

Dear Editor and Reviewers,

We would like to express our sincere appreciation for your handling of our manuscript and for the valuable comments and suggestions offered. We have carefully considered the comments and have revised the manuscript accordingly. The comments are laid out below in italicized font. Our response is given in normal font and changes/additions to the manuscript are given in the blue text.

### **Responses to reviewer three's comments**

1. *#L512-514: This is an interesting observation, but please check whether you have enough data points for  $M > 6$  to support this statement. If not, clarify that it is not statistically significant. Be careful with using the word “significantly” if no statistical tests were done.*

**Response:** We sincerely thank the reviewer for the cautious approach regarding the statistical significance of the  $M \geq 6$  events. We have revised the manuscript on line 533 to avoid the word “significantly” and have explicitly characterized the observed dominance of magnitude over distance as a preliminary trend due to the limited sample size.

For earthquakes where  $M \geq 6.0$ , **magnitude appears to be a more dominant driver than distance, although limited data renders this a preliminary trend rather than a statistically significant finding. This is potentially due to high-energy release triggering regional-scale crustal stress perturbations that induce hydrochemical anomalies even at remote distances. The overall findings align** with the positive correlation between the scale of earthquake energy release and the number of anomalies, as confirmed by the hydrochemical monitoring results (Li et al., 2022).

2. *#L538-540: The assumption of isotropic structure is a spatial factor, not a temporal factor. It is uncertain why this results in “exhibits limited adaptability to long-term*

*data trends”.*

**Response:** Thanks! we have revised this section on line 560 to focus strictly on the spatial limitation imposed by the isotropic assumption:

This highlights another key limitation: the current algorithm relies on an effective and widely used isotropic underground structure assumption, *which restricts model’s adaptability to account for directional tectonic complexities and spatial heterogeneity.*

3. *#L541-549: Are all these compared literature findings concerning related problems? For example the reference on ionospheric anomalies may be too unrelated to the present manuscript. If they are not related to this manuscript, then comparing the accuracy levels of 60% to 83% may not be meaningful. It is better to discuss 2 strongly related literature in detail, rather than to gloss shallowly over 4 literature with weak relevance to the present manuscript.*

**Response:** Thanks! We agree that the previous version included references with less direct relevance. In the revised manuscript, we have removed the cross-disciplinary references (e.g., Mukherjee et al., 2025; Baselga, 2024) to focus on the hydrogeochemical domain. We have also expanded the comparative discussion on line 563, providing a more nuanced analysis of the differences in indicator logic and methodological perspectives relative to the studies by Skelton et al. (2024) and Zhu et al. (2024).

Despite these constraints, the multi-component anomaly detection model developed here demonstrates robust performance, achieving a POD exceeding 83% and a TS ranging from 59% to 70%. *Within the field of geochemical forecasting, these results are comparable to the positive predictive values of 62% to 85% derived from the statistical analysis reported by Skelton et al. (2024) in Iceland. While that study utilized the oscillatory behaviors of  $\delta D$  and  $\delta^{18}O$ , and deuterium excess to trace groundwater mixing associated with crustal dilation and fracture mineralization before earthquakes, the current approach seeks to complement these isotopic insights by*

integrating them with multiple major ions. Synchronous fluctuations in ion concentrations can provide additional anomaly signals when isotopic shifts are subtle, reflecting a wider range of water-rock interaction processes. The model results also align with the 70% POD and R-score of 0.6 described by Zhu et al. (2024) in the southeast coastal region of China, where ion ratios and PCA indicators were combined with the best-performing LOF algorithm to identify outliers. In contrast, the present model focuses on refining anomaly identification performance and classifying anomaly intensity by assessing the number of components with synchronous anomalies across multiple components. By leveraging the differentiated responses of various components to crustal stress, this synergistic analysis significantly enhances the identification of subtle pre-earthquake signals. Mechanistically, this multi-indicator integration provides a more objective reflection of pre-earthquake stress changes and fluid mixing processes while offering certain advantages in reducing environmental interference, consistent with the tectonic stress drivers discussed by Yakupoğlu et al. (2025).

4. #L549: Use “synergistic” or “synergetic” consistently in the paper.

**Response:** Thanks! We have standardized the terminology throughout the manuscript, using “synergistic” consistently to replace any instances of “synergetic”.

5. #L556: *How does the temperature influence the amount of information in the hydrochemical signal (citation?)? Or is high temperature used in this sentence as a correlate of “deep-circulated”? In L213 it is stated that the influence of temperature is negligible.*

**Response:** We appreciate the reviewer’s insightful comment. We would like to clarify that “high temperature” is used as a proxy for deep circulation depth. This is distinct from the temporal temperature fluctuations (seasonal changes) mentioned in L213, which do not significantly alter ion concentrations. In contrast, the discharge

temperature is a static criterion used to ensure a deep-seated fluid origin. We have clarified this distinction in the revised manuscript on line 582:

Priority should be given to *thermal springs with high discharge temperatures, as such temperatures typically indicate deep circulation and their hydrochemical components may carry deep-seated information while minimizing interference from meteoric water. Furthermore, the hydrogeochemical background of the spring should be clearly defined to distinguish its hydrogeochemical characteristics from shallow water bodies and to facilitate the monitoring of fluid mixing processes.*

6. *#L563-565: However, in such sites with site-specific hydrochemical indicators, would it be reasonable to use an anomaly detection method that gives equal weight to anomalies in different indicators, as you have done in this study? This statement could be improved with more nuance.*

**Response:** We thank the reviewer for this insightful comment. In this study, equal weighting was adopted to make the model more versatile. Without precise data on how sensitive each component is at different locations, using the same weight for all helps avoid subjective errors. We agree that for sites incorporating high-sensitivity indicators, an optimized weighting scheme could further enhance the model's performance. We have added a brief statement to the manuscript on line 590 to discuss this possibility.

Additionally, sites that could incorporate specific crustal deformation-sensitive isotopic monitoring indicators are ideal, such as He isotopes in the Noto Peninsula (Kagoshima et al., 2025) or H-O-C isotopes in the Xianshuihe fault (Yu et al., 2026). *In such cases, the implementation of differentiated weighting based on indicator sensitivity offers potential for further refining the precision of anomaly detection.*

7. *#Response 1 to reviewer 3: This is a good response that provides interesting insight. Adding it to the discussion could improve the paper.*

**Response:** Thank! We have incorporated the discussed content into Section 4.5

(Evaluation of forecasting performance, Lines 472–491) of the revised manuscript. Additionally, the detailed statistical data of anomaly directions has been included as Table S7 in the Supplementary Information.

Owing to variations in the geochemical behaviors of hydrochemical components, their response direction and magnitudes to earthquakes differ. The direction of anomalous variations (increase/decrease) is crucial for understanding the response mechanisms and improving event identification. A detailed statistical analysis of the anomaly directions for components across the two hot springs was performed (Table S7). The results reveal that certain pre-seismic anomalies manifest as intense bidirectional fluctuations. However, a definitive criterion to define a “dominant direction” for such oscillatory signals is lacking. Initial statistical results indicate that the frequency of anomalous increases (57%~77%) is higher than that of decreases for most ions (e.g.,  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ). While these findings provide tentative support for the assumption that “increases” possess a stronger correlation with seismic events (Skelton et al., 2014, 2024; Yakupoğlu et al., 2025), the current sample size and statistical depth remain exploratory. Additionally, as observed regarding  $\delta\text{D}$  and  $\delta^{18}\text{O}$ , statistics indicate that these components exhibit a higher frequency of decreases (60%~75%), and change directions can vary across different components for the same earthquake event. Such divergence among different components suggests that a simplistic, unidirectional criterion might overlook vital precursor information. Based on these findings, the existing detection model is primarily oriented toward identifying sustained positive anomalies (increases), while also accounting for recovery phases following a decline (i.e., the increase following a decrease). Nevertheless, effectively incorporating these specific directional variations into earthquake forecasting remains a challenge at this stage.

Given these complexities and the lack of academic consensus on response mechanisms, this study evaluates the forecasting efficacy of individual components through the anomaly detection model to identify the effective indicators for the study area. A comparison of the TS values of each component’s warning results in QJ and

WN (Figure 10) reveals that in the two thermal springs of the study area, the TS values for  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\delta\text{D}$ , and  $\delta^{18}\text{O}$  detection (around 0.50) are relatively high.

**Table S7: Statistical analysis of anomalous variations in hydrochemical components.**

Component	Qujiang spring				Wana spring			
	Increase ( <i>n</i> )	Decrease ( <i>n</i> )	Increase Ratio	Decrease Ratio	Increase ( <i>n</i> )	Decrease ( <i>n</i> )	Increase Ratio	Decrease Ratio
$\text{Na}^+$	10	4	0.71	0.29	3	1	0.75	0.25
$\text{K}^+$	8	4	0.67	0.33	3	2	0.60	0.40
$\text{Ca}^{2+}$	10	6	0.63	0.38	4	5	0.44	0.56
$\text{Cl}^-$	11	7	0.61	0.39	6	3	0.67	0.33
$\text{SO}_4^{2-}$	10	3	0.77	0.23	5	5	0.50	0.50
$\text{HCO}_3^-$	6	4	0.60	0.40	5	2	0.71	0.29
$\text{F}^-$	11	4	0.73	0.27	4	3	0.57	0.43
$\delta\text{D}$	1	3	0.25	0.75	2	3	0.40	0.60
$\delta^{18}\text{O}$	1	2	0.33	0.67	2	3	0.40	0.60

## Responses to reviewer four's comments

### # Major comments:

1. *#The anomaly detection framework lacks sufficient statistical rigor and uncertainty quantification. The proposed model (Section 3.4) relies heavily on multiple empirical thresholds ( $p1-p5$  and  $T_{th}$ ), yet no uncertainty propagation is provided for these parameters. In addition, key evaluation metrics such as POD, FAR, and TS are reported without confidence intervals, and there is no robustness assessment using approaches such as bootstrap or cross-validation. The absence of comparison with null or baseline models further weakens the credibility of the reported performance. As a result, the high detection performance (e.g.,  $POD = 0.95$  at  $QJ$ ) may be partly due to overfitting to relatively short datasets (5 years and 2.5 years). Without uncertainty bounds, it is difficult to assess the true predictive capability of the model, which is particularly problematic given the inherently probabilistic nature of earthquake forecasting.*

**Response:** We sincerely thank the reviewer for this insightful comment. We agree that reporting parameters and evaluation metrics without uncertainty bounds or baseline comparisons could lead to concerns regarding overfitting and the true forecasting capability of the model.

To address the concern about uncertainty propagation for parameters and the empirical nature of the thresholds, we would like to clarify that the selection of  $p1-p5$  and  $T_{th}$  was not based on arbitrary assignment but was the result of a systematic optimization and sensitivity analysis. In the Supplementary Material (Tables S2 and S3), we presented the model's performance (POD, FAR, and TS) under various combinations of  $p1$  and  $p3$  key parameters. This analysis functions as a form of uncertainty propagation, showing how changes in these core parameters affect the final forecasting metrics. The results show that the model maintains stable performance within a reasonable range of these parameters.

**Table S2: Model performance metrics (NA, NB, NC, FAR, MAR, POD, TS) under varying parameters p1 and p3 for Na<sup>+</sup> at Qujiang Spring.**

p1	p3	NB	NA	NA+NC	FAR	MAR	POD	TS
1	1	19	20	22	0.49	<b>0.09</b>	<b>0.91</b>	0.49
1	1.01	12	18	22	0.40	0.18	0.82	0.53
1	1.02	6	14	22	0.30	0.36	0.64	0.50
1	1.03	5	11	22	0.31	0.50	0.50	0.41
1	1.04	4	9	22	0.31	0.59	0.41	0.35
1	1.05	3	7	22	0.30	0.68	0.32	0.28
1	1.06	2	5	22	0.29	0.77	0.23	0.21
1	1.07	1	4	22	0.20	0.82	0.18	0.17
1	1.08	1	4	22	0.20	0.82	0.18	0.17
1.01	1	14	19	22	0.42	0.14	0.86	0.53
1.01	1.01	8	16	22	0.33	0.27	0.73	0.53
1.01	1.02	5	12	22	0.29	0.45	0.55	0.44
1.01	1.03	5	5	22	0.50	0.77	0.23	0.19
1.01	1.04	4	6	22	0.40	0.73	0.27	0.23
1.01	1.05	1	6	22	<b>0.14</b>	0.73	0.27	0.26
1.01	1.06	1	5	22	0.17	0.77	0.23	0.22
1.01	1.07	1	4	22	0.20	0.82	0.18	0.17
1.01	1.08	1	2	22	0.33	0.91	0.09	0.09
1.02	1	8	16	22	0.33	0.27	0.73	0.53
1.02	1.01	4	14	22	0.22	0.36	0.64	<b>0.54</b>
1.02	1.02	4	11	22	0.27	0.50	0.50	0.42
1.02	1.03	4	7	22	0.36	0.68	0.32	0.27
1.02	1.04	2	6	22	0.25	0.73	0.27	0.25
1.02	1.05	1	5	22	0.17	0.77	0.23	0.22
1.02	1.06	1	5	22	0.17	0.77	0.23	0.22
1.02	1.07	1	4	22	0.20	0.82	0.18	0.17
1.02	1.08	1	2	22	0.33	0.91	0.09	0.09
1.03	1	4	13	22	0.24	0.41	0.59	0.50
1.03	1.01	4	12	22	0.25	0.45	0.55	0.46
1.03	1.02	2	10	22	0.17	0.55	0.45	0.42
1.03	1.03	2	6	22	0.25	0.73	0.27	0.25
1.03	1.04	2	6	22	0.25	0.73	0.27	0.25
1.03	1.05	1	5	22	0.17	0.77	0.23	0.22
1.03	1.06	1	5	22	0.17	0.77	0.23	0.22
1.03	1.07	1	3	22	0.25	0.86	0.14	0.13
1.03	1.08	1	2	22	0.33	0.91	0.09	0.09
1.04	1	2	10	22	0.17	0.55	0.45	0.42
1.04	1.01	2	8	22	0.20	0.64	0.36	0.33
1.04	1.02	2	7	22	0.22	0.68	0.32	0.29

1.04	1.03	2	6	22	0.25	0.73	0.27	0.25
1.04	1.04	2	5	22	0.29	0.77	0.23	0.21
1.04	1.05	1	5	22	0.17	0.77	0.23	0.22
1.04	1.06	1	5	22	0.17	0.77	0.23	0.22
1.04	1.07	1	3	22	0.25	0.86	0.14	0.13
1.04	1.08	1	2	22	0.33	0.91	0.09	0.09
1.05	1	2	6	22	0.25	0.73	0.27	0.25
1.05	1.01	2	6	22	0.25	0.73	0.27	0.25
1.05	1.02	2	6	22	0.25	0.73	0.27	0.25
1.05	1.03	2	6	22	0.25	0.73	0.27	0.25
1.05	1.04	2	5	22	0.29	0.77	0.23	0.21
1.05	1.05	1	5	22	0.17	0.77	0.23	0.22
1.05	1.06	1	5	22	0.17	0.77	0.23	0.22
1.05	1.07	1	3	22	0.25	0.86	0.14	0.13
1.05	1.08	1	2	22	0.33	0.91	0.09	0.09
1.06	1	2	5	22	0.29	0.77	0.23	0.21
1.06	1.01	2	5	22	0.29	0.77	0.23	0.21
1.06	1.02	2	5	22	0.29	0.77	0.23	0.21
1.06	1.03	2	5	22	0.29	0.77	0.23	0.21
1.06	1.04	2	4	22	0.33	0.82	0.18	0.17
1.06	1.05	1	4	22	0.20	0.82	0.18	0.17
1.06	1.06	1	4	22	0.20	0.82	0.18	0.17
1.06	1.07	1	3	22	0.25	0.86	0.14	0.13
1.06	1.08	1	2	22	0.33	0.91	0.09	0.09

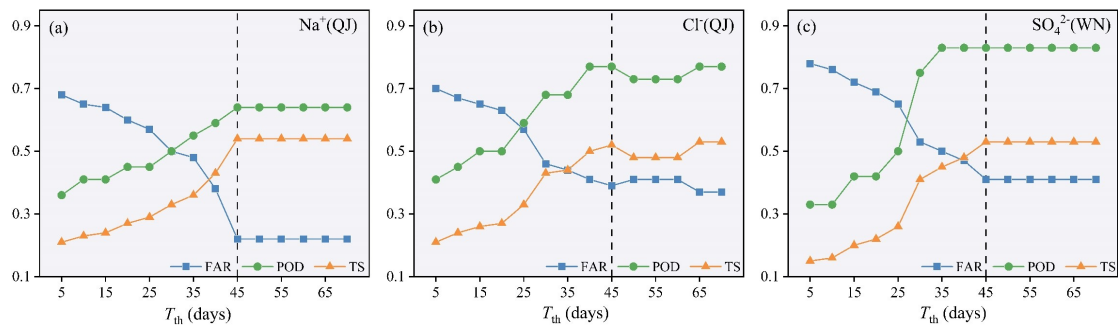
**Table S3: Model performance metrics (NA, NB, NC, FAR, MAR, POD, TS) under varying parameters p1 and p3 for SO<sub>4</sub><sup>2-</sup> at Wana Spring.**

p1	p3	NB	NA	NA+NC	FAR	MAR	POD	TS
1	1	12	10	12	0.55	<b>0.17</b>	<b>0.83</b>	0.42
1	1.01	11	10	12	0.52	<b>0.17</b>	<b>0.83</b>	0.43
1	1.02	11	10	12	0.52	<b>0.17</b>	<b>0.83</b>	0.43
1	1.03	11	10	12	0.52	<b>0.17</b>	<b>0.83</b>	0.43
1	1.04	9	8	12	0.53	0.33	0.67	0.38
1	1.05	8	8	12	0.50	0.33	0.67	0.40
1	1.06	7	8	12	0.47	0.33	0.67	0.42
1	1.07	7	6	12	0.54	0.50	0.50	0.32
1	1.08	6	6	12	0.50	0.50	0.50	0.33
1.01	1	11	10	12	0.52	<b>0.17</b>	<b>0.83</b>	0.43
1.01	1.01	11	10	12	0.52	<b>0.17</b>	<b>0.83</b>	0.43
1.01	1.02	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.01	1.03	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.01	1.04	8	8	12	0.50	0.33	0.67	0.40
1.01	1.05	8	8	12	0.50	0.33	0.67	0.40

1.01	1.06	7	8	12	0.47	0.33	0.67	0.42
1.01	1.07	7	6	12	0.54	0.50	0.50	0.32
1.01	1.08	6	6	12	0.50	0.50	0.50	0.33
1.02	1	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.02	1.01	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.02	1.02	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.02	1.03	10	10	12	0.50	<b>0.17</b>	<b>0.83</b>	0.45
1.02	1.04	8	8	12	0.50	0.33	0.67	0.40
1.02	1.05	8	8	12	0.50	0.33	0.67	0.40
1.02	1.06	7	8	12	0.47	0.33	0.67	0.42
1.02	1.07	7	6	12	0.54	0.50	0.50	0.32
1.02	1.08	6	6	12	0.50	0.50	0.50	0.33
1.03	1	9	10	12	0.47	<b>0.17</b>	<b>0.83</b>	0.48
1.03	1.01	8	10	12	0.44	<b>0.17</b>	<b>0.83</b>	0.50
1.03	1.02	7	9	12	0.44	0.25	0.75	0.47
1.03	1.03	7	9	12	0.44	0.25	0.75	0.47
1.03	1.04	7	8	12	0.47	0.33	0.67	0.42
1.03	1.05	7	8	12	0.47	0.33	0.67	0.42
1.03	1.06	6	8	12	0.43	0.33	0.67	0.44
1.03	1.07	6	6	12	0.50	0.50	0.50	0.33
1.03	1.08	5	6	12	0.45	0.50	0.50	0.35
1.04	1	9	10	12	0.47	<b>0.17</b>	<b>0.83</b>	0.48
1.04	1.01	7	10	12	<b>0.41</b>	<b>0.17</b>	<b>0.83</b>	<b>0.53</b>
1.04	1.02	7	9	12	0.44	0.25	0.75	0.47
1.04	1.03	7	9	12	0.44	0.25	0.75	0.47
1.04	1.04	7	8	12	0.47	0.33	0.67	0.42
1.04	1.05	7	8	12	0.47	0.33	0.67	0.42
1.04	1.06	6	8	12	0.43	0.33	0.67	0.44
1.04	1.07	6	6	12	0.50	0.50	0.50	0.33
1.04	1.08	5	6	12	0.45	0.50	0.50	0.35
1.05	1	9	9	12	0.50	0.25	0.75	0.43
1.05	1.01	8	9	12	0.47	0.25	0.75	0.45
1.05	1.02	8	9	12	0.47	0.25	0.75	0.45
1.05	1.03	7	9	12	0.44	0.25	0.75	0.47
1.05	1.04	7	8	12	0.47	0.33	0.67	0.42
1.05	1.05	7	8	12	0.47	0.33	0.67	0.42
1.05	1.06	6	8	12	0.43	0.33	0.67	0.44
1.05	1.07	6	6	12	0.50	0.50	0.50	0.33
1.05	1.08	5	6	12	0.45	0.50	0.50	0.35
1.06	1	7	8	12	0.47	0.33	0.67	0.42
1.06	1.01	7	8	12	0.47	0.33	0.67	0.42
1.06	1.02	6	7	12	0.46	0.42	0.58	0.39
1.06	1.03	6	7	12	0.46	0.42	0.58	0.39

1.06	1.04	6	6	12	0.50	0.50	0.50	0.33
1.06	1.05	6	6	12	0.50	0.50	0.50	0.33
1.06	1.06	5	6	12	0.45	0.50	0.50	0.35
1.06	1.07	5	5	12	0.50	0.58	0.42	0.29
1.06	1.08	4	5	12	0.44	0.58	0.42	0.31

Figure S6 in the Supplementary Materials provides a detailed sensitivity analysis of the evaluation metrics (FAR, POD, and TS) across different  $T_{th}$  values. By identifying the plateau where the metrics converge, we ensured that the selected  $T_{th}$  is statistically justified within the specific hydrogeological context of our study.



**Figure S6: Variations in the false alarm rate (FAR), probability of detection (POD), and threat score (TS) as a function of the seismic response time threshold ( $T_{th}$ ).**

Based on the above supplementary materials, we have revised the manuscript on line 415:

The results for all evaluation metrics under varying parameters ( $p1$  and  $p3$ ) for  $\text{Na}^+$  at QJ and for  $\text{SO}_4^{2-}$  at WN are provided in Table S2 and Table S3, respectively. [These sensitivity analyses demonstrate a stable model performance, confirming its robustness against parameter uncertainty.](#) According to this method, the complete set of model parameters for all hydrochemical components at QJ and WN is provided in Table S4.

Regarding the robustness assessment of evaluation metrics, we performed 1,000 bootstrap resamples to calculate the 95% confidence intervals (CIs) for all evaluation metrics (POD, FAR, and TS) across individual hydrochemical components and the synergistic alarm system (Table S6). At QJ, the comprehensive alarms (DA) reached a TS of 0.70 with a 95% CI of 0.53–0.87, while individual components such as  $\text{Ca}^{2+}$  (TS = 0.52 [0.31, 0.69]) and  $\text{Cl}^-$  (TS = 0.53 [0.33, 0.70]) also exhibited consistent

forecasting capabilities. Similarly, at WN, the DA achieved a TS of 0.59 [0.35, 0.82], with stable contributions from components like  $\text{SO}_4^{2-}$  (TS = 0.53 [0.33, 0.78]). This analysis demonstrates that even at the lower bound of the 95% CI, our model maintains a reliable detection capability. Furthermore, the CIs of the DA are relatively narrower compared to those of individual components, indicating that the multi-component synergistic approach effectively enhances the stability and reduces the uncertainty of the detection results.

Regarding the stochastic baseline model comparison, we conducted 1,000 Monte Carlo simulations to establish a stochastic baseline (null model) of the TS. As shown in Figure S7, the actual TS values for QJ (0.70) and WN (0.59) are significantly higher than the mean null TS (0.39 and 0.41, respectively), with p-values of  $p < 0.001$  (QJ) and  $p = 0.025$  (WN). The results demonstrate that our model's performance is significantly better than random chance, mitigating concerns about overfitting.

While we acknowledge the limitations associated with the relatively short dataset and the finite number of seismic events, our additional analyses statistically confirm that the model's performance is not merely a product of overfitting. Instead, the results reflect a predictive capability that significantly outperforms random chance. Accordingly, we have included these robustness assessments in the Supplementary Material and revised the manuscript on line 433:

**Table S6: Model performance evaluation and uncertainty analysis of hydrochemical components at Qujiang and Wana springs based on 1,000 bootstrap resamples.**

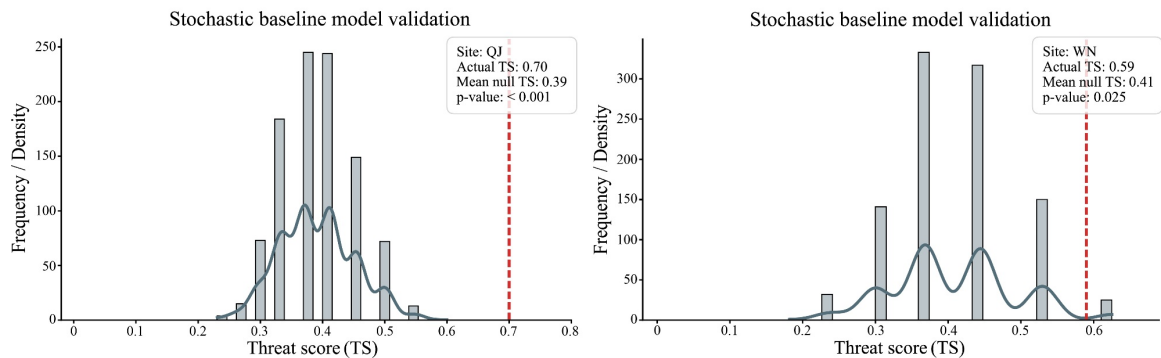
Comp- onent	Qujiang spring			Wana spring		
	POD (95% CI)	FAR (95% CI)	TS (95% CI)	POD (95% CI)	FAR (95% CI)	TS (95% CI)
$\text{Na}^+$	0.64 (0.43, 0.83)	0.22 (0.05, 0.44)	0.54 (0.35, 0.73)	0.42 (0.17, 0.67)	0.00 (0.00, 0.00)	0.42 (0.17, 0.67)
$\text{K}^+$	0.55 (0.35, 0.74)	0.37 (0.18, 0.61)	0.41 (0.23, 0.57)	0.50 (0.22, 0.79)	0.33 (0.07, 0.67)	0.40 (0.13, 0.67)
$\text{Ca}^{2+}$	0.73 (0.54, 0.91)	0.36 (0.19, 0.60)	0.52 (0.31, 0.69)	0.75 (0.50, 1.00)	0.36 (0.14, 0.63)	0.53 (0.29, 0.76)
$\text{Cl}^-$	0.77 (0.58, 0.95)	0.37 (0.21, 0.58)	0.53 (0.33, 0.70)	0.75 (0.50, 1.00)	0.36 (0.12, 0.60)	0.53 (0.29, 0.76)
$\text{SO}_4^{2-}$	0.59 (0.38, 0.79)	0.28 (0.10, 0.50)	0.48 (0.30, 0.67)	0.83 (0.58, 1.00)	0.41 (0.13, 0.62)	0.53 (0.33, 0.78)

HCO <sub>3</sub> <sup>-</sup>	0.41 (0.21, 0.62)	0.44 (0.24, 0.71)	0.31 (0.17, 0.47)	0.58 (0.31, 0.86)	0.22 (0.00, 0.50)	0.50 (0.29, 0.79)
F <sup>-</sup>	0.55 (0.35, 0.75)	0.43 (0.13, 0.56)	0.39 (0.25, 0.61)	0.58 (0.30, 0.86)	0.36 (0.09, 0.67)	0.44 (0.19, 0.69)
δD	1.00 (1.00, 1.00)	0.20 (0.00, 0.60)	0.80 (0.40, 1.00)	1.00 (1.00, 1.00)	0.50 (0.20, 0.80)	0.50 (0.20, 0.80)
δ <sup>18</sup> O	0.75 (0.25, 1.00)	0.40 (0.00, 0.83)	0.50 (0.17, 0.83)	1.00 (1.00, 1.00)	0.44 (0.11, 0.78)	0.56 (0.22, 0.89)
DA	0.95 (0.85, 1.00)	0.28 (0.10, 0.45)	0.70 (0.53, 0.87)	0.83 (0.60, 1.00)	0.33 (0.12, 0.60)	0.59 (0.35, 0.82)

Note: Values are original estimated metrics; values in parentheses are their corresponding 95% confidence intervals (CI).

Figures 8 and 9 present the 15-day moving average time series of hydrochemical components, anomaly detection results, and earthquake events for the anomaly detection model at QJ and WN. For each component, the model successfully identifies varying numbers of pre-earthquake anomalies and triggered alarms. The model activates comprehensive alarms when anomalies are detected in three or more components simultaneously. Based on the sensitivity analysis results presented in Table S5, requiring three or more anomalous components was identified as the preferred criterion for balancing the POD and the FAR. Furthermore, this choice is hydrochemically justified by its ability to suppress random dual-component interferences stemming from a single source, as discussed previously, thereby leading to the highest TS among all tested numbers of components. As detailed in Table S6, the 1,000-iteration bootstrap resampling results provide a robust evaluation of each individual component's performance alongside the comprehensive alarm (DA). This analysis demonstrates that even at the lower bound of the 95% confidence interval (CI), the model maintains a reliable detection capability. At QJ, the model provided 21 effective alarms for 22 earthquake events (POD = 0.95, 95% CI: 0.85~1.00), with 8 false alarms (FAR = 0.28, 95% CI: 0.10~0.45), and a TS of 0.70 with a 95% CI of 0.53~0.87. At WN, the model generated 10 accurate alarms for 12 events (POD = 0.83, 95% CI: 0.60~1.00), 5 false alarms (FAR = 0.33, 95% CI: 0.12~0.60), and a TS of 0.59 with a 95% CI of 0.35~0.82. To further evaluate the robustness of these results and exclude potential overfitting, the model performance was compared against a stochastic baseline model through 1,000 Monte Carlo iterations (Figure S7). The actual TS values

for QJ (0.70) and WN (0.59) are significantly higher than the mean null TS (0.39 and 0.41, respectively), with p-values of  $p < 0.001$  for QJ and  $p = 0.025$  for WN. This statistical evidence confirms that the model’s forecasting performance is statistically significant and demonstrably superior to random chance. Although the obtained FAR remains relatively high, which is a common challenge in earthquake anomaly detection that prioritises detection sensitivity, the TS provides a comprehensive metric that balances both POD and FAR. Compared with the internal single-component anomaly detection results from our model, the multicomponent synergistic alarm results exhibit higher TS values (Figures 8, 9, 10). Their CIs are relatively narrower than those of individual components (Table S6), indicating that the synergistic approach effectively enhances the stability and reduces the uncertainty. The multicomponent synergy mitigates the effects of geochemical behavior differences among components, reduces environmental interference on individual ions/ion pairs, and consequently enhances the accuracy of the anomaly detection model. These results indicate the practical value of the multicomponent model for anomaly identification, though its practical application would benefit from integration with other geophysical, geodetic, and geological data to further reduce the false alarm burden.

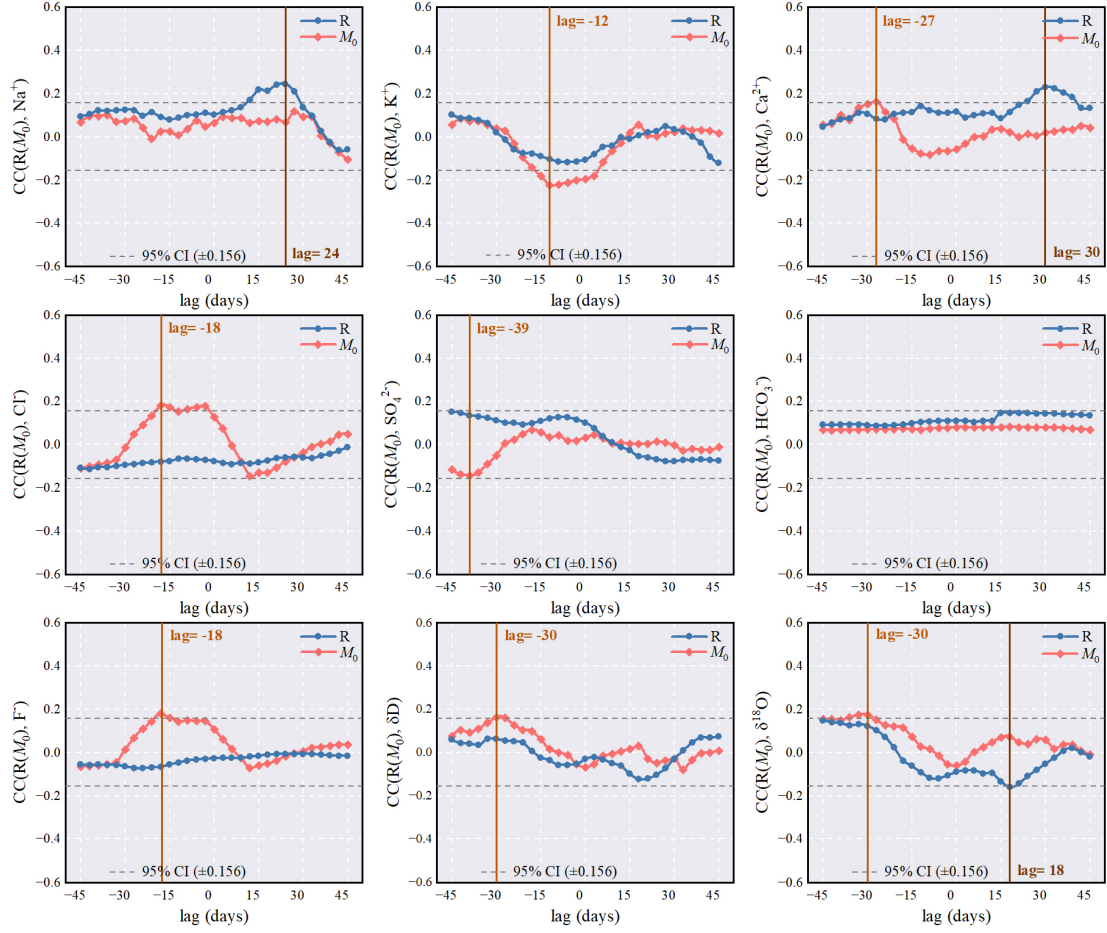


**Figure S7: Statistical significance and model performance reliability validation. The histograms illustrate the TS distribution of the stochastic baseline model (1,000 Monte Carlo iterations), while the red dashed lines mark the actual performance of the proposed model at sites Quijiang (QJ) and Wana (WN) springs, respectively.**

2. *#The selection of the 45-day response time threshold ( $T_{th}$ ) is not well justified. The manuscript identifies this threshold as “optimal” based on cross-correlation peaks,*

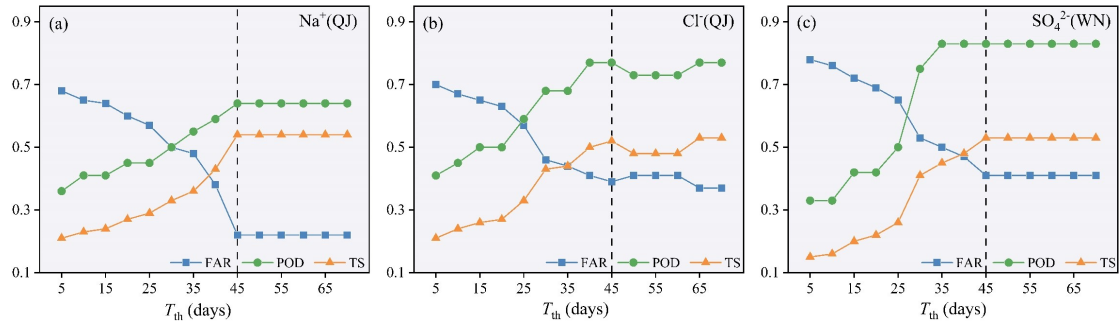
*BCP detection windows, and metric plateau behavior. However, the reported cross-correlation values are weak (approximately  $\pm 0.2$ ) and no statistical significance testing (e.g.,  $p$ -values or confidence intervals) is provided. Furthermore, the plateau behavior of evaluation metrics is not a robust criterion for determining optimality, and no sensitivity analysis is conducted to evaluate whether this threshold remains stable across different datasets or regions. Therefore, the conclusion that 45 days is optimal appears empirical and potentially site-specific rather than generalizable.*

**Response:** We are grateful to the reviewer for the insightful comments regarding the selection logic of the  $T_{th}$  threshold. We agree that a robust threshold requires both statistical validation and an evaluation of its sensitivity, and have accordingly updated Figure 5 to include a 95% Confidence Interval (CI). While the absolute cross-correlation values are relatively modest (around  $\pm 0.2$ ), the revised figure demonstrates that the peaks for key components clearly exceed the 95% CI threshold of  $\pm 0.156$ . In the field of earthquake forecasting, a correlation of  $\pm 0.2$  carries significant reference value for capturing weak precursor signals, and the fact that these peaks stand above the CI ( $p < 0.05$ ) confirms a non-random temporal link. Consequently, these lag times are statistically credible and are consistently contained within the 45-day window.



**Figure 5: Cross-correlation function analysis of the 15-day moving average time series of hydrochemical components during stable isotope monitoring, with rainfall and distance-corrected seismic moment.**

Furthermore, Figure S6 in the Supplementary Materials provides a detailed sensitivity analysis of the evaluation metrics (FAR, POD, and TS) across different  $T_{th}$  values. The selection of 45 days is based on a deliberate trade-off between model performance and temporal precision. As shown in Figure S6, increasing  $T_{th}$  from 5 to 45 days yields a sharp improvement in POD and TS while reducing FAR; beyond this point, the metrics reach a plateau where further increasing the window does not yield significant performance gains but would decrease the temporal resolution of the forecast. Thus, 45 days represents the key threshold of optimal model performance for this dataset.



**Figure S6: Variations in the false alarm rate (FAR), probability of detection (POD), and threat score (TS) as a function of the seismic response time threshold ( $T_{th}$ ).**

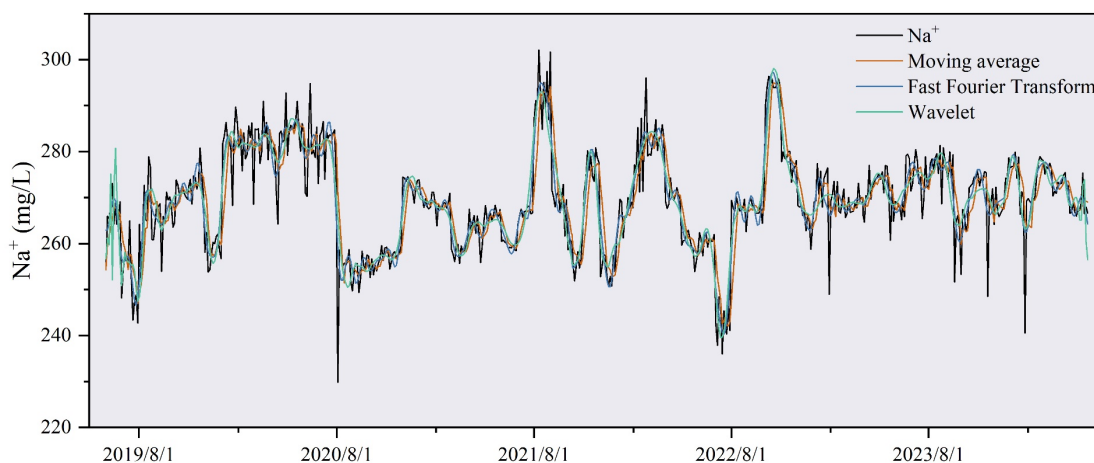
Finally, we acknowledge the reviewer’s point regarding generalizability and have revised the Abstract and Conclusions to clarify that this threshold is likely more applicable to the specific hydrogeological conditions of the study region.

3. *#The manuscript assumes a causal relationship between hydrochemical anomalies and earthquakes without adequately addressing confounding factors. Although rainfall effects are considered, the filtering approach is limited to a 15-day moving average, which may not sufficiently remove environmental noise. There is no application of multivariate statistical methods or causal inference frameworks to disentangle seismic signals from other influences. Moreover, anthropogenic effects are assumed negligible without quantitative evidence, and the correlation between hydrochemical variations and seismic moment is weak (around  $\pm 0.2$ ). These limitations raise concerns about whether the detected anomalies truly represent seismic precursors or are partially driven by non-seismic processes.*

**Response:** We sincerely appreciate the reviewer’s insightful comments. Distinguishing seismic precursors from environmental noise remains a fundamental challenge in hydrogeochemical monitoring. We have addressed these concerns by elaborating on the effectiveness of our filtering approach, the integrative logic of our synergistic detection model, and the environmental characteristics of our sampling sites.

While the 15-day moving average effectively reduces high-frequency noise related to rainfall pulses, we agree that additional validation strengthens methodological

robustness. Accordingly, we have further applied both Fast Fourier Transform (FFT) low-pass filtering and a 3-level wavelet denoising technique (using DB5 wavelet with 20% threshold) to the same dataset (Figure S5, using the  $\text{Na}^+$  series from Qujiang as an example). The results demonstrate approximately consistent signal smoothing across all three methods, thereby reinforcing the reliability of the moving average approach for denoising process. Although the backward moving average introduces a slight phase lag, this method is better suited to the real-time anomaly detection framework of our detection model. Furthermore, cross-correlation analysis between the processed datasets and environmental variables (e.g., rainfall) shows a significant reduction in correlation, validating the effectiveness of our denoising process. Although we acknowledge that eliminating complex environmental noise is inherently difficult, our current workflow maximizes the signal-to-noise ratio for potential pre-earthquake signals.



**Figure S5: Comparison of denoising results using 15-day moving average, Fast Fourier Transform, and 3-level Wavelet (using DB5 wavelet with 20 % threshold).**

Regarding the suggestion for multivariate statistical methods or causal inference, the core of our study relies on a “synergistic detection model” for multiple components to achieve a similar objective. Instead of relying on a single ion’s fluctuations, we utilize the synchronized temporal anomalies across multiple hydrochemical components to filter out stochastic noise and single-factor interference. This synergistic approach functions as a multivariate cross-validation mechanism. Although the linear correlation between seismic moment and individual

components is around  $\pm 0.2$ , such values are statistically meaningful within the highly complex and heterogeneous crustal systems. Our primary intent is to capture atypical signals that evolve synergistically with the earthquake preparation process, particularly when conventional precursor identification is challenging.

Regarding the lack of quantitative evidence for anthropogenic effects, it is important to note that our sampling sites are remote thermal springs with continuous natural flow. The hydrogeochemical characteristics of such flowing systems are generally resilient to localized surface activities. Given the geographical isolation and the dynamic nature of these springs, we considered anthropogenic influences to be negligible. While logistically challenging to provide real-time quantitative data on human activity in these remote areas, our long-term observations show no discernible characteristic of human interference.

We would like to express our sincere gratitude to the reviewer for the forward-looking suggestions regarding causal inference frameworks and multivariate statistical methods. While this study focuses on the synergistic detection of anomalies, future research will incorporate advanced causal inference frameworks (e.g., SEM or CCM) to further quantify the coupling between seismic stress and hydrochemical variations.

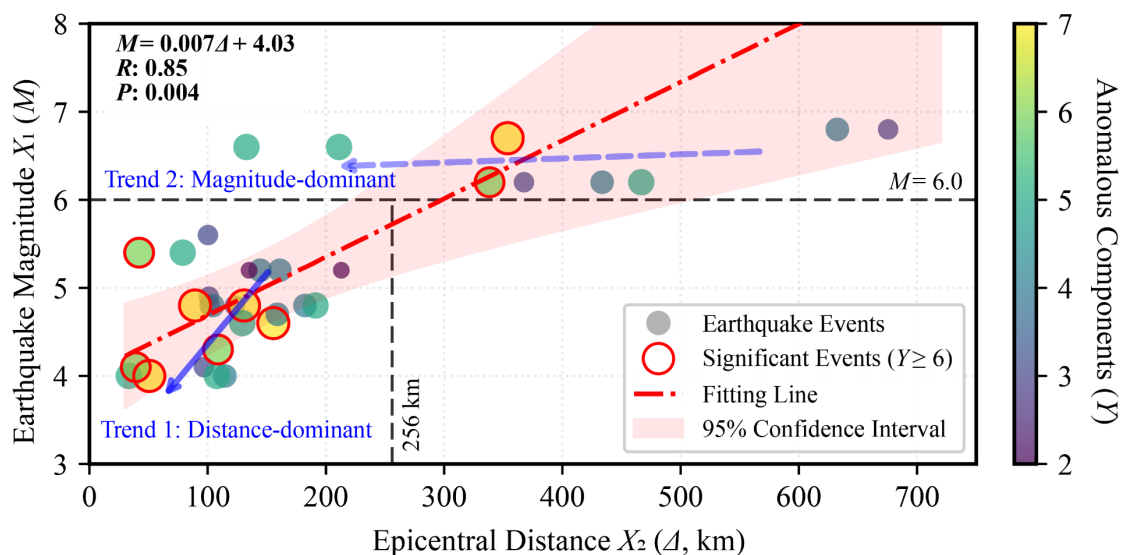
4. *#The definition of the multicomponent alarm criterion (i.e., at least three components showing synchronous anomalies) lacks theoretical justification. This threshold plays a central role in controlling model outputs, yet it appears to be heuristic. The manuscript does not provide statistical reasoning (e.g., false discovery rate control), nor does it include sensitivity analyses comparing alternative thresholds (e.g.,  $\geq 2$ ,  $\geq 4$ ,  $\geq 5$  components). Additionally, no probabilistic framework (such as Bayesian fusion or likelihood-based integration) is used to combine multi-component information. As a result, the chosen rule may bias the results and reduce reproducibility.*

**Response:** We sincerely thank the reviewer for this critical and constructive comment. We agree that the selection of the number of synchronous anomalous

components ( $N$ ) requires both statistical evidence and a hydrochemical rationale. In response, we have performed a comprehensive sensitivity analysis and expanded the theoretical justification as follows:

We have added a new supplementary table (Table S5) comparing model performance across different counts of synchronous anomalous components ( $N \geq 2$  to  $N \geq 6$ ). The results demonstrate that  $N \geq 3$  yields the optimal TS for both Qujiang (TS = 0.70) and Wana (TS = 0.59) springs. This provides a data-driven justification for our choice, as  $N \geq 3$  represents the optimal key point that maximizes the balance between the POD and the FAR.

We have clarified the theoretical basis for  $N \geq 3$  as a denoising mechanism. As manuscript on line 290, external disturbances (e.g., contamination or anthropogenic inputs) typically follow the charge balance principle, which often induces synchronous changes in only paired cations and anions ( $N = 2$ ). By requiring at least three synchronous components, the model effectively filters out these dual-component fluctuations, thereby controlling the FAR more robustly than an  $N \geq 2$  criterion.



**Figure 11: Relationship between the number of synchronous anomalous hydrochemical components, earthquake magnitude, and epicentral distance.**

The criterion  $N \geq 3$  serves as the foundation of a graded warning system. As shown in Table S5, while  $N \geq 3$  provides the optimal performance balance, further increases in  $N$  (4, 5, 6) correspond to higher signal intensity. This allows the model to categorize

the severity of the anomaly, where a higher N indicates a more intense tectonic disturbance. Our results (Figure 11) reveal that the number of synchronous anomalous components correlates with earthquake magnitude or epicentral distance. Consequently, the number of components exhibiting synchronous anomalies can be used as an effective criterion for forecasting, with higher count of anomalous components generally corresponding to larger earthquake magnitudes or shorter epicentral distances.

We have revised manuscript on line 435:

The model activates comprehensive alarms when anomalies are detected in three or more components simultaneously. Based on the sensitivity analysis results presented in Table S5, requiring three or more anomalous components was identified as the preferred criterion for balancing the POD and the FAR. Furthermore, this choice is hydrochemically justified by its ability to suppress random dual-component interferences stemming from a single source, as discussed previously, thereby leading to the highest TS among all tested numbers of components.

**Table S5: Model performance comparison across different numbers of synchronous anomalous components.**

Components (N)	Qujiang spring				Wana spring			
	POD	FAR	MAR	TS	POD	FAR	MAR	TS
$N \geq 2$	1	0.37	0	0.63	1.00	0.43	0	0.57
$N \geq 3$	0.95	0.27	0.05	0.70	0.83	0.33	0.17	0.59
$N \geq 4$	0.77	0.29	0.23	0.59	0.75	0.31	0.25	0.56
$N \geq 5$	0.55	0.29	0.45	0.44	0.67	0.20	0.33	0.57
$N \geq 6$	0.36	0.11	0.64	0.35	0.58	0.13	0.42	0.54

5. *#Model validation is insufficient and limits the generalizability of the conclusions. The model is evaluated only at two sites (QJ and WN) with relatively short*

*monitoring periods, and no independent validation is performed. There is no train-test split, no forward prediction test to simulate real forecasting conditions, and no spatial validation using external datasets. Furthermore, comparisons with existing machine learning or statistical approaches remain qualitative rather than quantitative. Therefore, the claim that the model can be “universally applied” is not adequately supported by the current analysis.*

**Response:** We sincerely thank the reviewer for these insightful comments regarding model validation and the generalizability of our findings. We completely agree that rigorous independent validation and quantitative comparisons represent the benchmark standard for evaluating predictive models.

Regarding the lack of a train-test split, we would like to clarify the objective constraints of our dataset. The decision to use the entire dataset was primarily driven by its limited size, a common challenge in earthquake monitoring applications. Acquiring additional data is often prohibitively expensive or operationally infeasible, particularly for rare target events such as earthquakes. Specifically, for one hydrochemical component analyzed over a five-year monitoring period, the dataset consists of only 604 samples, which include just 22 potential earthquake events. Under such constraints, partitioning the data into a hold-out test set could easily result in a highly unrepresentative subset that might contain 0 earthquake events, thereby rendering performance evaluation unreliable or meaningless. Using the entire dataset enables a more stable and meaningful evaluation of model performance.

To further mitigate the risk of overfitting given our limited sample size, we avoided complex models or deep learning architectures that require massive datasets. Instead, we developed a parsimonious model with a limited number of parameters, in accordance with industry constraints. The model framework we adopted was designed to incorporate seismological interpretability and empirical predictive experience where feasible, thereby enhancing practical applicability and reducing the model’s propensity to overfit compared to more complex models.

Furthermore, we acknowledge the reviewer’s concern that the term “universally applied” was an overstatement. Our original intention was to emphasize the transferability of the methodological framework rather than claiming the model itself is a universal model. As discussed in the “Limitations and prospects” section, the effective transfer of this model depends on specific criteria, including site selection in active tectonic regions, the use of thermal springs with high discharge temperatures to minimize meteoric interference and so on.

We have revised manuscript on line 28:

This model **has the potential to be applied to** thermal spring monitoring across diverse **active tectonic regions** through targeted parameter optimisation, offering a valuable reference for earthquake forecasting.

**# Minor comments:**

1. #Line 3–4, “*a synergetic detection model*” → better use “*a synergistic detection model*”.

**Response:** Thank you for the suggestion. We have changed “a synergetic detection model” to “a synergistic detection model” on lines 1–2.

2. #Line 15–16, “*encounters intense tectonic activity*” → awkward, use “*experiences*”.

**Response:** Thanks! We have changed “encounters” to “experiences” on lines 15.

3. #Line 19, “*the occurrence timing for  $M \geq 4$  earthquakes*” → “*the timing of  $M \geq 4$  earthquakes*”.

**Response:** Thanks! We have removed “occurrence” as suggested to make the phrasing more concise.

4. #Line 24–25, “*which can serve as sensitive indicators*” → unclear subject, revise.

**Response:** Thank you. We have revised the sentence to explicitly define the subject. The unclear “which” has been replaced with “these components” to improve clarity.

5. #Line 26–27, “using the number of hydrochemical components...” → *missing subject consistency.*

**Response:** Thanks! We have revised the sentence to ensure subject consistency. By using the “where” clause and splitting the statement into two sentences, we have clarified that the “number of components” is the primary subject and the criterion for forecasting.

6. #Line 27, “componentstypically” → *typo (missing space).*

**Response:** Thanks! We have carefully performed a full-text search and checked all figures and tables for the string “componentstypically”. However, we were unable to locate this specific typo in the current manuscript. Nevertheless, we have re-checked the entire manuscript for any similar spacing or formatting issues to ensure technical accuracy.

7. #Line 33, “relies on detecting sensitive and reliable precursor anomalies” → *redundant wording.*

**Response:** Thank you for the observation. While seemingly similar, “sensitive” and “reliable” address different performance metrics (sensitivity and precision) in the context of seismic monitoring. Therefore, we have opted to retain both terms.

8. #Line 37–38, “respond rapidly to deviations” → *unclear phrasing.*

**Response:** Thanks! We have revised “deviations” to “variations” to better reflect the stress loading process during the seismogenic period.

9. #Line 49–50, “remains a gigantic challenge” → informal, use “a major challenge”.

**Response:** Thanks! We have revised “gigantic” to “major”.

10. #Line 54–55, “display convoluted nonlinear dynamic behaviors” → overly wordy.

**Response:** Thank you for the suggestion. We have simplified the phrase to “exhibit complex nonlinear fluctuations” to make the sentence more concise.

11. #Line 148–150, Equation (1) formatting incorrect and unclear.

**Response:** Thanks! We have thoroughly revised Equation (1) to address the formatting and clarity issues on line 144:

To ensure data accuracy, cation–anion **electrical balance (EB)** tests were performed for each sample as a quality control measure, with all **EB** kept within  $\pm 5\%$ . Data that fulfilled this criterion were included in the subsequent analysis. The **EB** (Appelo and Postma, 2005) is calculated as below:

$$EB = \frac{\sum m_c - \sum m_a}{\sum m_c + \sum m_a} \times 100\% \quad (1)$$

where  $\sum m_c$  represents the sum of cation concentrations (in milliequivalents per liter, meq/L), and  $\sum m_a$  represents the sum of anion concentrations.

12. #Line 156–157, missing variable definition in Eq. (2).

**Response:** Thank you for the observation. We have updated the text accompanying Equation (2) to provide a more rigorous definition of the variables.

13. #Line 175–176, “ $\omega$  is the weighting factor” → no physical explanation.

**Response:** We appreciate the reviewer’s comment regarding the physical

interpretation of  $\omega$ . Following the methodology established by Piersanti et al. (2016),  $\omega$  is employed as a weighting factor that characterizes the relationship between the seismic moment ( $M_0$ ) and the epicentral distance ( $\Delta$ ). While it originates from an empirical formulation, it functionally serves to modulate the relative contribution of distant and local earthquakes to the observed hydrogeochemical variations. We have retained this original designation to maintain consistency with the cited literature and revised the manuscript on line 181:

where  $\omega$  is the weighting factor to modulate the relative contribution of distant and local earthquakes to the observed hydrogeochemical variations.

14. #Line 218–220, Eq. (5) unclear notation (MA definition inconsistent).

**Response:** Thanks! We have corrected Equation (5) by introducing a proper summation index ( $i$ ) to distinguish the moving window's range from the specific time point  $t$ . The revised notation now clearly defines the backward moving average calculation. On line 222:

$$MA(t) = \frac{1}{15} \sum_{i=t-14}^t Dr(i) \quad (5)$$

where  $MA(t)$  is the 15-day moving average at day  $t$ , and  $Dr(i)$  is the daily raw data on day  $i$ .

15. #Line 230–231, Eq. (6) incomplete expression.

**Response:** Thanks! We have revised Equation (6) to provide a complete and normalized expression of the cross-correlation function. On line 231:

The cross-correlation function (Chatfield, 1975; Brockwell and Davis, 1991) is represented as follows:

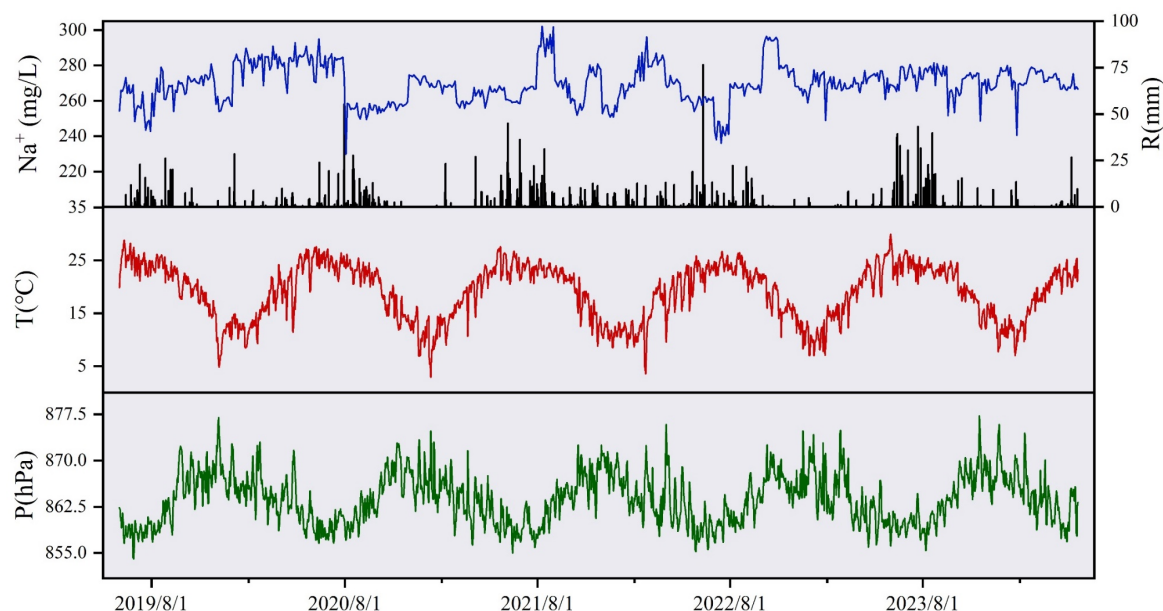
$$CC_{xy}(k) = \frac{\sum_{t=1}^{N-|k|} (x_t - \bar{x})(y_{t+k} - \bar{y})}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2 \sum_{t=1}^N (y_t - \bar{y})^2}} \quad (6)$$

where  $x$  and  $y$  are two time series,  $\bar{x}$  and  $\bar{y}$  represent their sample means,  $N$  is

the series length, and  $k$  is the lag. The lag  $k$  denotes the time shift, where  $k > 0$  indicates  $y$  lags  $x$ , and  $k < 0$  indicates  $x$  lags  $y$ . Considering the seasonal effects of rainfall and the reliability of seismic precursor response times,  $k$  is set within a range of  $-45$  to  $45$  days.

16. #Line 205–210, statement “temperature and pressure negligible” lacks quantitative support.

**Response:** We appreciate the reviewer’s constructive feedback. To provide quantitative support, we have revised the manuscript to explicitly refer to the observational evidence in Figure S3.



**Figure S3: Time series of  $\text{Na}^+$ , alongside corresponding rainfall (R), temperature (T), and atmospheric pressure (P) for Qujiang spring.**

The long-term monitoring shows that while ambient temperature and atmospheric pressure exhibit significant seasonal cycles, the hydrochemical concentrations remain stable without corresponding periodic fluctuations. This lack of correlation, contrasted with the sharp, pulsed response following rainfall events, demonstrates that the impacts of temperature and pressure are relatively minor in this system. We have revised manuscript on line 205:

The thermal spring water in the study area originates from atmospheric precipitation recharge. It circulates deeply through faults, is heated by geothermal energy, and then discharges at the surface, with its hydrochemical composition primarily determined by the lithology of the surrounding rocks (Shao et al., 2024). **Due to the deep-circulation nature of this hydrothermal system**, ambient temperature and atmospheric pressure at the spring outlet exert a **relatively minor effect on the hydrochemistry**. However, rainfall serves not only as the primary water source but also accelerates groundwater circulation, fosters shallow infiltration, and mixes with thermal waters (Taylor et al., 2012; Hosono et al., 2020; Colman et al., 2021). This process can potentially obscure deep-seated earthquake preparatory signals carried by the thermal spring. The comparative analysis of hydrochemical and meteorological data (Figure S3) reveals that rainfall is the dominant interfering factor, accompanied by a partial decrease in ion concentrations following events. **In contrast, as shown in Figure S3, the significant seasonal fluctuations in ambient temperature and atmospheric pressure do not induce corresponding periodic variations in the hydrochemical concentrations.** Consequently, this study focuses on assessing the potential perturbations induced by rainfall on thermal spring hydrochemistry.

17. *#Line 233–235, moving average justification too weak (no comparison metrics).*

**Response:** We appreciate the reviewer’s feedback. In fact, we have already provided quantitative metrics in the manuscript to justify the moving average approach. As shown in Figure 5, the cross-correlation coefficients between the denoised series and rainfall are within  $\pm 0.2$ , which confirms the effectiveness of the denoising. Additionally, Figure S5 provides a comparative analysis with FFT and Wavelet methods, showing that MA yields consistent results while being more suitable for our real-time detection framework.

18. *#Line 243–245, “appear to adhere to specific distribution patterns” vague*

*statement.*

**Response:** Thanks! We have revised the manuscript on line 244 to clarify that the analysis relies on the detection of shifts in these statistical properties.

Assuming that the denoised hydrochemical component time series [exhibit stable statistical characteristics](#).

19. #Line 270–274, *algorithm description overly dense and difficult to follow.*

**Response:** We sincerely appreciate the reviewer’s suggestion to improve the clarity of the algorithm description. Following this advice, we have carefully revised the manuscript on line 274:

The improved detection algorithm (Figure 4) is described as follows: [real-time monitoring data for each component are loaded on day  \$i\$ . An alarm is triggered on day  \$i\$  if the daily value on day  \$i-p\_2\$  exceeds  \$p\_1\$  times the 15-day moving average on day  \$i-p\_2-1\$ , and simultaneously, the 15-day moving average on day  \$i\$  surpasses  \$p\_3\$  times that on day  \$i-p\_2\$ . This alarm is considered a successful early warning if an  \$M \geq 4\$  earthquake occurs within  \$i+T\_{th}\$ . Following an earthquake, the algorithm enters a threshold adjustment period to suppress redundant false alarms caused by ongoing post-earthquake anomalies. \[During this period of  \\$p\\_5^M\\$  days \\(where  \\$M\\$  is the magnitude\\), the parameters are updated to  \\$p\\_1' = p\\_4 \times p\\_1\\$  and  \\$p\\_3' = p\\_4 \times p\\_3\\$ . The functional form  \\$p\\_5^M\\$  is adopted based on an exponential relationship between earthquake magnitude and geochemistry anomaly duration \\(Fleischer et al., 1985; Rikitake, 1988; Ilić et al., 2005; Elmaghraby et al., 2009\\). Treating the base  \\$p\\_5\\$  as a tunable parameter allows the algorithm to dynamically optimize this duration for specific regional data. \\[If subsequent earthquakes occur within this period, the adjustment period is recalculated based on the maximum magnitude of the subsequent earthquakes.\\]\\(#\\) If no  \\$M \geq 4\\$  earthquakes occur during this period, parameters  \\$p\\_1\\$  and  \\$p\\_3\\$  automatically revert to their initial values. The algorithm incorporates five adjustable parameters \\( \\$p\\_1\\$ – \\$p\\_5\\$ \\) and a seismic response time threshold \\( \\$T\\_{th}\\$ \\), with  \\$p\\_1\\$ ,  \\$p\\_3\\$ , and  \\$T\\_{th}\\$  being key parameters.\]\(#\)](#)

20. #Line 292–293, “at least three components” → no justification.

**Response:** Thanks! As this point aligns with our response to Comment #4, we have addressed this issue in detail in that section.

21. #Line 310–312, evaluation metrics description redundant.

**Response:** We thank the reviewer for the suggestion. To eliminate redundancy, we have streamlined the description of the evaluation metrics by defining the fundamental variables at the beginning and removing repetitive phrasing for each formula. The revised manuscript now focuses more concisely on the functional significance of these metrics in handling imbalanced datasets. On line 297:

To evaluate the model’s performance in identifying low-probability earthquake events within imbalanced datasets, four quantitative metrics are applied based on the number of correct alarms (NA), false alarms (NB), and missed alarms (NC):

False alarm rate (FAR): Measures the proportion of incorrect warnings relative to the total number of alarms issued.

$$FAR = NB/(NA + NB) \quad (7)$$

Missed alarm rate (MAR): Quantifies the risk of undetected seismic events.

$$MAR = NC/(NA + NC) \quad (8)$$

Probability of detection (POD): Assesses the model’s sensitivity in identifying actual events.

$$POD = NA/(NA + NC) \quad (9)$$

Threat score (TS): A comprehensive accuracy index that integrates NA, NB, and NC to provide a holistic evaluation of forecasting performance.

$$TS = NA/(NA + NB + NC) \quad (10)$$

22. #Line 325–330, “effectively mitigates rainfall-induced noise” → overstatement given weak correlation.

**Response:** Thanks! We apologize for any confusion caused by our initial phrasing. The reviewer may have misunderstood our logic: the “weak correlation” is not a sign of a failed analysis, but rather the target objective of our noise mitigation process. An effective noise-reduction method should result in a processed time series that is decoupled from rainfall events. Therefore, the achieving of a weak correlation coefficient after the 15-day moving average treatment directly demonstrates that the rainfall-induced variance has been effectively suppressed, ensuring that the remaining anomalies are more likely to be of seismic or tectonic origin. To clarify this point for the general reader, we have revised the manuscript to explicitly state that the reduction in correlation is the indicator of effective noise removal. On line 321:

The cross-correlation analysis findings (Figure 5) show that after the 15-day moving average treatment, the correlations (blue dotted lines) between rainfall and hydrochemical components are reduced to a weak level (within  $\pm 0.2$ ). This result indicates that the moving average treatment effectively mitigates rainfall-induced noise.

23. #Line 345–350, retaining low-probability signals (e.g., 0.06) is questionable.

**Response:** Thanks! We agree with the reviewer that very low-probability signals may lack statistical conviction. Accordingly, we have revised the manuscript to replace the previous example with a more robust case: the *M*4.6 earthquake on 3 March 2023. In this revised version, the posterior probabilities for  $\text{Ca}^{2+}$  (0.20) and  $\text{SO}_4^{2-}$  (0.19) show more distinct deviations. By focusing on these higher-probability, synchronized signals across multiple components, we believe the geochemical evidence for pre-seismic anomalies is now more clearly and reliably presented in the updated manuscript. On line 342:

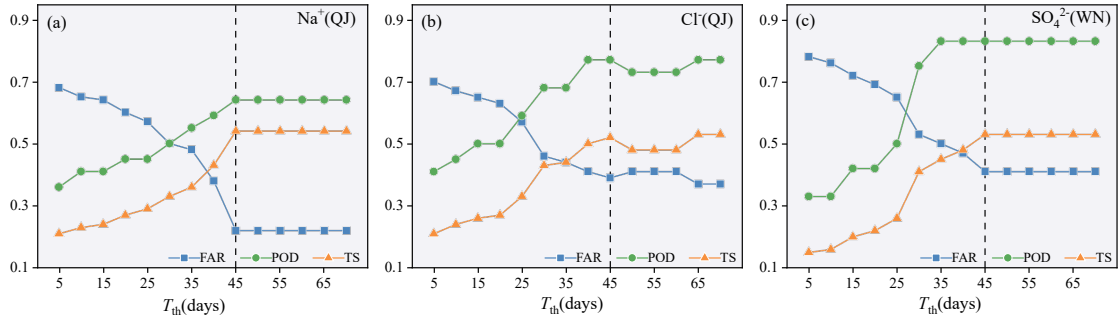
For example, before the *M*4.6 earthquake on 3 March 2023, the posterior probability for a  $\text{Ca}^{2+}$  change point at QJ was 0.20 at 20 days before the earthquake, while  $\text{SO}_4^{2-}$  exhibited a posterior probability of 0.19 at 28 days before the earthquake.

24. #Line 379–381, *limitation of BCP is correctly stated but not addressed methodologically.*

**Response:** We sincerely appreciate the reviewer’s insightful observation regarding the gap between the identified limitation and its methodological resolution. Regarding the methodological treatment, we would like to clarify that the BCP analysis in our framework is intentionally positioned as a supplementary tool rather than a primary predictive component for real-time forecasting. Its primary role is to provide a reliable statistical benchmark for anomaly timing localization and parameter optimization through retrospective analysis. However, we acknowledge that addressing this limitation methodologically is crucial for transitioning from scientific discovery to practical application. In future research, we plan to adapt this retrospective framework into an “online” or sliding-window detection mode, which will be a key step toward practical forward-forecasting.

25. #Line 395–400, *“turning point” interpretation lacks statistical test.*

**Response:** We appreciate the reviewer’s perspective on the statistical rigor of threshold identification. The 45-day threshold was determined through a systematic sensitivity analysis (5 to 70 days), a standard procedure for parameter optimization in anomaly detection models. Rather than being derived from a single data point, this “turning point” is evidenced by the synchronized convergence of three independent metrics (FAR, POD, and TS) across multiple hydrochemical components ( $\text{Na}^+$ ,  $\text{Cl}^-$ , and  $\text{SO}_4^{2-}$ , as shown in Figure S6). This consistent plateau demonstrates that the model reaches a stable optimum beyond 45 days.



**Figure S6: Variations in the false alarm rate (FAR), probability of detection (POD), and threat score (TS) as a function of the seismic response time threshold ( $T_{th}$ ).**

We have revised the manuscript to clarify that the 45-day threshold is a systematic inflection point observed consistently across tested hydrochemical components. On line 389:

This increase determines statistically key inflection points [across multiple hydrochemical components](#) during threshold optimisation (Figure S6).

26. #Line 410–412, parameter range selection unclear (why 1.00–1.20?).

**Response:** Thanks! The optimization range for parameters p1 and p3 is governed by both the model’s operational logic and its functional detection boundaries.

The lower bound of 1.00 is a logical prerequisite. Since p1 and p3 function as multipliers to define deviations from the sliding window baseline, a value exceeding 1.00 is essential to identify a “positive” anomaly. Setting these parameters below 1.00 would erroneously trigger alarms for fluctuations falling below the average, which is physically inconsistent with the target earthquake hydrochemical precursors. The upper bound of 1.20 was determined through iterative optimization; preliminary tests consistently showed that model performance degraded significantly beyond this value due to an excessive MAR. Consequently, this range effectively encapsulates the optimal performance window while maintaining the physical relevance of the seismic anomaly detection.

In summary, the lower bounds for parameters are defined by the model’s foundational logic, while the upper bounds are determined by actual model

performance test results and fundamental constraints of seismology. We have revised the manuscript on line 403:

The parameter ranges were based on specific seismological rationale, operational framework of model, and empirical performance limits. Specifically, the lower bound is a logical prerequisite for identifying positive anomalies, while the upper bound was determined through iterative testing.

27. #Line 440–445, *high FAR acknowledged but not mitigated.*

**Response:** We appreciate the reviewer’s comment regarding the FAR. The mitigation of FAR is intrinsically addressed in the manuscript through the multicomponent synergistic framework, which leverages the consistency of multiple geochemical indicators to filter out non-seismic noise, as evidenced by the improved TS in Figures 8–10. Furthermore, we explicitly discuss the integration of multidisciplinary data (geophysical and geodetic) as a practical strategy to further suppress the false alarm burden. In the context of earthquake forecasting, maintaining high sensitivity (POD) is prioritized, and these collaborative mechanisms provide the necessary balance to ensure the model’s reliability.

28. #Line 465–470, *“TS around 0.50” claimed as good without benchmark.*

**Response:** We appreciate the reviewer’s suggestion. After a thorough review of the current literature, we would like to clarify our position as follows. To our knowledge, there is currently no universally accepted “standard benchmark” for TS values in hydrogeochemical earthquake forecasting. The performance of such models is highly region-specific and data-dependent. In the absence of an external baseline, we employed an internal comparative approach. By calculating the TS values for all monitored components (as shown in Figure 10), we identified those with relatively higher performance (around 0.50–0.80) as sensitive indicators. This selection is statistically sound, as it distinguishes high-performing indicators from low-performing

ones. In seismic event detection, the prior probability of an event is extremely low. Under these conditions, a TS of 0.50 represents a result where the hits significantly outweigh the false alarms and misses, which is a performance level far superior to the stochastic baseline.

29. #Line 500–503, “robust linear correlation ( $R = 0.85$ )” → *sample size not reported.*

**Response:** We appreciate the reviewer’s suggestion. The “significant events ( $Y \geq 6$ )” mentioned in the manuscript indeed refer to the 9 specific data points highlighted with red circles in Figure 11. To ensure statistical transparency, we have now explicitly included the sample size ( $n = 9$ ) in the revised manuscript. Given the high correlation coefficient ( $R = 0.85$ ) and the low p-value ( $P = 0.004$ ), this relationship remains statistically significant even with the specific sample size of 9. On line 525:

Among these significant events ( $Y \geq 6$ ), a robust linear correlation exists between magnitude and distance ( $n = 9$ ,  $R = 0.85$ ,  $P = 0.004$ ).

30. #Line 540–542, “accuracy exceeding 83%” *unclear definition (not standard metric).*

**Response:** Thanks! To ensure statistical rigor and follow standard practices in anomaly detection, we have replaced “accuracy” with Probability of Detection (POD) in the revised manuscript.

31. #*Inconsistent use of isotope notation ( $\delta D$ ,  $\delta 18O$ ,  $\delta^{18}O$ ).*

**Response:** Thanks! We have conducted a thorough search for isotope notations throughout the manuscript. While we did not find a mixture of all the mentioned formats, we realized there was an inconsistency between “ $\delta^2H$ ” and the standard “ $\delta D$ ”. We have now standardized all isotopic notations to  $\delta D$  and  $\delta^{18}O$  throughout the manuscript to ensure rigorous formatting.

32. *#Inconsistent spacing for ions (Cl<sup>-</sup> vs Cl -).*

**Response:** Thanks! We have carefully re-examined all ion notations throughout the manuscript. While no instances of inconsistent spacing (such as “Cl -”) were identified in the current version, we have performed a global “find and replace” to ensure that all ionic charges are correctly formatted as superscripts immediately following the element symbol without any intervening spaces.

33. *#“M ≥ 4” sometimes written as “M≥4”.*

**Response:** Thank you for the reminder. We have carefully re-checked the entire manuscript and confirmed that all expressions (e.g., “M ≥ 4”) consistently include spaces around the operators. While we did not identify any instances where spaces were missing around operators, we have performed an additional review to ensure full compliance with this formatting standard.

34. *#Units formatting inconsistent (μS/cm vs mg/L spacing).*

**Response:** Thanks! While we have maintained spaces between values and units throughout the manuscript, we noticed that slight visual variations existed due to different character shapes. We have now re-verified and standardized all these intervals throughout the manuscript to ensure absolute consistency in formatting.

35. *#References formatting inconsistent (some missing spaces, punctuation).*

**Response:** Thank you for the reminder. We have conducted a thorough audit of the reference list and corrected all formatting inconsistencies, including missing spaces and punctuation errors, to ensure full compliance with the journal’s style guidelines.