## Dear editor and reviewers:

We sincerely thank the Editor for the prompt handling of our manuscript and for the opportunity to revise our submission, “A Multi-chain Surrogate-assisted Hybrid Optimization Framework for Joint Identification of Groundwater Contaminant Sources and Hydrogeological Parameters” (EGUSPHERE-2025-6140). **Please note that the author name previously shown as “Yang Xu” has been corrected to “Xu Yang” in the revised manuscript.** This is a name-order correction only and does not involve any change in authorship or contributions.

We are also very grateful to the Reviewers for their careful evaluation and constructive comments, which have helped us improve the manuscript. In response to the reviewers’ comments, we have thoroughly revised the manuscript and made the necessary changes. We believe that this revision version has significantly improved the clarity and overall quality of the manuscript. Our detailed point-by-point responses to the comments of each reviewer are provided below:

## Comments and Reponses regarding Anonymous Referee 1 (RC1)

### General comments:

A new efficient calculation procedure was developed for estimating the input parameters for groundwater model simulation to explore the locations of contamination sources. The authors demonstrated the step-by-step performance evaluation of the procedure through hypothetical field conditions to near real-world situations. Although the manuscript contains several points of uncertainty or illegibility, as noted below, the overall structure makes it possible to understand the authors' great work as described above.

・In Subchapter 4.3, the authors described Case 3 as a real-world groundwater problem; however, insufficient information was provided for the regional groundwater flow system. At a minimum, "aquifer structure (number of layers)," "number of grid cells in the depth direction," "water input to the aquifer (i.e., presence or absence of local precipitation)," and "reasonable information to define the regions as four; for example, geological information except the northern and southern boundary areas as mentioned" would make it easier to agree with the process by which the model was discretized from the actual groundwater flow field to the structure necessary to achieve the objectives.

・The title in Chapter 3 appears irregular with the other chapter titles because it includes the specific names of algorithms (SA-CSA-TS). Abstracting the title, for example, to "Overview of the Developed Algorithms," would make the overall context of the article easier to understand.

・The manuscript contained many figures, which reduced readability. It would be more readable to consolidate Figures 14, 15, and 17, which are similar; a multi-panel figure with branch numbers 14a, 14b, and 14c can be used. This approach could also be applied to pairs of figures that are "of the same type but depict different experimental cases," such as Figures 12 and 13 or Figures 10 and 11. If possible, the authors should consider categorizing the figures into those that are strongly relevant to the derivation of conclusions and those that are not; the latter should be moved to the supplement.

・Limitations of this approach at the present time must be addressed. The estimation of three-dimensional flow and dispersion of contamination, which is one of the major purposes of groundwater flow modeling, is not captured by this approach.

・In this manuscript, a model with a simplified depth layer is constructed. If the layer (as grids) is increased for depth, would the computation time remain within a practical range? I believe that the importance of this question is related to the applicability of the developed method beyond the estimation of contamination locations under the simplified aquifer system (mentioned in the last line of conclusion).

**General responses:**

We would like to express our profound gratitude for your review and the constructive feedback provided on our manuscript. Your insights regarding the hydrogeological characterization of the real-world case, the conceptual limitations of the current algorithm, and the overall presentation logic have been invaluable in elevating the scientific rigor of our work. We acknowledge that the original manuscript required more depth in justifying the transition from complex field conditions to simplified numerical models. We have made a major revision. The detailed responses are listed following.

**Comment 1:** In Subchapter 4.3, the authors described Case 3 as a real-world groundwater problem; however, insufficient information was provided for the regional groundwater flow system. At a minimum, "aquifer structure (number of layers)," "number of grid cells in the depth direction," "water input to the aquifer (i.e., presence or absence of local precipitation)," and "reasonable information to define the regions as four; for example, geological information except the northern and southern boundary areas as mentioned" would make it easier to agree with the process by which the model was discretized from the actual groundwater flow field to the structure necessary to achieve the objectives.

**Response 1:** Thank you for this constructive comment. We agree that providing detailed hydrogeological characterization is crucial for validating the model discretization process in Case 3. In the revised manuscript, we have significantly expanded **Subchapter 3.3 (original Subchapter 4.3)** to include a comprehensive description of the regional groundwater flow system. Specifically, we have addressed your concerns as follows:

1. **Aquifer structure and layers:** we have clarified that the aquifer is conceptualized as a single-layer unconfined system consisting of weathered and fractured granite (average thickness of 30 m), overlying impermeable fresh granite bedrock.

2. **Grid discretization:** The model uses a two-dimensional single-layer structured grid with a cell size of 30 m × 30 m. Since it is modeled as a 2D flow system based on the horizontal dominance of the flow field, there is one grid cell in the depth direction

representing the saturated thickness.

3. **Precipitation:** We have added information regarding the recharge source. Groundwater recharge is driven by vertical infiltration of local precipitation, with an average annual precipitation of 650 mm and a recharge coefficient of 0.12.

4. **Geological Justification for Partitioning:** We have provided the geological basis for dividing the domain into four zones: Zone I (alluvial sand/gravel), Zones II and III (highly and moderately weathered granite, respectively), and Zone IV (a localized tectonic fracture zone). This partitioning is based on exploration-stage geological reports.

**The relevant descriptions in Subchapter 3.3 now read as follows:**

This case study is designed as a realistic numerical experiment based on the hydrogeological conditions of a mining area in Henan Province, China. The study area covers approximately $2.67 \times 3$ km. According to exploration-stage geological archives and field investigations, the aquifer is conceptualized as a single-layer unconfined system composed mainly of weathered and fractured granite, with an average saturated thickness of about 30 m. The underlying fresh granite is considered impermeable and therefore forms the basal boundary of the model. The groundwater flow system is represented by a two-dimensional single-layer numerical model. In plan view, the model domain is discretized using a structured grid with a uniform cell size of $30m \times 30m$, and the irregular outer boundary is represented by active and inactive cells, as shown in Fig. 8. The rivers along the western and eastern margins are treated as constant-head boundaries, whereas the northern and southern margins are specified as no-flow boundaries because they are bounded by relatively intact, low-permeability fresh granite. Groundwater recharge occurs primarily through vertical infiltration of precipitation and is represented using an average annual precipitation of 650 mm and a recharge coefficient of 0.12. To capture spatial heterogeneity, the aquifer is divided into four hydraulic-conductivity zones based on the exploration-stage geological archives: Zone I corresponds to alluvial sand and gravel near the riverbanks, Zones II and III represent highly weathered and moderately weathered granite, respectively, and Zone IV represents a localized tectonic fracture zone. The main hydrogeological parameters adopted in the model are summarized in Table 2.

We once again thank the reviewer for this suggestion, which have been instrumental in improving the overall clarity of our work.

**Comment 2:** The title in Chapter 3 appears irregular with the other chapter titles because it includes the specific names of algorithms (SA-CSA-TS). Abstracting the title, for example, to "Overview of the Developed Algorithms," would make the overall context of the article easier to understand.

**Response 2:** Thank you for this helpful suggestion. We agree that, in the original manuscript, the title of Section 3 was overly specific and not fully consistent with the other section headings. In the revised manuscript, this issue has been addressed through the broader structural reorganization made in response to Comment 1 of the RC2. Specifically, the methodological content originally presented in Section 3 has been moved into Section 2, and the overview subsection is now presented as Section 2.3, entitled "Overview of the proposed algorithm." Section 3 has been reassigned to the case studies. We believe that this revision improves the consistency of the section titles and makes the overall structure of the manuscript clearer for readers."

**Comment 3:** The manuscript contained many figures, which reduced readability. It would be more readable to consolidate Figures 14, 15, and 17, which are similar; a multi-panel figure with branch numbers 14a, 14b, and 14c can be used. This approach could also be applied to pairs of figures that are "of the same type but depict different experimental cases," such as Figures 12 and 13 or Figures 10 and 11. If possible, the authors should consider categorizing the figures into those that are strongly relevant to the derivation of conclusions and those that are not; the latter should be moved to the supplement.

**Response 3:** We appreciate the reviewer's suggestion regarding the presentation of our figures. We agree that the large number of figures previously hindered the manuscript's readability and that consolidating similar results facilitates easier comparison. Following your recommendation, we have performed a comprehensive reorganization of the figures, specifically:

1. Original **Figures 10** and **11** have been integrated into a single multi-panel figure (now **Figure 10** in the revised version).
2. Original **Figures 12** and **13** have been merged into **Figure 11**.
3. Original **Figures 14**, **15**, and **17**, which share similar themes, have been combined into a unified **Figure 12**.
4. Original **Figures 16** and **18** have been merged into **Figure 13**.

**Comment 4:** Limitations of this approach at the present time must be addressed. The estimation of three-dimensional flow and dispersion of contamination, which is one of the major purposes of groundwater flow modeling, is not captured by this approach.

**Response 4:** We thank the reviewer for pointing out this limitation. We fully agree that three-dimensional (3D) flow and hydrodynamic dispersion are fundamental components of groundwater modeling and that their omission in the current study represents a significant boundary of our approach. We acknowledge that the manuscript lacked a transparent discussion of these limitations. Therefore, we have incorporated **Subchapter 6.4 (Limitations)** in the **Chapter 6 (Discussion).** In this new section, we explicitly state that the current study focuses on 2D systems and that capturing 3D migration and dispersion patterns remains a boundary of our present work. Furthermore, we clarify that while the search strategy of SA-CSA-TS is intrinsically model-agnostic, meaning it is guided by response discrepancies at monitoring locations rather than being tied to a specific model dimensionality, its practical performance in complex 3D transport dynamics still requires further investigation. We believe these additions provide the necessary transparency and context for readers to understand both the current scope and the future potential of our study.

**The relevant descriptions in Subchapter 6.4 now read as follows:**

Despite the promising performance and robustness of SA-CSA-TS, some limitations should be further discussed. First, this study evaluates the proposed algorithm only using two-dimensional groundwater systems. The search strategy of SA-CSA-TS, however, is guided by the difference between simulated and observed responses at monitoring locations, rather than by the assumption tied to a specific groundwater model dimensionality. This gives the SA-CSA-TS potential for extension to three-dimensional groundwater models. Nevertheless, its applicability to three-dimensional flow and hydrodynamic dispersion systems has not yet been demonstrated in this study. Furthermore, an increase in vertical resolution (depth layers) not only raises the computational cost of groundwater simulation but also poses a challenge to the predictive fidelity of the surrogate model in complex cases, which should be confirmed through further investigation. Second, while the robustness analysis demonstrated resilience against Gaussian noise, real-world field conditions often involve more complex uncertainties. These include sparse monitoring networks, systematic measurement biases, and structural model errors. Therefore, future work should focus on testing SA-CSA-TS in three-dimensional systems and under combined uncertainty sources, in order to establish its robustness and reliability for complex groundwater inverse problems.

**Comment 5:** In this manuscript, a model with a simplified depth layer is constructed. If the layer (as grids) is increased for depth, would the computation time remain within a practical range? I believe that the importance of this question is related to the applicability of the developed method beyond the estimation of contamination locations under the simplified aquifer system (mentioned in the last line of conclusion).

**Response 5:** Thank you for this valuable comment. We agree that the practical applicability of the proposed framework depends on its scalability to more realistic multilayer aquifer systems. In the current manuscript, the validation cases, including the most realistic case, are still based on a simplified single-layer aquifer representation. Therefore, the present study does not directly quantify the wall-clock time for a fully multilayer discretization. Nevertheless, our current results provide useful evidence regarding computational scalability. Figure R1 below demonstrates that the total runtime is overwhelmingly dominated by the high-fidelity simulations rather than the algorithm's internal overhead. Since the SA-CSA-TS framework is model-agnostic and reduces the required number of simulations by 85–88% compared to GA and CSA, its efficiency advantage is expected to become even more pronounced as the forward model becomes more computationally intensive (e.g., in 3D systems). We have clarified this potential in **Subchapter 6.1**, noting that the framework's primary strength lies in minimizing expensive evaluations. Meanwhile, we have also updated **Subchapter 6.4** (Limitations) and **Chapter 7** (Conclusions) to acknowledge the challenges of maintaining surrogate accuracy and managing simulation costs when transitioning to higher-dimensional multilayer systems, which remains a key focus for our future research.

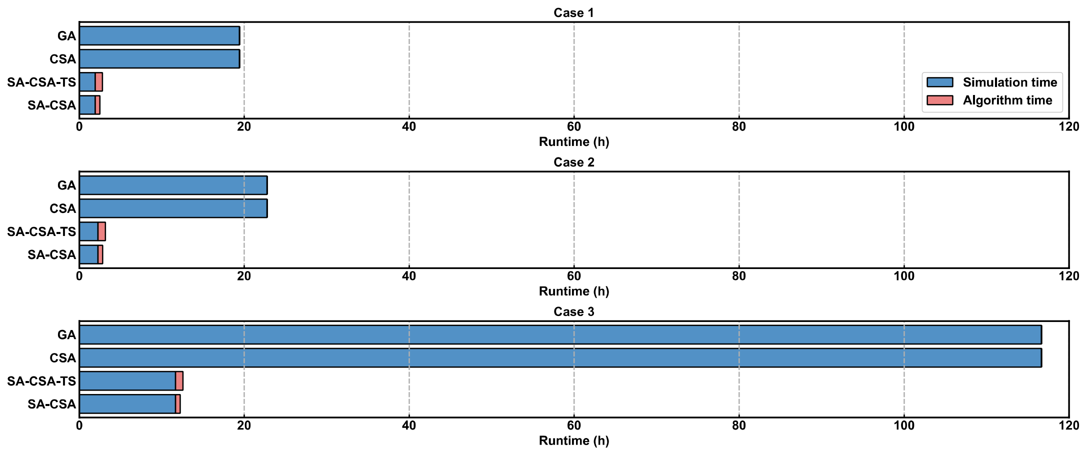


**Figure R1.** Runtime breakdown of all algorithms across three cases.

We once again thank you for your time and for the constructive feedback that has helped improve the quality of this paper. We hope that the revisions we have made successfully address your concerns and meet with your approval.

## Comments and Reponses regarding Dr. Wei Gong (RC2)

### General comments:

This paper proposed a new synergistic method called SA-CSA-TS based on (1) multi-chain (2) surrogate-assisted (3) hybrid optimization framework integrating the Cooperative Search Algorithm (CSA) and Tabu Search (TS) for Groundwater Contamination Source Identification (GCSI) problem. Three cases with synthetic and practical data were carried out, demonstrating that the proposed SA-CSA-TS method can consistently identify both the values of hydrogeological parameters and the locations and time-varying release fluxes of contaminant sources. The novel contributions of this paper are significant enough to be published on HESS. Only a few minor revisions are required.

1. Section 2.2.2, Page 5. Why was CSA chosen instead of other heuristic optimization algorithms, such as the well known Genetic Algorithm, Simulated Annealing, etc.? The reasons for choosing CSA need to be explained. After reading [Feng et.al., 2021], I got that CSA is a new optimization algorithm with outstanding performance, but it's not well known in the hydrology community. Please provide more information about CSA.
2. Section 2.3, Page 6 - 7. This paper established a surrogate-assisted optimization method for GCSI problem, which need to construct multiple surrogate models for each well. The computational burden of surrogate model might be too heavy, if the surrogate model itself is too expensive. Consequently, the surrogate model selected in this method should be cheap and effective. The reason of selecting RBF has been sufficiently demonstrated in the discussion part in section 5.2, line 385-390, page 17. But the introduction of the compared surrogate models in section 2.3 is somehow insufficient and misleading.

- The relationship between the Gaussian Processes Regression (GPR) and Kriging interpolation should be explained. The Gaussian Processes Regression originated from machine learning, and Kriging interpolation originated from geostatistics. The two are essentially the same method in mathematics, and the main difference between them is that they have different expressions of the same concept, because they grew up in difference community. It is necessary to introduce what kind of covariance function is used in the Gaussian Processes Regression here and how the hyper-parameters are set. If the mathematical equations are the same with Kriging, only one should be kept.
- Radial Basis Functioncan be used as a covariance function in Gaussian Processes Regression (GPR), and a kernel in Support Vector Regression (SVR). Please list out the setup details of GPR and SVR (e.g. the parameters assigned in scikit-learn function call. Please double check the details if the default parameters were used.), in order to demonstrated what were actually compared in this work.
- GPR(merge with Kriging), SVR and RBF methods has many hyper-parameters, which may have significant influence to the final fitting performance. The hyper-parameters in GPR (mainly in the covariance function) can be manually specified, or automatically optimized, as elaborated in [Rasmussen and Williams, 2006]. The hyper-parameters of SVR (use RBF kernel) C, gamma, epsilon, are usually optimized with a crude grid-search, or manually specified according to expert knowledge. Please double check the hyper-parameters used in the comparison, are they automatically optimized? Or had been specified with default values? The fitting RMSE varies greatly, possibly because the default hyper-parameters are used, which do not appropriate for the problem in this paper. A good practice is automatically optimize, or manually tune the hyper-

parameters after initial sampling and first construction of surrogate models.

1. Section 3, page 8. This section is also methodology section, and this is actually the novel contribution of this paper. Some content in this section overlaps with section 2. It is better to move this section as section 2, in order to highlight the research significance. Move the introduction of surrogate models, CSA and TS, as sub-sections within this section.
2. Table 3, page 16. Add RMSE values in this table.
3. These two references are the same book? Please double check.

Rasmussen, C.E., Williams, C.K.I., 2006a. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA.

Rasmussen, C.E., Williams, C.K.I., 2006b. Gaussian processes for machine learning, Adaptive computation and machine learning. MIT Press, Cambridge, Mass.

**General responses:**

We sincerely thank the reviewer for the careful reading of our manuscript and for the positive and encouraging evaluation of this work. We also greatly appreciate the reviewer's constructive suggestions, which have helped us further improve the clarity, rigor, and presentation of the manuscript. In the revised version, we have carefully addressed all comments and made corresponding revisions throughout the manuscript, including refining the manuscript structure, clarifying the reasons for selecting CSA, improving the description and comparison of surrogate models, providing more details on hyperparameter settings and tuning strategies, and carefully checking the references for consistency. Our point-by-point responses are provided below.

**Comment 1:** Section 3, page 8. This section is also methodology section, and this is actually the novel contribution of this paper. Some content in this section overlaps with section 2. It is better to move this section as section 2, in order to highlight the research significance. Move the introduction of surrogate models, CSA and TS, as sub-sections within this section.

**Response 1:** We thank the reviewer for this valuable suggestion. We agree that, in the original manuscript, the core methodological contribution of this study is distributed across Sections 2 and 3, which could reduce the coherence of the methodological presentation and weaken the prominence of the proposed framework. In response, we have substantially reorganized the manuscript structure in the revised version. Specifically, the methodological content originally presented in Section 3 has been moved into Section 2 and integrated into the overall methodology section. The revised Section 2 now includes the overview of the proposed SA-CSA-TS framework, surrogate modelling in SA-CSA-TS, surrogate-assisted CSA for global exploration, and surrogate-assisted TS for local refinement. Meanwhile, the case studies have been moved to Section 3. In addition, the descriptions of surrogate modelling, CSA, and TS have been reorganized so that they are presented more closely in connection with their specific roles within the proposed framework. We believe that this restructuring reduces overlap between sections, improves the logical flow of the manuscript, and better highlights the novelty and methodological significance of the proposed SA-CSA-TS framework.

**Comment 2:** Section 2.2.2, Page 5. Why was CSA chosen instead of other heuristic optimization algorithms, such as the well known Genetic Algorithm, Simulated Annealing, etc.? The reasons for choosing CSA need to be explained. After reading [Feng et.al., 2021], I got that CSA is a new optimization algorithm with outstanding performance, but it's not well known in the hydrology community. Please provide more information about CSA.

**Response 2:** We thank the reviewer for this helpful comment. We agree that the reasons for choosing CSA should be stated more clearly. In this study, CSA is selected not simply because it is a recent method, but because its search mechanism is well matched to the role of the first stage in our framework. The GCSI problem considered here is high-dimensional, nonlinear, mixed-variable, and strongly multimodal. Therefore, the first stage of SA-CSA-TS requires a population-based global search method that can maintain diversity while efficiently exchanging information among candidate solutions. Compared with a classical GA, CSA provides a more explicit cooperative learning structure through team communication, reflective learning, and internal competition; compared with a single-trajectory method such as simulated annealing, it is better suited to broad global exploration across multiple chains. In addition, following the reviewer's suggestion, we have expanded the manuscript to provide more background information on CSA and have added relevant references reporting its applications in water-resources and hydrological studies, including cascade reservoir operation, discharge simulation, and streamflow/flood forecasting. These additions were intended to better position CSA for readers in the hydrology community and to demonstrate that, although CSA is relatively new, it has already shown encouraging performance in related domains.

**The relevant descriptions in the revised manuscript (Section 2.5) now read as follows:**

In SA-CSA-TS, the first stage focuses on global exploration, where each chain independently executes the Cooperative Search Algorithm (CSA) with surrogate-based fitness evaluation. The CSA, proposed by Feng et al. (2021), is a population-based metaheuristic inspired by cooperative behaviours in social systems. Previous studies (Feng et al., 2022; Feng et al., 2024) have already shown its feasibility in related water-resources and hydrological applications, including cascade reservoir operation, discharge simulation, streamflow and flood forecasting. Here, CSA is adopted for GCSI because it emphasizes team communication, reflective learning and internal competition among individuals. These mechanisms are well suited to the high-dimensional, nonlinear, and potentially multimodal nature of the inverse problem, and are expected to identify promising regions for subsequent local refinement.

In CSA, a population of candidate solutions $\{x_i\}_{i=1}^{N}$ is initially generated. During the optimization process, individuals improve their positions by learning from others within the population. For example, at iteration $t$, the update of the $i$-th individual typically follows a team communication rule:

$$u_i^{t+1} = x_i^t + A_i^t + B_i^t + C_i^t \tag{3}$$
$$A_i^t = log\left(1/\phi(0,1)\right) \cdot (g_{ind}^t - x_i^t)$$
$$B_i^t = \alpha \cdot \phi(0,1) \cdot (gm^t - x_i^t)$$
$$C_i^t = \beta \cdot \phi(0,1) \cdot (pm^t - x_i^t)$$

where $A_i^t$, $B_i^t$ and $C_i^t$ denote the knowledge components from the chairman, board of directors, and board of supervisors, respectively. $g_{ind}^t$ is the $ind$-th global best individual at iteration $t$. The $gm^t$ represents the mean position of the top $M$ global best individuals. The $pm^t$ is the mean position of the $i$th personal best individual. In addition to this team-communication update, CSA also employs reflective learning and internal competition to maintain population diversity and retain superior individuals; other detailed algorithmic formulations can be found in Feng et al. (2021).

In the proposed algorithm, CSA is embedded in a surrogate-assisted manner. As illustrated in Fig. 3, the objective values of candidate solutions are predicted by the trained surrogate models instead of being repeatedly evaluated by the computationally expensive groundwater simulator. This substitution substantially improves the efficiency of the global exploration stage. The superior solutions generated by CSA are then used to update the current position of each chain (Line 08 in Algorithm 1). For comparison purposes, this surrogate-assisted CSA module is also implemented as a standalone benchmark algorithm, denoted SA-CSA, so that the specific contributions of the multi-chain architecture and the subsequent Tabu Search can be explicitly assessed against the complete SA-CSA-TS framework.

**Comment 3:** The relationship between the Gaussian Processes Regression (GPR) and Kriging interpolation should be explained. The Gaussian Processes Regression originated from machine learning, and Kriging interpolation originated from geostatistics. The two are essentially the same method in mathematics, and the main difference between them is that they have different expressions of the same concept, because they grew up in difference community. It is necessary to introduce what kind of covariance function is used in the Gaussian Processes Regression here and how the hyper-parameters are set. If the mathematical equations are the same with Kriging, only one should be kept.

**Response 3:** We thank the reviewer for this insightful comment. We agree that Kriging and Gaussian Process Regression (GPR) are mathematically closely related and mainly differ in terminology and implementation conventions that evolved in the geostatistics and machine-learning communities, respectively. In the original manuscript, presenting them separately with individual equations may have overstated their distinction. Following the reviewer's suggestion, we have revised the manuscript to clarify this relationship and simplified the theoretical presentation. Specifically, the separate mathematical derivations of KRG and GP have been removed from the main text, and we now describe them more concisely while referring readers to the standard references for detailed theory.

Nevertheless, we retained both KRG and GP in the comparative experiments because the purpose of this part of the study is to compare the practical performance of commonly used surrogate implementations in UQPyL. Although KRG and GP are theoretically closely related, differences in implementation details, hyperparameter estimation and model construction may still lead to different predictive performance in practice. These considerations have also been clarified in the revised manuscript, where the close theoretical relationship between KRG and GP, as well as the practical motivation for retaining both models in the benchmark comparison, are now explicitly stated.

**The relevant descriptions in the revised manuscript (Section 2.4) now read as follows:**

To alleviate the computational burden of repeated groundwater simulations, surrogate modelling is embedded into the proposed SA-CSA-TS framework. In GCSI, candidate parameters should be evaluated by the groundwater simulator to quantify the mismatch between simulated and observed concentrations at the monitoring wells (see the dashed line of Fig. 2). However, the entire optimization process typically requires thousands of forward simulations. To alleviate the computational demand, the SA-CSA-TS incorporates a surrogate modelling technique (see the solid line of Fig. 2). In this study, four commonly used surrogate models, namely Kriging (KRG), Gaussian Process (GP), Support Vector Regression (SVR), and Radial Basis Function (RBF), are considered as candidate approximators; detailed theoretical backgrounds of these models are available in Lophaven et al. (2002), Rasmussen and Williams (2006), Smola and Schölkopf (2004), and Buhmann (2003), respectively. Although KRG and GP are theoretically closely related, both are considered in this study because differences in practical implementation and hyperparameter estimation may still lead to different predictive performance. All four models are implemented in UQPyL, and their predictive performance is compared in Section 4. Rather than treating surrogate modelling as an independent component, the present study embeds it directly into the optimization workflow so that surrogate predictions can guide both the global exploration and the local refinement stage of SA-CSA-TS.

<u>**Comment 4:**</u> Radial Basis Functioncan be used as a covariance function in Gaussian Processes Regression (GPR), and a kernel in Support Vector Regression (SVR). Please list out the setup details of GPR and SVR (e.g. the parameters assigned in scikit-learn function call. Please double check the details if the default parameters were used.), in order to demonstrated what were actually compared in this work. GPR(merge with Kriging), SVR and RBF methods has many hyper-parameters, which may have significant influence to the final fitting performance. The hyper-parameters in GPR (mainly in the covariance function) can be manually specified, or automatically optimized, as elaborated in [Rasmussen and Williams, 2006]. The hyper-parameters of SVR (use RBF kernel) C, gamma, epsilon, are usually optimized with a crude grid-search, or manually specified according to expert knowledge. Please double check the hyper-parameters used in the comparison, are they automatically optimized? Or had

been specified with default values? The fitting RMSE varies greatly, possibly because the default hyper-parameters are used, which do not appropriate for the problem in this paper. A good practice is automatically optimize, or manually tune the hyper-parameters after initial sampling and first construction of surrogate models.

**Response 4:** We thank the reviewer for this important comment. We agree that the configuration details of the surrogate models should be reported more explicitly, since the predictive performance of KRG, GP, SVR, and RBF can be sensitive to the choice of covariance/kernel functions and hyperparameters. In response to the reviewer's suggestion, we have revised the manuscript to clarify the practical setup used in this study. Following the cost-benefit perspective on online model tuning in surrogate-assisted optimization proposed by Ahrari and Verstraete (2023), we did not rely entirely on untuned default settings. Instead, only selected influential hyperparameters are optimized using grid search, while the remaining parameters are retained at their default values in UQPyL. The tuned hyperparameters, their search ranges, and the corresponding covariance/kernel settings are now summarized in Table S4.

This design is adopted to balance fairness, transparency, and the computational cost associated with repeated surrogate reconstruction within the SA-CSA-TS framework. Therefore, the comparison presented in this study is not based on a purely untuned all-default configuration, but rather on a limited yet systematic tuning strategy for the most influential model parameters. These considerations have also been clarified in the revised manuscript. We believe that these revisions improve the transparency, reproducibility, and interpretability of the surrogate-model comparison, and more clearly demonstrate what was actually compared in this work.

**Reference:**

Ahrari, A., Verstraete, D., 2023. Online model tuning in surrogate-assisted optimization - an effective approach considering the cost-benefit tradeoff. Swarm Evol. Comput. 82, 101357. https://doi.org/10.1016/j.swevo.2023.101357

**The relevant descriptions in the revised manuscript (Section 4.1 and Table S4) now read as follows:**

This study employs four commonly used surrogate models to investigate their performance in predicting the discrepancy between observed and simulated data for a given set of solutions: a. Kriging (KRG); b. Gaussian Process (GP); c. Support Vector Regression (SVR); d. Radial Basis Function (RBF).

To ensure a fair comparison, all surrogate models are constructed using UQPyL on a computer equipped with 12th Gen Intel(R) Core (TM) i5-12490F CPU, and 32.0 GB of RAM. Motivated by the cost–benefit perspective of surrogate tuning discussed by Ahrari and Verstraete (2023), only selected influential hyperparameters are tuned in this study using grid-search, whereas the remaining hyperparameters are retained at their default values in UQPyL. The tuned hyperparameters and their search ranges are summarized in Table S4.

For sample generation, Latin Hypercube Sampling (LHS) is used in Cases 1-3 to produce a set of parameter samples, which are subsequently input into the groundwater models to obtain contaminant concentrations. For each sample, the RMSE between the simulated and observed concentrations at all monitoring wells is calculated. RMSE is selected here because it provides a steeper and more informative gradient, which is advantageous for optimization. The generated parameter sets and their corresponding RMSE values constitute the full input-output datasets.

To evaluate model performance, four training datasets, denoted as DS1-DS4 with sample sizes of 100, 200, 300, and 500, respectively, are constructed. An independent set of 50 samples is generated for testing.

**Table S4.** Tuned hyperparameters and candidate grids of the four surrogate models.

| Surrogate | Tuned hyperparameters | Candidate grid |
|---|---|---|
| KRG | Kernel $k$; Regression $r$;<br>Length-scale $l$; | $k \in \{Gaussian, Cubic, Exponential\}$;<br>$r \in \{Constant, Linear, Quadratic\}$;<br>$l \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$; |
| GP | Kernel $k$;<br>$\nu$ (Matérn only);<br>$\alpha$ (RQ only);<br>Length-scale $\theta$ | $k \in \{Gaussian, Matérn, Rational\ Quadratic\ (RQ)\}$;<br>$\nu \in \{0.5, 1.5, 2.5\}$;<br>$\theta \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$;<br>$\alpha \in \{0.1, 0.3, 1, 3, 10\}$; |
| $\varepsilon$ −SVR | Kernel $k$; $\gamma$; $C$;<br>$d_{poly}$(Polynomial only) | $k \in \{Gaussian, Polynomial\}$;<br>$r \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$;<br>$C \in \{10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$;<br>$d_{poly} \in \{1,2,3,4\}$; |
| RBF | Kernel $k$;<br>Shape parameter $\epsilon$ | $k \in \{Linear, Thin\ Plate\ Spline\ (TPS), Cubic, Multiquadric\ (MQ), Gaussian\}$;<br>$\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$; |

**Comment 5:** Table 3, page 16. Add RMSE values in this table; These two references are the same book? Please double check.

Rasmussen, C.E., Williams, C.K.I., 2006a. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA.

Rasmussen, C.E., Williams, C.K.I., 2006b. Gaussian processes for machine learning, Adaptive computation and machine learning. MIT Press, Cambridge, Mass.

**Response 5:** We thank the reviewer for this careful comment. In response, the RMSE values have been added to Table 3 to provide a clearer quantitative comparison of surrogate-model performance. In addition, we carefully rechecked the references and confirm that the two cited entries for Rasmussen and Williams (2006a, 2006b) refer to the same book. This duplication has been corrected in the revised manuscript, and only one standardized reference entry is now retained.

We sincerely thank the reviewer and the editor for their careful reading of the manuscript and for the constructive comments and suggestions. We have carefully revised the manuscript accordingly and believe that these changes have significantly improved its clarity, rigor, and overall quality. We hope that the revised version adequately addresses all concerns and is now suitable for publication in the journal.

## Comments and Reponses regarding Dr. Giacomo Medici (CC1)

**General comments:**

Good research that needs some improvement. See my specific comments that should improve the manuscript.

Specific comments

Lines 34-35. “Groundwater contamination has become an increasingly critical issue, posing significant risks to environmental safety and public health”. Insert recent literature on groundwater contamination with an evident worldwide angle:

- Agbotui, P. Y., Firouzbehi, F., Medici, G. 2025. Review of effective porosity in sandstone aquifers: insights for representation of contaminant transport. Sustainability, 17(14), 6469.
- Sauvé, S.,Desrosiers, M. 2014. A review of what is an emerging contaminant. Chemistry Central Journal, 8(1), 15.

Line 91. You need to disclose the general aim of the research.

Line 91. You need to describe the specific objectives of your research by using numbers (e.g., i, ii, and iii).

Line 92-onwards. You need to add more information on the boundary conditions.

Line 92-onwards. Add more detail on the nature of the geological material modelled.

Line 109. Overall, 9 equations in the manuscript are too many, not all of the are necessary. Equation 2 is very well known.

Line 175. Equations on kriging (very well-known method) not necessary.

Line 515. Assign a number to this equation.

Figures and tables

Figure 1-5. Room to make the figures larger.

Figure 8. You need to discuss boundary conditions in more detail in the main body.

Figure 8. Increase the graphic resolution of the figure.

**General responses:**

We sincerely thank the reviewer for the thoughtful and constructive comments. We agree that the manuscript could be further improved in several aspects, and we have revised it carefully following all of the suggestions. Specifically, we have:

**Comment 1:** Lines 34-35. “Groundwater contamination has become an increasingly critical issue, posing significant risks to environmental safety and public health”. Insert recent literature on groundwater contamination with an evident worldwide angle:

- Agbotui, P. Y., Firouzbehi, F., Medici, G. 2025. Review of effective porosity in sandstone aquifers: insights for representation of contaminant transport. Sustainability, 17(14), 6469.

- Sauvé, S., Desrosiers, M. 2014. A review of what is an emerging contaminant. Chemistry Central Journal, 8(1), 15.

**Response 1:** We sincerely thank the reviewer for this helpful suggestion. We carefully reviewed the recommended literature and agree that it helps strengthen the background and motivation of the study. In the revised manuscript, we incorporated Agbotui et al. (2025), which is directly relevant to groundwater contaminant transport and the hydrogeological context of the present study, and revised the opening sentence of the Introduction accordingly. In addition, we appreciate the reviewer's recommendation of Sauvé and Desrosiers (2014), which provides a broader perspective on emerging contaminants.

**The relevant descriptions in Introduction now read as follows:**

Groundwater contamination has become an increasingly critical issue, posing significant risks to environmental safety and public health (Gorelick and Zheng, 2015; Li et al., 2021a; Agbotui et al., 2025).

**Comment 2:** Line 91. You need to disclose the general aim of the research.

Line 91. You need to describe the specific objectives of your research by using numbers (e.g., i, ii, and iii).

Line 92-onwards. You need to add more information on the boundary conditions.

Line 92-onwards. Add more detail on the nature of the geological material modelled.

Line 109. Overall, 9 equations in the manuscript are too many, not all of the are necessary. Equation 2 is very well known.

Line 175. Equations on kriging (very well-known method) not necessary.

Line 515. Assign a number to this equation.

**Response 2:** We thank the reviewer for this helpful suggestion. We agree that the original manuscript included more mathematical detail than necessary for several well-established surrogate models. In the revised manuscript, we add more detailed geological material in Case 3. In addition, we streamlined the mathematical presentation of the manuscript (**merge original Section 2.3 and 3.2 into revised Section 2.4**). The detailed equations of well-established surrogate models, such as Kriging and Gaussian Process, have been removed, and the surrogate-modelling subsection now provides only a concise description together with appropriate references to the original literature. Finally, the previously unnumbered equation in the robustness-analysis section has now been assigned a number.

We believe these revisions improve the clarity, structure, and readability of the manuscript, while keeping the methodological description sufficiently focused on the novel contribution of this study.

**Comment 3:**

Figure 1-5. Room to make the figures larger.

Figure 8. You need to discuss boundary conditions in more detail in the main body.

Figure 8. Increase the graphic resolution of the figure.

**Response 3:** We sincerely thank the reviewer for these helpful comments. In response, we have improved the presentation of the figures in the revised manuscript. Figures 1-5 have been enlarged where possible to enhance readability. We have added more detailed geological material in Case 3 with Figure 8. In addition, the resolution of Fig. 8 has been improved to achieve better graphic quality in the revised version.

Once again, we sincerely thank Dr. Giacomo Medici for the careful reading of our manuscript and for the constructive and valuable suggestions.

## Comments and Reponses regarding Dr. Nima Zafarmomen (CC2)

**General comments:**

The study introduces a significant advancement in the simulation-optimization (S−O) framework for groundwater contamination diagnosis. The novelty of the SA-CSA-TS framework lies in its synergistic multi-chain architecture. Unlike traditional single-population algorithms, the integration of the Cooperative Search Algorithm (CSA) for global exploration and Tabu Search (TS) for local refinement—guided by a shared tabu list—effectively addresses the "equifinality" and multimodality inherent in mixed-variable groundwater problems.

Furthermore, the systematic evaluation of surrogate models and the implementation of a dynamic reconstruction strategy (achieving an 85–88% reduction in computational demand) provides a highly practical blueprint for real-world remediation efforts where time and computing resources are limited.

Minor Comments:

1. In Section 3.4, the authors describe the update rules for the tabu list. It would be beneficial to briefly clarify the "tenure" or size of the tabu list. Does the list have a maximum capacity, or does it grow indefinitely throughout the FEmax iterations?

2. While the authors mention using default settings in UQPyL, the performance of Kriging and Gaussian Processes is often highly sensitive to the choice of kernel/correlation functions. A brief sentence justifying the choice of the Cubic RBF kernel over others (like Thin Plate Spline) would add more depth to the surrogate comparison section.

3. In the discussion of the "parameter-compensation effect" (Section 7.2), the authors correctly identify that multiple locations can yield similar concentrations. It might be helpful to suggest how monitoring well placement (optimal experimental design) could potentially reduce this equifinality in future iterations of the framework.

4. In Figure 16 and Figure 18 (Radar Charts), the overlap of the GA and CSA lines can be difficult to distinguish. Consider using slightly different line textures (e.g., dashed vs. dotted) to improve accessibility for the reader.

5. To broaden the impact of the study, the authors should consider how this framework interacts with broader hydrological cycles and diverse data sources. I strongly recommend the authors consider and potentially reference studies such as: "Assimilation of sentinel-based leaf area index for modeling surface-ground water interactions in irrigation districts" This would help contextualize how satellite-derived data and surface-water interactions might provide additional constraints to the groundwater simulation models, potentially refining the identification of hydrogeological parameters.

**General responses:**

We sincerely thank the reviewer for the highly positive and insightful evaluation of our work. We are also grateful for these comments, which have helped us further improve the clarity, completeness, and presentation of the manuscript. In the revised version, we have carefully considered all of these suggestions and made corresponding revisions in the relevant sections. Detailed point-by-point responses are provided below:

**Comment 1:** In Section 3.4, the authors describe the update rules for the tabu list. It would be beneficial to briefly clarify the "tenure" or size of the tabu list. Does the list have a maximum capacity, or does it grow indefinitely throughout the FEmax iterations?

**Response 1:** We sincerely thank the reviewer for this helpful suggestion. We agree that a clearer description of the tabu list setting is important for better understanding the implementation of the TS component. Specifically, the maximum capacity of the tabu list should be equal to the total number of candidate source areas. In the revised manuscript, we have added a brief clarification in **Section 2.6 (original Section 3.4)** regarding the tabu list size (Lines 210-215 of the revised manuscript).

**Comment 2:** While the authors mention using default settings in UQPyL, the performance of Kriging and Gaussian Processes is often highly sensitive to the choice of kernel/correlation functions. A brief sentence justifying the choice of the Cubic RBF kernel over others (like Thin Plate Spline) would add more depth to the surrogate comparison section.

**Response 2:** Thank you for this helpful comment. We agree that our previous wording did not make this point sufficiently clear. In the revised manuscript, we now clarify that the surrogate models were not used with entirely default settings. Instead, to ensure a fair comparison, and following the cost–benefit perspective of surrogate tuning discussed by Ahrari and Verstraete (2023), only a small number of influential hyperparameters were tuned by grid search, while the remaining hyperparameters were retained at their default values in UQPyL. Among the tuned hyperparameters, the kernel/correlation function was explicitly examined for GP, KRG, and RBF. The tuned hyperparameters and their search ranges are now summarized in Table S4.

## Reference:

Ahrari, A., Verstraete, D., 2023. Online model tuning in surrogate-assisted optimization - an effective approach considering the cost-benefit tradeoff. Swarm Evol. Comput. 82, 101357. https://doi.org/10.1016/j.swevo.2023.101357

## The relevant descriptions in the revised manuscript (Section 4.1 and Table S4) now read as follows:

This study employs four commonly used surrogate models to investigate their performance in predicting the discrepancy between observed and simulated data for a given set of solutions: a. Kriging (KRG); b. Gaussian Process (GP); c. Support Vector Regression (SVR); d. Radial Basis Function (RBF).

To ensure a fair comparison, all surrogate models are constructed using UQPyL on a computer equipped with 12th Gen Intel(R) Core (TM) i5-12490F CPU, and 32.0 GB of RAM. Motivated by the cost–benefit perspective of surrogate tuning discussed by Ahrari and Verstraete (2023), only selected influential hyperparameters are tuned in this study using grid-search, whereas the remaining hyperparameters are retained at their default values in UQPyL. The tuned hyperparameters and their search ranges are summarized in Table S4.

For sample generation, Latin Hypercube Sampling (LHS) is used in Cases 1-3 to produce a set of parameter samples, which are subsequently input into the groundwater models to obtain contaminant concentrations. For each sample, the RMSE between the simulated and observed concentrations at all monitoring wells is calculated. RMSE is selected here because it provides a steeper and more informative gradient, which is advantageous for optimization. The generated parameter sets and their corresponding RMSE values constitute the full input-output datasets.

To evaluate model performance, four training datasets, denoted as DS1-DS4 with sample sizes of 100, 200, 300, and 500, respectively, are constructed. An independent set of 50 samples is generated for testing.

**Table S4.** Tuned hyperparameters and candidate grids of the four surrogate models.

| Surrogate | Tuned hyperparameters | Candidate grid |
|---|---|---|
| KRG | Kernel $k$; Regression $r$;<br>Length-scale $l$; | $k \in \{Gaussian, Cubic, Exponential\}$;<br>$r \in \{Constant, Linear, Quadratic\}$;<br>$l \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$; |
| GP | Kernel $k$;<br>$\nu$ (Matérn only);<br>$\alpha$ (RQ only);<br>Length-scale $\theta$ | $k \in \{Gaussian, Matérn, Rational\ Quadratic\ (RQ)\}$;<br>$\nu \in \{0.5, 1.5, 2.5\}$;<br>$\theta \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$;<br>$\alpha \in \{0.1, 0.3, 1, 3, 10\}$; |
| $\varepsilon$ −SVR | Kernel $k$; $\gamma$; $C$;<br>$d_{poly}$(Polynomial only) | $k \in \{Gaussian, Polynomial\}$;<br>$r \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$;<br>$C \in \{10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$;<br>$d_{poly} \in \{1,2,3,4\}$; |
| RBF | Kernel $k$; | $k \in \{Linear, Thin\ Plate\ Spline\ (TPS), Cubic, Multiquadric\ (MQ), Gaussian\}$; |

| | |
|---|---|
| Shape parameter $\epsilon$ | $\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{-3}, 1, 10^{1}, 10^{2}, 10^{3}\}$; |

**Comment 3:** In the discussion of the "parameter-compensation effect" (Section 7.2), the authors correctly identify that multiple locations can yield similar concentrations. It might be helpful to suggest how monitoring well placement (optimal experimental design) could potentially reduce this equifinality in future iterations of the framework.

**Response 3:** We sincerely thank the reviewer for this insightful suggestion. We agree that monitoring well placement and optimal experimental design are highly relevant to reducing equifinality and improving parameter identifiability in groundwater contaminant source identification. In the revised manuscript, we have added that future developments of the proposed framework could incorporate optimal monitoring well placement or experimental design to provide stronger spatial constraints on source location and hydrogeological parameters, thereby helping to reduce equifinality and improve inversion reliability.

**Comment 4:** In Figure 16 and Figure 18 (Radar Charts), the overlap of the GA and CSA lines can be difficult to distinguish. Consider using slightly different line textures (e.g., dashed vs. dotted) to improve accessibility for the reader.

**Response 4:** We sincerely thank the reviewer for this practical suggestion. We agree that improving the visual distinction between the GA and CSA curves would enhance the readability of the radar charts. In the revised manuscript, we have adjusted the line styles in Figures 13a-b (original Figure 16 and Figure 18) to make the overlapping curves easier to distinguish and to improve the accessibility of the figures for readers.

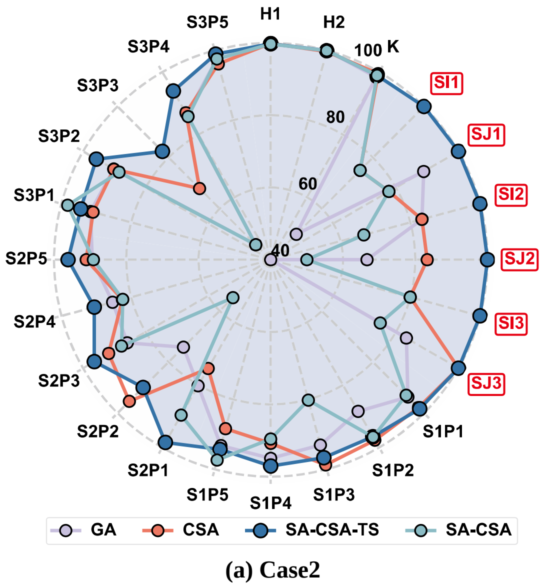


**Figure R1.** Radar chart comparing the optimal solutions obtained by four algorithms: (a) Case2 and (b) Case3.

**Comment 5:** To broaden the impact of the study, the authors should consider how this framework interacts with broader hydrological cycles and diverse data sources. I strongly recommend the authors consider and potentially reference studies such as: "Assimilation of sentinel-based leaf area index for modeling surface-ground water interactions in irrigation districts" This would help contextualize how satellite-derived data and surface-water interactions might provide additional constraints to the groundwater simulation models, potentially refining the identification of hydrogeological parameters.

**Response 5:** We sincerely thank the reviewer for this thoughtful and constructive suggestion. In the revised manuscript, we have added a statement in the conclusion section to note that future work will explore the incorporation of surface-groundwater interactions and multi-source observations, including satellite-derived data, to provide additional constraints for hydrogeological parameter identification in more complex cases.

Once again, we sincerely thank Dr. Nima Zafarmomen for the careful reading of our manuscript and for the constructive and valuable suggestions.