu within the optimisation process ensured that both overall isotope migration pattern responses contributed to the deposit location history reconstruction process. This represented a more complex scenario of groundwater system behaviour compared to that which would be obtained using a single species concentration history alone. The objective function value served as the primary measure of solution quality throughout the optimisation process. Lower computational objective function values indicated improved agreement between observed and predicted groundwater responses, while the minimum objective function identified by the ASA algorithm represented the optimal release flux history for the investigated groundwater system.

6. Optimisation Results and Deposit Location Flux Reconstruction

The optimisation process was achieved through the coupling of the Adaptive Simulated Annealing (ASA) algorithm with the trained GPR surrogate model, with the objective function as mentioned earlier. During each optimisation iteration, ASA generated a candidate deposit flux history consisting of twenty decision variables corresponding to the four candidate deposit locations and five stress periods of 320 days each. The 320-day stress period was reasonably taken by the decision model as the best interval to mathematically converge to the best solution. This stress period value of 320days was optimally favourable, as the tracer element of copper had three species and all three species needed to undergo normalisation. These fluxes from locations were supplied to the trained GPR metamodel, which rapidly estimated the corresponding isotopic signature responses at all monitoring wells and observation times. The optimal estimate of the deposit location, as explained above, was based on the best match between the measured and simulated isotope signatures. The candidate locations are pre-defined and once a particular deposit

location is identified, the corresponding coordinates of locations are obtained. The estimated isotopic concentrations were subsequently compared with the observed isotopic concentrations, which were evaluated to quantify the agreement between the two datasets. The optimisation algorithm continuously updated the candidate flux histories until the minimum objective function value was achieved, indicating convergence to the optimal deposit location configuration.

6.1. Reconstructed Deposit Locations Flux Histories

Table 4 summarises the source-flux histories reconstructed using the coupled Adaptive Simulated Annealing–Gaussian Process Regression (ASA–GPR) optimisation framework. Figure 2 directly compares the actual input fluxes used in the forward groundwater-transport simulation with the corresponding fluxes reconstructed by the ASA–GPR framework. Four candidate source locations were included in the inverse analysis. Locations 1–3 represented active sources, whereas the dummy location was assigned zero actual flux throughout the simulation to test whether the framework could distinguish an inactive candidate location from the active sources. Although the GPR surrogate was trained using data generated over a 30-year transport period, the inverse reconstruction considered 1600 days from the entire data set of the 30-year period. This interval was divided into five equal stress periods (SP1–SP5), each lasting 320 days. Consequently, the inverse problem contained 20 decision variables representing the temporally varying fluxes of the four candidate locations.

Table 4. Twenty deposit locations, fluxes and the stress periods with the corresponding durations.

Location	Stress Period	Estimated Flux mg/Day	Actual Flux mg/Day
Deposit Location 1	1	66.3	64.2
	2	20.1	18.2
	3	23.0	16.1
	4	28.4	25.2
	5	12.9	13.1
Deposit Location 2	1	3.5	7.9
	2	30.3	28.3
	3	39.8	37.0
	4	36.8	43.0
	5	64.7	53.0
Deposit Location 3	1	42.4	30.1
	2	92.4	88.2
	3	99.8	93.2
	4	25.9	23.1
	5	99.9	89.2
Deposit Location 4	1	0.0	0.0
	2	0.0	0.0
	3	0.0	0.0
	4	0.0	0.0
	5	0.0	0.0

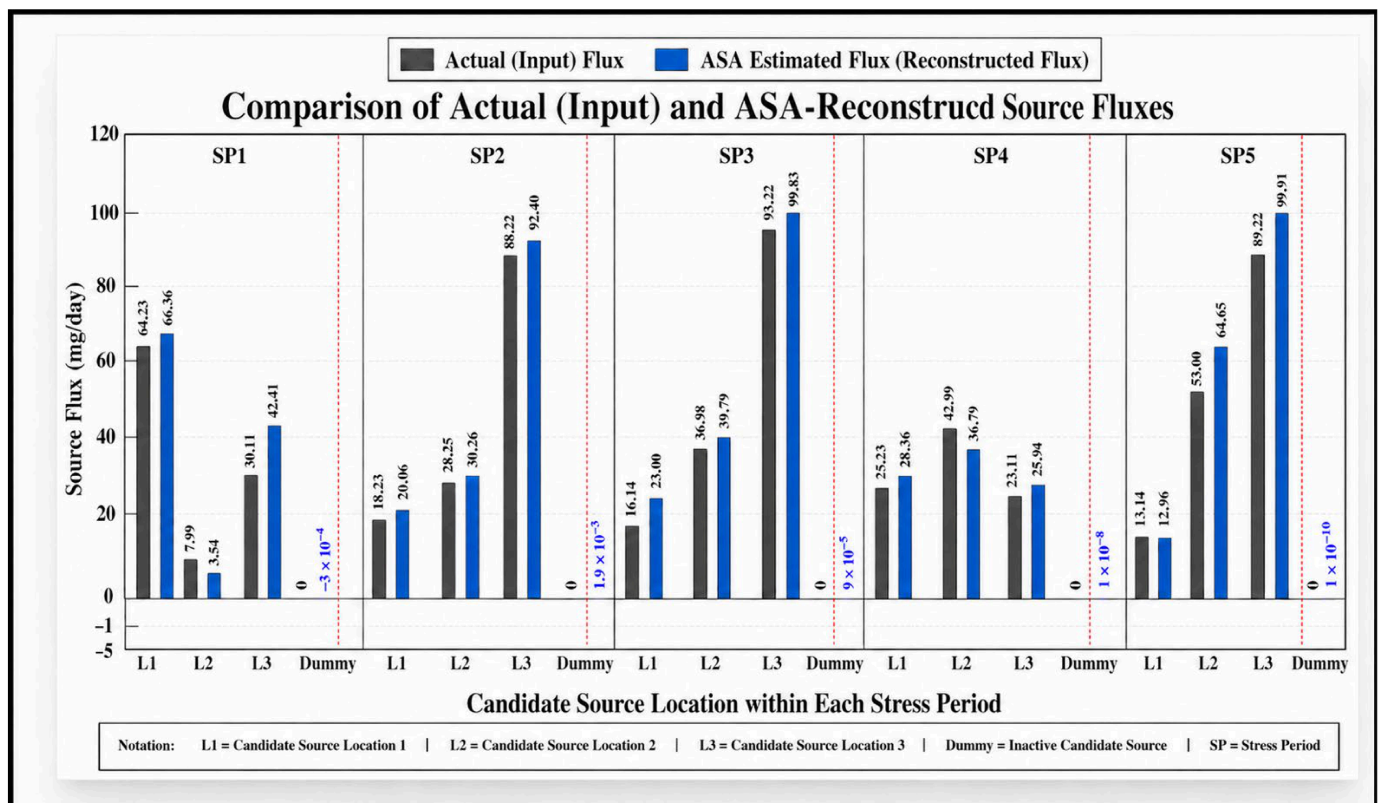


Figure 2. The optimised flux for three active deposit locations and one dummy location with their five time-step periods.

As shown in Figure 2, the reconstructed fluxes generally corresponded with the actual input fluxes in terms of both magnitude and temporal variation, although the level of agreement differed among locations and stress periods. For Location 1, the reconstructed fluxes were 66.36, 20.06, 23.00, 28.36 and 12.96 mg/day, compared with actual values of 64.23, 18.23, 16.14, 25.23 and 13.14 mg/day, respectively. The reconstruction was particularly close in SP1 and SP5, with absolute differences of only 2.13 and 0.18 mg/day. The largest discrepancy occurred in SP3, where the flux was overestimated by 6.86 mg/day. Nevertheless, the reconstruction correctly reproduced the dominant high flux in SP1, subsequent reduction in SP2–SP3, modest increase in SP4 and the further decline in SP5.

For Location 2, the reconstructed values were 3.54, 30.26, 39.79, 36.79 and 64.65 mg/day, whereas the actual input values were 7.99, 28.25, 36.98, 42.99 and 53.00 mg/day. The agreement was closest in SP2 and SP3, where the absolute differences were approximately 2.01 and 2.81 mg/day. Larger deviations occurred in SP1, SP4 and SP5. In particular, the reconstruction underestimated the SP1 and SP4 fluxes by approximately 4.45 and 6.20 mg/day, respectively, and overestimated the SP5 flux by approximately 11.66 mg/day. Thus, the ASA–GPR framework captured the overall increase in the contribution of Location 2 from the early to the later stress periods, but it did not reproduce every stress-period value with the same degree of accuracy. For Location 3, reconstructed fluxes of 42.41, 92.40, 99.83, 25.94 and 99.91 mg/day were obtained, compared with actual values of 30.11, 88.22, 93.22, 23.11 and 89.22 mg/day. The closest agreement occurred in SP2–SP4, with absolute differences ranging from approximately 2.83 to 6.60 mg/day. More noticeable overestimations occurred in SP1 and SP5, where the reconstructed values exceeded the actual values by approximately 12.30 and 10.68 mg/day, respectively. Despite these differences, the reconstruction reproduced the sharp increase from SP1 to SP2, the high flux during SP2–SP3, the pronounced decrease in SP4 and the renewed increase in SP5. The comparison

also correctly identified Location 3 as the largest source in SP2, SP3 and SP5. However, Location 1 was the dominant source in SP1, while Location 2 had the largest actual flux in SP4.

The dummy location provides an additional assessment of the source-identification capability of the framework. Its actual input flux was zero in all five stress periods, while the reconstructed values remained between approximately -0.0003 and 0.0019 mg/day. These values were negligible compared with the fluxes reconstructed for the three active locations and could be regarded as effectively zero within the numerical tolerance of the optimisation. The close correspondence between the zero actual flux and the near-zero reconstructed flux indicated that the framework correctly classified the dummy location as inactive and did not assign it a substantial false-positive contribution. This approach was taken to ensure that the methodology could verify the proof of concept with the control experiment ideology to be able to identify the dummy sources. This correct identification of the dummy source provided assurance that false source identification has a lower possibility of being picked up by the optimiser before the methodology could roll out in real field scenarios.

Overall, Figure 2 demonstrates that the ASA-GPR framework reproduced the principal magnitudes and values, temporal changes and relative contributions of the three active locations while correctly identifying the inactive dummy location. The comparison also reveals that reconstruction accuracy was not uniform across all stress periods, with the largest differences occurring for Location 1 in SP3, Location 2 in SP5 and Location 3 in SP1 and SP5. Therefore, the results support the ability of the framework to reconstruct the main characteristics of temporally varying source-flux histories, while also showing the stress periods in which uncertainty or reconstruction error was comparatively greater. The anomalies for Location 1/SP3 and Location 2/SP5 were examined explicitly. At Location 1 during SP3, ASA-GPR reconstructed a flux of 23.00 mg/day compared with the actual input flux of 16.14 mg/day, representing an overestimation of 6.86 mg/day. At Location 2 during SP5, the reconstructed flux was 64.65 mg/day compared with the actual value of 53.00 mg/day, representing an overestimation of 11.65 mg/day. These results indicate that the framework correctly identified both locations as active sources and reproduced their broader temporal behaviour, although it did not recover every stress-period flux with equal accuracy. These discrepancies reflected the non-uniqueness of the inverse problem, overlapping concentration responses from different sources, limitations in the spatial and temporal information provided by the monitoring network, GPR approximation error, and the stochastic convergence behaviour of ASA. Consequently, successful identification of an active source location should be distinguished from exact reconstruction of its temporal flux magnitude. Uncertainty in the reconstructed source fluxes may produce false-positive or false-negative exploration targets when the framework is applied under field conditions. A false-positive result may occur when a non-mineralised candidate location is assigned a substantial reconstructed flux, whereas a false-negative result may occur when an active mineralised location is assigned a negligible flux. Such errors may arise from measurement noise, hydrogeological uncertainty, overlapping plume responses, model structural error, surrogate approximation, and the non-uniqueness of the inverse problem. Therefore, the proposed framework should primarily be considered a target-ranking and decision-support tool rather than an independent method for confirming mineral deposits. Candidate locations identified by the framework must ultimately be evaluated alongside geological, geophysical and geochemical evidence and verified through field investigation and exploration drilling.

The reconstructed deposit location histories provided insight into both the magnitude and temporal evolution of plume release, which was generated from the concealed mineral deposit location. Stress periods exhibiting larger optimised flux values indicated

intervals during which loading exerted a greater influence on the observed groundwater concentrations. Conversely, lower flux values suggest periods of reduced release activity or diminished contribution to groundwater contamination.

6.2. Objective Function Convergence

The effectiveness of the optimisation framework was evaluated using the difference in the objective function values obtained following convergence as a proof of concept from other research [27,52]. The minimum overall objective function value achieved by the optimisation process was 0.075, representing the cumulative residual error between the observed groundwater concentrations and those estimated by the GPR metamodel. To provide a physically meaningful interpretation of the objective function value, a normalisation analysis was performed using the observed groundwater concentrations. The observed concentrations for Cu_{Tot} , ^{63}Cu and ^{65}Cu from all 20 monitoring wells over 5 sampling times (100 observations per species) were summed and averaged individually. The total observed concentrations were 2630.61 mg/L for Cu Total, 2455.13 mg/L for ^{63}Cu and 2462.15 mg/L for ^{65}Cu , corresponding to average concentrations of 26.31 mg/L, 24.55 mg/L and 24.62 mg/L, respectively. The combined average concentration of the three species was therefore 75.48 mg/L, representing the overall concentration magnitude against which the optimisation error was evaluated. Since the objective function was based on squared residuals, its square root was first determined to restore the units to concentration. The square root of the minimised final objective function value was 0.274 mg/L. The square root of the minimum objective-function value was first calculated as 0.274 mg/L. The mean concentrations of Cu_{Tot} , ^{63}Cu , and ^{65}Cu were 26.31, 24.55, and 24.62 mg/L, respectively, giving a combined reference concentration of 75.48 mg/L. Dividing 0.274 mg/L by 75.48 mg/L produced a normalised ratio of 0.00363. Multiplying this ratio by 100 gave a normalised objective-function error of approximately 0.36%. This indicated that the residual error remaining after optimisation was less than 1% of the average observed groundwater concentration across all three copper species. The lower error of less than 1% demonstrated that the ASA–GPR optimisation framework achieved an excellent agreement between the estimated and observed concentration datasets with a very high degree of accuracy, providing strong confidence in the reconstructed source flux. Table 5 below presents the total of 100 observations and its corresponding averages, which were used to normalise.

Table 5. Total observed concentrations and corresponding average concentrations of Cu, ^{65}Cu , and ^{63}Cu calculated from 100 observations (20 monitoring wells $\times$ 5 time steps) for use in the normalisation of the objective function.

Species	Total of 100 Observations (mg/L)	Average (mg/L)
Cu_{Tot}	2630.61	26.30
^{65}Cu	2462.15	24.62
^{63}Cu	2455.13	24.55

The relatively small residual values obtained for all three species indicated that the optimised flux histories provided a close agreement between the observed and estimated groundwater concentrations, as mentioned earlier. Beyond an objective function value of 0.075, successive optimisation iterations produced negligible reductions in the residual error, indicating that the ASA algorithm had converged to a stable solution as shown in Figure 3. ASA was selected because the inverse source-identification problem was treated as nonlinear, potentially multimodal, and potentially lacking analytical gradients. ASA performed a global search and could probabilistically accept inferior trial solutions,

allowing it to escape local minima. Its adaptive temperature and step-size mechanisms also progressively shifted the search from broad exploration to local refinement. Furthermore, the convergence and the optimised results revealed that ASA could potentially be coupled directly with the GPR surrogate and can efficiently handle the bounds and time-varying source-flux decision variables. In contrast, population-based methods such as PSO generally require evaluating many candidate solutions during each iteration and may converge prematurely without careful parameter tuning. However, ASA is not claimed to be universally superior; it was selected as a suitable and previously established method for groundwater inverse problems, while a formal comparison with PSO and other algorithms remains an important subject for future research. The convergence behaviour demonstrated the capability of the optimisation framework to efficiently explore the high-dimensional decision-variable space and identify a flux configuration that best reproduces the observed groundwater responses [7].

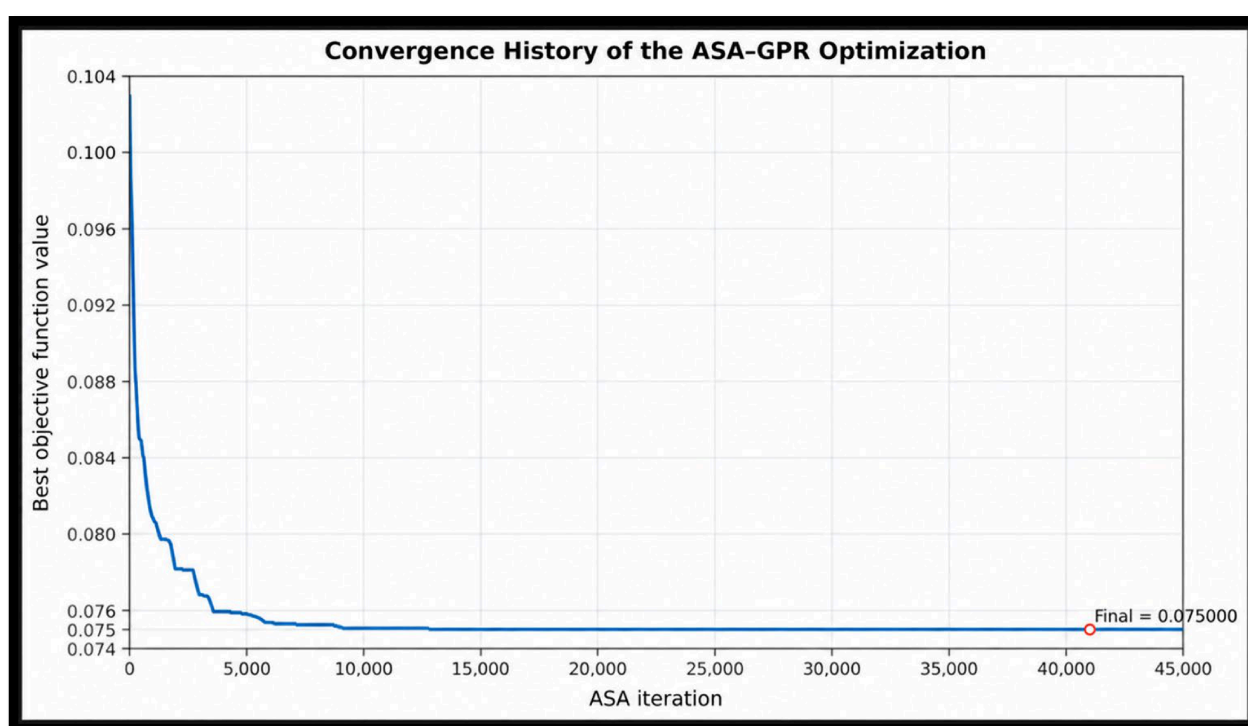


Figure 3. Shows the ASA-GPR linking, which produced multiple iterations before converging to a solution.

Figure 3 illustrates the convergence history of the ASA–GPR optimisation framework implemented in the MATLAB Version R2020b Optimisation Toolbox. The convergence process comprised approximately 41,000 iterations, reflecting the extensive search performed by the inverse optimisation algorithm across the solution space. The Adaptive Simulated Annealing parameters used in the inverse problem solving are mentioned in Table 6 below. The optimisation problem contained 15 decision variables, corresponding to three candidate source locations with five stress periods and 5 decision variables corresponding to 1 dummy source location. The source-flux values were constrained between 0 and 100 mg/day for all 20 decision variables. The lower bound of 0 mg/day prevented physically unrealistic negative source fluxes representing dummy sources, while the upper bound of 100 mg/day defined the maximum permissible release rate. The ASA algorithm generated the initial flux values randomly within these bounds, and any candidate value outside this range during the search was returned to the nearest allowable boundary. These

iterations represent the continuous exploration and refinement of candidate solutions until the optimal objective function value was identified. The optimisation required approximately 1080s to converge using the high-speed CPU supercomputer, demonstrating computational efficiency considering the size and complexity of the dataset, as shown in Figure 3.

Table 6. Adaptive Simulated Annealing (ASA) parameter settings used in the GPR–ASA source-flux reconstruction framework.

ASA Parameters	Values
Number of Decision Variables	20
Flux Bounds	0–100 mg/day
Random Number Seeds	42
Initial Solution	Uniformly sampled within bounds
Initial Temperature	
Cooling Schedule	$T_K = T_0 \exp \left[-8 \sqrt{\frac{k}{4100}} \right]$
Cooling Coefficient	8
Initial Step Fraction	0.2
Proposal Distribution	Normal distribution
Step Adaptation Interval	250 iterations
Target Acceptance Range	0.2–0.5
Minimum Step Fraction	0.005
Acceptance Rule	0.5
Minimum Numerical Temperature	Metropolis Criterion
Maximum Iterations	10^{-12}
Termination Conditions	41,000
Retained Results	Best solution encountered

6.3. Deposit Locations Contribution Analysis

The cumulative contribution analysis indicated that the mineral deposit for Location 3 was the dominant release contributor throughout the reconstruction period. Despite exhibiting temporal variations in flux magnitude between successive stress periods, deposit Location 3 consistently maintained relatively high reconstructed fluxes, resulting in the largest cumulative contributor of isotopes. This indicates that Location 3 exerted the strongest long-term influence on plume evolution and copper concentration splitting behaviour within the groundwater system, making it the significant contributor to groundwater deterioration. In contrast, deposit Location 1 exhibited the highest reconstructed source flux during the initial stress period but declined considerably during subsequent stress periods. Consequently, although deposit Location 1 played a significant role during the early stages of isotope release, its overall cumulative contribution was substantially lower than that of deposit Location 3. This behaviour suggested that deposit Location 1 represented an early-stage release whose influence diminished with time rather than a persistent long-term source. Deposit Location 2 displayed an intermediate contribution to the overall release of the isotopic signature of the concealed mineral location. While its reconstructed flux was relatively small during the initial stress period, progressively larger reconstructed fluxes were obtained during the later stress periods, indicating a sustained release behaviour. The cumulative contribution therefore suggests that Loca-

tion 2 represents a continuing release of isotope signatures whose relative importance increases over time and contributes significantly to the long-term persistence of groundwater pollutants. An important outcome of the cumulative source contribution analysis is the successful identification of the dummy candidate, which is Location 4. This location was intentionally assigned an actual flux of zero throughout the forward simulation to represent an inactive candidate mineralised deposit zone and was termed a dummy source. The dummy candidate location was included as a controlled discrimination test to determine whether the inverse optimisation framework could reduce the estimated contribution of an inactive candidate towards zero rather than assigning non-zero fluxes to every possible source location. The near-zero reconstructed flux obtained for the dummy location demonstrates this fundamental capability under the controlled synthetic conditions considered. However, this represents only an initial test of source discrimination. Future studies should evaluate more challenging configurations involving multiple inactive candidates, closely spaced active and inactive sources, hydrogeologically similar candidate positions, heterogeneous hydraulic-conductivity fields, measurement uncertainty, and alternative observation-network designs.

Optimisation result show, ASA–GPR framework correctly reconstructed flux values of essentially zero reflecting that the plausible deposit location was a dummy or not a real location. These negligible reconstructed values confirm that the optimisation framework did not falsely attribute source location to a non-existing deposit. This result demonstrates the robustness of the proposed inverse modelling methodology in discriminating between active deposit locations and non-contributing candidate locations, thereby reducing the likelihood of false-positive source identification. These results are in the preliminary stage of the methodology development. Implementation of this methodology will require further refinement and evaluation for more complex hydrogeochemical scenarios. The dummy candidate location was included as a controlled discrimination test to determine whether the inverse optimisation framework could reduce the estimated contribution of an inactive candidate toward zero rather than assigning non-zero fluxes to every possible source location.

6.4. Comparison Between Observed and Estimated Isotope Signatures in Groundwater

The effectiveness of the proposed decision model framework was further assessed by comparing the synthetically simulated isotopic signatures with the estimated signatures for the optimal deposit locations. Evaluation results were obtained for Cu_{Tot} , ^{63}Cu , and ^{65}Cu at all monitoring wells and observation times as shown in Figure 4 below. Overall, a close correspondence was observed between the measured and estimated concentration responses. This demonstrated that the reconstructed source histories successfully reproduced the isotopic behaviour generated by the numerical groundwater model. The agreement obtained across the three copper species indicates that the surrogate-assisted optimisation framework accurately captured isotope fractionation behaviour represented within the synthetic dataset. Minor discrepancies between observed and estimated concentrations were expected due to the approximation introduced by the surrogate model and the inherent complexity of the inverse concealed mineral deposit identification. Nevertheless, these differences remain sufficiently small and convergence to the optimal solution was achieved. These results demonstrate the robustness and predictive capability of the proposed optimisation framework.

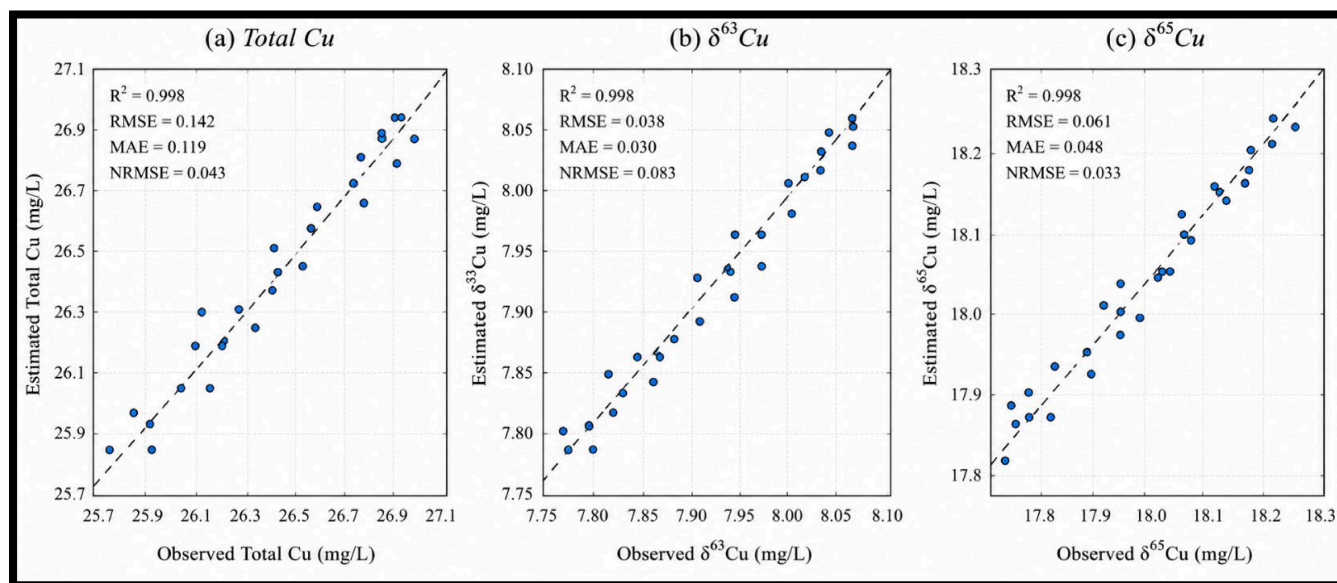


Figure 4. (a) Comparison of the observed and estimated concentrations for Cu_{Tot} ; (b) Comparison of the observed and estimated concentrations for ^{63}Cu ; (c) Comparison of the observed and estimated concentrations for ^{65}Cu respectively.

6.5. Performance Evaluation of the Proposed Methodology

The results presented in the preceding sections collectively demonstrate the capability of the proposed linked Adaptive Simulated Annealing–Gaussian Process Regression (ASA–GPR) framework to reconstruct varying mineral deposit locations. ASA–GPR maintained stable convergence and satisfactory agreement with the observed groundwater concentrations, as discussed earlier. The progressive reduction in the objective function throughout the optimisation process indicated that the Adaptive Simulated Annealing (ASA) algorithm effectively explored the twenty-dimensional search space and consistently improved the candidate deposit location flux estimates until a stable solution was obtained. The smooth convergence towards the optimum objective function suggests the surrogate-assisted optimisation framework efficiently navigated the highly nonlinear inverse problem without exhibiting unstable convergence behaviour or oscillatory search patterns. These results demonstrated the reliability and effectiveness of coupling ASA with a GPR metamodel to solve computationally demanding mineral deposit identification. The reconstructed source flux histories in Figure 2 further demonstrate that the optimisation framework successfully distinguished different mineral locations among the potential zones. Rather than converging towards a uniform loading randomly, the optimisation identified distinctly different release patterns from each potential location of isotopic signatures. An important outcome of the optimisation framework was the successful discrimination between active and inactive candidate deposit mineral locations. The comparison between observed and estimated groundwater concentrations is presented in Figure 4, which further validates the optimised deposit location reconstruction. For the majority of observation wells, the reconstructed groundwater concentrations closely followed the observed responses for Cu_{Tot} , ^{65}Cu and ^{63}Cu , indicating that the optimised source fluxes successfully reproduced the transport behaviour simulated by the numerical groundwater model. Although several observation wells exhibited relatively larger deviations, particularly at higher concentration levels, the overall agreement demonstrates that the optimised solution adequately captured the dominant spatial and temporal concentration trends. These results indicate that the optimisation framework was capable of reconstructing deposit location histories that remained consistent with observed groundwater responses despite the inherent

non-uniqueness associated with groundwater inverse problems. By replacing repeated MODFLOW–MT3DMS simulations with rapid GPR predictions, the computational cost of the inverse modelling process was substantially reduced and the time required was drastically reduced. Thus, the linking of the metamodel with the decision model as ASA–GPR provided computational efficiency, which enabled extensive global exploration of the solution space. Without this concept, it would have been computationally impractical to solve the inverse-based problem using direct numerical simulation alone [53,54]. Unlike conventional groundwater source identification approaches that rely solely on concentrations, the proposed methodology simultaneously reconstructed Cu_{Tot} , ^{65}Cu and ^{63}Cu responses during the optimisation process, which is a new methodology related to mineral exploration. The successful convergence of the optimisation, together with the accurate reconstruction of deposit location histories, cumulative source contributions and correct rejection of the dummy location, demonstrated that incorporating isotopic information provided additional constraints for copper concealed mineral identification without compromising optimisation stability. Overall, the proposed ASA–GPR framework provided an efficient, computationally robust, and scientifically reliable methodology for solving complex inverse problems by optimally identifying multiple unknown mineral deposit locations. Furthermore, the successful identification of active mineral locations while simultaneously rejecting non-actual candidate locations demonstrated the applicability of the proposed methodology for detection of potential mineral zones using isotope signatures.

7. Computational Efficiency

One of the primary challenges associated with groundwater inverse modelling is the substantial computational effort required to repeatedly execute numerical flow and transport simulations during the optimisation process [38,55]. In a conventional simulation–optimisation (SO) framework, each candidate solution generated by the optimisation algorithm required a complete numerical model execution. As the number of decision variables, observation points and simulation periods increase, the computational burden becomes prohibitive, often limiting the practical application of inverse modelling techniques to real-world problems [56]. In the present study, this limitation was addressed through the development of a surrogate-assisted optimisation framework. The GPR metamodel was employed as a computationally efficient replacement for the MODFLOW–MT3DMS simulator during optimisation. Consequently, thousands of candidate source configurations were evaluated within a fraction of the computational time required by a conventional simulation-based optimisation framework. The computational advantages of the surrogate-assisted framework were particularly evident during the Adaptive Simulated Annealing (ASA) optimisation process. The optimisation algorithm required numerous evaluations of the objective function while searching for the best solution. Direct coupling of ASA with the numerical simulator would have resulted in extremely long execution times due to the repeated solution of groundwater flow and transport equations. By replacing the numerical simulator with the trained GPR surrogate, the optimisation process became computationally feasible while maintaining a high degree of predictive accuracy. In addition to computational efficiency, the surrogate-assisted approach provided several practical advantages. The framework enabled rapid exploration of alternative source configurations, facilitated sensitivity analysis and allowed extensive optimisation experiments to be performed within reasonable computational times. The ability to evaluate a large number of candidate solutions improved the likelihood of identifying and reconstructing flux histories that closely reproduced the observed groundwater responses. The optimisation results further demonstrated the effectiveness of combining machine learning and inverse modelling methodologies within a unified groundwater investigation framework. The

proposed inverse modelling framework was developed by coupling multiple modelling engines within a unified computational workflow. Approximately 2000 numerical simulations were automatically executed using a macro-based command engine incorporating a logic-driven architecture and file-path control system. Owing to the computational demands of the numerical simulator, generation of the training dataset required several weeks. However, once the input–output database had been established, the GPR surrogate model was trained within only a few minutes. The trained GPR was subsequently coupled with the Adaptive Simulated Annealing (ASA) optimiser, where convergence was achieved in approximately 1080s after evaluating nearly 41,000 iterations. This substantial reduction in computational time demonstrates the efficiency of the surrogate-assisted optimisation framework, enabling rapid evaluation of thousands of candidate solutions without repeatedly executing the computationally expensive numerical simulator. The high predictive accuracy achieved by the GPR metamodel ensured that surrogate predictions remained representative of the underlying numerical simulator, while the global search characteristics of ASA enabled effective exploration of complex and highly nonlinear solution spaces. The GPR-assisted ASA optimisation was executed on a high-performance computing cluster using one compute node equipped with 32 CPU cores and 128 GB RAM under a Linux operating system. Approximately 41,000 optimisation iterations were completed in 1080s. The reported runtime represents only the surrogate-assisted optimisation stage and excludes the offline generation of MODFLOW–MT3DMS simulations and GPR training. Because an equivalent direct ASA–MODFLOW–MT3DMS optimisation direct numerical coupling was not performed, no quantitative speed-up factor is claimed, and this is beyond the scope of the current work.

From a mineral exploration perspective, the proposed methodology could potentially provide a novel approach for utilising groundwater responses as indirect indicators of concealed mineralisation. By reconstructing source flux histories from groundwater concentration and isotope observations, the framework offers a mechanism for investigating potential subsurface source zones without requiring direct exposure of the mineralised body. The integration of isotope-informed transport modelling, surrogate modelling and optimisation therefore extends traditional groundwater source identification concepts into the domain of groundwater-assisted mineral exploration. Despite the promising results obtained in this study, several limitations should be acknowledged. The present investigation was conducted using synthetic datasets generated from a numerical simulation framework and therefore the performance of the methodology under field conditions remains to be evaluated. In addition, uncertainty associated with hydrogeological parameters, monitoring network design and isotope transport assumptions may influence source reconstruction accuracy. Future research should focus on incorporating uncertainty analysis, field-scale validation and multi-objective optimisation strategies to further enhance the robustness of the proposed framework. Overall, the results demonstrated that surrogate-assisted optimisation provides an effective means of overcoming the computational limitations associated with conventional linked simulation–optimisation (SO) approaches. The methodology therefore represents a promising framework for future groundwater and concealed mineralisation investigations.

Limitations and Field Applicability

The proposed study represents proof of concept, to the best of our knowledge, for the proposed new methodology development. Although the results have demonstrated the potential of isotope-informed inverse modelling for identifying candidate mineralised locations and reconstructing their time-varying source fluxes, several challenges must be addressed before field implementation. These include spatial heterogeneity and uncertainty

in various parameters such as hydraulic conductivity, porosity, dispersivity and recharge; uncertainty in mineralisation geometry and source-release mechanisms; and structural uncertainty in the groundwater-flow and transport model. There are various ways in which natural copper-isotope behaviour may also be influenced by mineral dissolution and precipitation, redox reactions, aqueous complexation, kinetic effects and fractionation processes that are more complex than the simplified differential adsorption mechanism represented in the present model. Although the study presents a synthetic proof of concept which has been utilised for the methodology development, field application would additionally need to be considered together with analytical errors in isotope measurements. Therefore, the present findings should not be interpreted as demonstrating immediate field capability. Rather, they establish that the mathematical inverse-identification concept is sufficiently promising to warrant progressive testing under increasingly realistic geological, hydrogeological and geochemical conditions, followed by validation using field observations.

8. Conclusions

This study presents a proof-of-concept using a surrogate-assisted simulation–optimisation framework for the development of a methodology for inverse identification of concealed copper-bearing mineral deposits using copper isotope signatures in groundwater. The proposed methodology was adapted from optimal groundwater contaminant source characterisation, which integrated groundwater flow and transport simulation. Further, the proposed methodology integrated GPR-based surrogate modelling and the Adaptive Simulated Annealing (ASA) optimisation algorithm to identify the optimal locations and temporal characteristics of concealed mineralisation while substantially reducing computational requirements. The GPR surrogate model accurately reproduced the responses of the MODFLOW–MT3DMS numerical simulator, enabling efficient optimisation without repeated execution of the computationally intensive numerical groundwater transport model. The ASA–GPR framework successfully reconstructed the temporal isotopic signature histories originating from plausible copper deposits, correctly identified the active deposit locations and rejected the intentionally introduced dummy (non-mineralised) location, demonstrating its capability to distinguish genuine mineralisation targets from false candidates. A key innovation of this work is the incorporation of the individual copper isotopes (^{63}Cu and ^{65}Cu) together with the derived $\delta^{65}\text{Cu}$ signatures within a unified inverse modelling framework. Unlike conventional groundwater investigations that consider dissolved copper as a single conservative species, the present methodology independently simulates the transport behaviour of both isotopes. The differences in their spatial and temporal responses provided additional hydrogeochemical information that can be exploited to identify concealed copper-bearing mineralised systems. The successful reconstruction of the synthetic mineralised locations demonstrated that isotope partitioning provided meaningful exploration-related information beyond total dissolved copper concentrations alone. Consequently, groundwater copper isotope signatures have the potential to serve as quantitative indicators for detecting concealed mineralisation beneath soil cover. The evaluation of the proposed methodology is based on synthetic groundwater measurements generated under multiple hypothetical mineralisation scenarios and therefore represents a preliminary validation of the proposed methodology. Although the hydrogeological system was simplified for this proof-of-concept evaluation, the results demonstrated the potential feasibility of extending groundwater inverse source characterisation concepts to groundwater-assisted mineral exploration using isotope signatures.

The current model is not a site-specific model to reproduce a specific porphyry, volcanogenic massive sulphide, or sediment-hosted copper deposit. The domain dimensions, hydraulic properties, boundary conditions, and transport parameters were selected to gen-

erate physically plausible groundwater flow and solute-transport behaviour, while retaining sufficient experimental control to assess source identification and flux-reconstruction accuracy. In real copper-bearing environments, plume development may be influenced by deposit-specific lithology, structural controls, fracture networks, hydraulic connectivity, mineralogy, weathering profiles, and groundwater geochemistry. Consequently, field implementation would require a site-specific hydrogeological and geochemical conceptual model supported by geological mapping, hydraulic testing, groundwater monitoring, isotope measurements, and appropriate model calibration and validation.

Transferability to real aquifer systems may be affected by uncertainty in hydraulic properties, geological heterogeneity, boundary conditions, recharge, source geometry, isotope fractionation parameters, monitoring errors, and geochemical processes not represented in the current model. Although recent machine learning-based mineral-prospectivity approaches identify exploration targets by integrating geological and geochemical evidence layers [56], the present study investigates concealed mineralisation indirectly through isotope-specific groundwater transport signals and inverse source identification. Future applications should therefore retrain and validate the surrogate using site-specific hydrogeological, geochemical, and isotopic observations, together with uncertainty and sensitivity analyses. Independent field validation would be required before the framework could be used for operational mineral exploration.

The novel contribution of this work is the use of isotope signatures to identify concealed mineral locations, and the methodology has been successful in determining plausible copper deposits utilising the inverse model. The principal scientific contribution is the transfer and extension of inverse groundwater source-identification concepts towards a new application in concealed mineral exploration, with isotope-specific groundwater information providing an additional source-discrimination signal.

Future research should focus on the incorporation of more complex hydrogeological, geochemical and geophysical scenarios. This should be followed by field-scale validation using measured groundwater isotope datasets from known mineralised terrains, incorporation of geological heterogeneity and more comprehensive isotope fractionation processes, uncertainty quantification, and multi-objective optimisation. These developments will further improve the robustness and applicability of the proposed methodology as a practical decision-support tool for groundwater-assisted exploration of concealed copper deposits and other critical mineral resources.

Author Contributions: Conceptualisation, B.D. and R.M.; methodology, B.D. and R.M.; software, B.D. and R.M.; validation, B.D., I.S. and R.M.; formal analysis, B.D. and R.M.; investigation, B.D., I.S. and R.M.; resources, B.D. and R.M.; data curation, R.M. and B.D.; writing—original draft preparation, R.M. and B.D.; writing—review and editing, B.D., R.M. and I.S.; visualisation, B.D., R.M. and I.S.; supervision, B.D. and I.S.; project administration, B.D.; funding acquisition, B.D. and I.S. All authors have read and agreed to the published version of the manuscript.

Funding: The Funding for this project was provided by James Cook University (50%) and the Critical Mineral Trailblazer Programme (50%).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors gratefully acknowledge the Resources Technology and Critical Minerals (RTCM) Trailblazer Programme for supporting this research. The Trailblazer Programme is funded by the Commonwealth Government of Australia and various mining-related industries. This research is particularly related to the project titled: Critical Mineral Trailblazer Programme Emerging Technologies for Advanced Geochemical Exploration Undercover (Bithin Datta, Alex

McCoy-West, Ioan Sanislav, Mahmood Sadat Noori, and Helen McCoy-West). All the above-listed project participants have provided suggestions at various stages of this continuing project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mahan, B.; Mathur, R.; Sanislav, I.; Rea, P.; Dirks, P. Cu isotopes in groundwater for hydrogeochemical mineral exploration: A case study using the world-class Mount Isa Cu–Pb–Zn deposit (Australia). *Appl. Geochem.* **2023**, *148*, 105519. [\[CrossRef\]](#)
2. Sullivan, K.V.; Kidder, J.A.; Junqueira, T.P.; Vanhaecke, F.; Leybourne, M.I. Emerging applications of high-precision Cu isotopic analysis by MC-ICP-MS. *Sci. Total Environ.* **2022**, *838*, 156084. [\[CrossRef\]](#)
3. Ahmadi, H.; Kilanehei, F.; Nazari-Sharabian, M. Impact of pumping rate on contaminant transport in groundwater—A numerical study. *Hydrology* **2021**, *8*, 103. [\[CrossRef\]](#)
4. Amir Abdollahian, M. *Development of Integrated Methodologies for Optimal Monitoring and Source Characterization in Contaminated Groundwater Systems Under Uncertainty*; James Cook University: Douglas, QLD, Australia, 2016.
5. Amirabdollahian, M.; Datta, B. Application of Simulated Annealing and Adaptive Simulated Annealing in Search for Efficient Optimal. In *Computation Optimization in Engineering. Paradigms and Applications*; IntechOpen: Rijeka, Croatia, 2017; pp. 133–150.
6. Aral, M.M.; Guan, J.; Maslia, M.L. Identification of contaminant source location and release history in aquifers. *J. Hydrol. Eng.* **2001**, *6*, 225–234. [\[CrossRef\]](#)
7. Asher, M.J.; Croke, B.F.; Jakeman, A.J.; Peeters, L.J. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* **2015**, *51*, 5957–5973. [\[CrossRef\]](#)
8. Atmadja, J.; Bagtzoglou, A.C. Pollution source identification in heterogeneous porous media. *Water Resour. Res.* **2001**, *37*, 2113–2125. [\[CrossRef\]](#)
9. Baijiali, W.; Clark, I.D.; Fritz, P. The artesian thermal groundwaters of northern Jordan: Insights into their recharge history and age. *J. Hydrol.* **1997**, *192*, 355–382. [\[CrossRef\]](#)
10. Battiti, R.; Brunato, M.; Mascia, F. *Reactive Search and Intelligent Optimization*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; Volume 45.
11. Breukelen, B.M.V.; Prommer, H. Beyond the Rayleigh equation: Reactive transport modeling of isotope fractionation effects to improve quantification of biodegradation. *Environ. Sci. Technol.* **2008**, *42*, 2457–2463. [\[CrossRef\]](#)
12. Carrera, J.; Neuman, S.P. Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. *Water Resour. Res.* **1986**, *22*, 228–242. [\[CrossRef\]](#)
13. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [\[CrossRef\]](#)
14. Chang, Z.; Guo, Z.; Chen, K.; Wang, Z.; Zhan, Y.; Lu, W.; Zheng, C. A comparison of inversion methods for surrogate-based groundwater contamination source identification with varying degrees of model complexity. *Water Resour. Res.* **2024**, *60*, e2023WR036051. [\[CrossRef\]](#)
15. Datta, B. Discussion of “identification of contaminant source location and release history in aquifers” by Mustafa M. Aral, Jiabao Guan, and Morris L. Maslia. *J. Hydrol. Eng.* **2002**, *7*, 399–400. [\[CrossRef\]](#)
16. Datta, B. Preliminary hydrogeologic modeling and optimal monitoring network design for a contaminated abandoned mine site area: Application of developed monitoring network design software. *J. Water Resour. Prot.* **2016**, *8*, 46–64.
17. Datta, B.; Chakrabarty, D.; Dhar, A. Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *J. Hydro-Environ. Res.* **2011**, *5*, 25–36. [\[CrossRef\]](#)
18. Saremi, M.; Maghsoudi, A.; Yousefi, M.; Hezarkhani, A. A confidence index-based ensemble technique for mineral prospectivity mapping: Incorporation of supervised and unsupervised learning approaches. *Earth Sci. Inform.* **2026**, *19*, 109. [\[CrossRef\]](#)
19. Datta, B.; Amirabdollahian, M.; Zuo, R.; Prakash, O. Groundwater contamination plume delineation using local singularity mapping technique. *GEOMATE J.* **2016**, *11*, 2435–2441.
20. Esfahani, H.K.; Datta, B. Fractal singularity-based multiobjective monitoring networks for reactive species contaminant source characterization. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018021. [\[CrossRef\]](#)
21. Mostafaei, K.; Hezarkhani, A.; Maleki, S.; Zamani Ahmad Mahmoudi, M.; Khalilidermani, M.; Knez, D. High-grade zone estimation using the SVM and BPPN algorithms in Chah Firuzeh porphyry copper deposit. *Int. J. Min. Geo-Eng.* **2026**, *60*, 149–159.
22. Mostafaei, K.; Kianpour, M.N.; Yousefi, M.; Saleki, M. Improving the results of the fractal model of geochemical mineralization probability index using the gray wolf algorithm on the stream sediments data of Sarduiyeh-Baft Area. *J. Min. Environ.* **2025**, *16*, 1165–1178.

23. Datta, B.; Prakash, O.; Amirabdollahian, M.; Jha, M.K.; Esfahani, H.K.; Hayford, M.S.; Sreenivasulu, C.; Naidu, R. Recent developments in identification of unknown contamination sources and monitoring network design for contaminated groundwater systems. In Proceedings of the 7th International Contaminated Site Remediation Conference, Melbourne, VIC, Australia, 10–14 September 2015.
24. Esfahani, H.K.; Datta, B. Linked optimal reactive contaminant source characterization in contaminated mine sites: Case study. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04016061. [\[CrossRef\]](#)
25. Esfahani, H.K.; Heggie, A.; Datta, B. Integrated Sequential Groundwater Contaminant Source Characterization and Pareto-Optimal Monitoring Network Design Application for a Contaminated Aquifer Site. *J. Water Resour. Prot.* **2022**, *14*, 542–570.
26. Forrester, A.; Sobester, A.; Keane, A. *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
27. Gao, K.; Cheng, Z.; Song, Y.; Yin, S.; Chen, Y. Machine learning–driven surrogate model development for geotechnical numerical simulation. *Geotech. Res.* **2025**, *12*, 71–84. [\[CrossRef\]](#)
28. Harbaugh, A.W. *MODFLOW-2005, the US Geological Survey Modular Ground-Water Model: The Ground-Water Flow Process*; US Department of the Interior, US Geological Survey: Reston, VA, USA, 2005; Volume 6.
29. Hayford, M.S.; Datta, B. Source characterization of multiple reactive species at an abandoned mine site using a groundwater numerical simulation model and optimization models. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4776. [\[CrossRef\]](#)
30. Hazrati Yadkori, S.; Datta, B. Adaptive surrogate model based optimization (ASMBO) for unknown groundwater contaminant source characterizations using self-organizing maps. *J. Water Resour. Prot.* **2017**, *9*, 193–214.
31. Ingber, L. Very fast simulated re-annealing. *Math. Comput. Model.* **1989**, *12*, 967–973. [\[CrossRef\]](#)
32. Ingber, L. Adaptive simulated annealing (ASA). In *Global Optimization C-Code*; Caltech Alumni Association: Pasadena, CA, USA, 1993.
33. Jha, M.; Datta, B. Three-dimensional groundwater contamination source identification using adaptive simulated annealing. *J. Hydrol. Eng.* **2013**, *18*, 307–317. [\[CrossRef\]](#)
34. Jin, B.; Xu, X. Predictive Modeling of Peanut Oil Prices Utilizing a Gaussian Process Regression-Based Machine Learning Framework. *Adv. Data Sci. Adapt. Anal.* **2025**, *17*, 1. [\[CrossRef\]](#)
35. Lal, A.; Datta, B. Multiple objective management strategies for coastal aquifers utilizing new surrogate models. *GEOMATE J.* **2018**, *15*, 79–85. [\[CrossRef\]](#)
36. Lal, A.; Datta, B. Optimal pumping strategies for the management of coastal groundwater resources: Application of Gaussian Process Regression metamodel-based simulation-optimization methodology. *ISH J. Hydraul. Eng.* **2021**, *27*, 136–145. [\[CrossRef\]](#)
37. Li, M.; Wang, G.; Qian, L.; Li, X.; Ma, Z. ASS-GPR: Adaptive sequential sampling method based on Gaussian process regression for reliability analysis of complex geotechnical engineering. *Int. J. Geomech.* **2021**, *21*, 04021192. [\[CrossRef\]](#)
38. Mahar, P.S.; Datta, B. Optimal identification of ground-water pollution sources and parameter estimation. *J. Water Resour. Plan. Manag.* **2001**, *127*, 20–29. [\[CrossRef\]](#)
39. Marrel, A.; Perot, N.; Mottet, C. Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 959–974. [\[CrossRef\]](#)
40. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE* **2015**, *58*, 1763–1785. [\[CrossRef\]](#)
41. Neupane, R.; Datta, B. Optimal characterization of unknown multispecies reactive contamination sources in an aquifer. *J. Hydrol. Eng.* **2021**, *26*, 04021035. [\[CrossRef\]](#)
42. Ostertag, J.; Petersen, A.; Schmidt, O.; Brüggemann, R. Application of an optimization tool for inverse modelling of thin-film silicon solar cells. *Energy Procedia* **2011**, *10*, 266–270. [\[CrossRef\]](#)
43. Prakash, O.; Datta, B. Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations. *Environ. Monit. Assess.* **2013**, *185*, 5611–5626. [\[CrossRef\]](#)
44. Prakash, O.; Datta, B. Optimal characterization of pollutant sources in contaminated aquifers by integrating sequential-monitoring-network design and source identification: Methodology and an application in Australia. *Hydrogeol. J.* **2015**, *23*, 1089–1107. [\[CrossRef\]](#)
45. Rasmussen, C.E.; Nickisch, H. Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.* **2010**, *11*, 3011–3015.
46. Razavi, S.; Tolson, B.A.; Burn, D.H. Numerical assessment of metamodeling strategies in computationally intensive optimization. *Environ. Model. Softw.* **2012**, *34*, 67–86. [\[CrossRef\]](#)
47. Razavi, S.; Tolson, B.A.; Burn, D.H. Review of surrogate modeling in water resources. *Water Resour. Res.* **2012**, *48*, W07401. [\[CrossRef\]](#)
48. Roy, D.K.; Datta, B. An ensemble meta-modelling approach using the Dempster-Shafer theory of evidence for developing saltwater intrusion management strategies in coastal aquifers. *Water Resour. Manag.* **2019**, *33*, 775–795. [\[CrossRef\]](#)
49. Roy, D.K.; Datta, B. Saltwater intrusion prediction in coastal aquifers utilizing a weighted-average heterogeneous ensemble of prediction models based on Dempster-Shafer theory of evidence. *Hydrol. Sci. J.* **2020**, *65*, 1555–1567. [\[CrossRef\]](#)

50. Saves, P.; Palar, P.S.; Robani, M.D.; Verstaavel, N.; Garouani, M.; Aligon, J.; Gaudou, B.; Shimoyama, K.; Morlier, J. Surrogate modeling and explainable artificial intelligence for complex systems: A workflow for automated simulation exploration. *arXiv* **2025**, arXiv:2510.16742.
51. Van Breukelen, B.M.; Rolle, M. Transverse hydrodynamic dispersion effects on isotope signals in groundwater chlorinated solvents' plumes. *Environ. Sci. Technol.* **2012**, *46*, 7700–7708. [[CrossRef](#)]
52. Willmott, C.J. On the validation of models. *Phys. Geogr.* **1981**, *2*, 184–194. [[CrossRef](#)]
53. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
54. Xiao, J.; Endo, K.; Muramatsu, M.; Nomura, R.; Moriguchi, S.; Terada, K. Application of factorization machine with quantum annealing to hyperparameter optimization and metamodel-based optimization in granular flow simulations. *Int. J. Numer. Anal. Methods Geomech.* **2024**, *48*, 3432–3451. [[CrossRef](#)]
55. Zhao, Y.; Lu, W.; Xiao, C. A Kriging surrogate model coupled in simulation–optimization approach for identifying release history of groundwater sources. *J. Contam. Hydrol.* **2016**, *185–186*, 51–60. [[CrossRef](#)]
56. Zheng, C.; Wang, P.P. *MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide*; Technical Report; U.S. Army Corps of Engineers, Engineer Research and Development Center: Vicksburg, MI, USA, 1999.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.