RESPONSE LETTER

Manuscript Number: egusphere-2026-1378

Title: Managed Aquifer Recharge in Confined Multi-Layer Aquifers: A Scalable Framework for Drought Resilience in Central Europe

Corresponding Author: Abdelrahman Ahmed Ali Abdelrahman
All Authors: Abdelrahman Ahmed Ali Abdelrahman, Hagen Koch, Mobarok Hossain, Ronjon Heim, Clara Hauke, and Irina Engelhardt

<u>Response to the Reviewer #1:</u>

This is an interesting study addressing a real gap in MAR research for confined aquifers in Central Europe. The integration of three methodological components into a single transferable framework is a good contribution. The manuscript is generally well-written and clearly structured. However, several issues require attention before publication.

Reviewer's comments:

1) The authors report a high pearson correlation between SPEI-12 and SGI with "no lag". This is an interesting finding for a confined multi-layer aquifer system, where pressure propagation rather than direct recharge drives head changes. The manuscript does not sufficiently explain the hydrogeological mechanism behind this near-instantaneous coupling. In confined systems, piezometric responses can be rapid, but this is a different mechanism than moisture storage response observed in unconfined systems. This distinction must be clearly discussed and reconciled with the literature.

It would be more correct, rather to state that the absence of lag between SPEI 12 and SGI "indicates that groundwater levels in this confined aquifer respond almost immediately to climatic anomalies" (Discussion section, lines 487-488) to state that the time scale of the response process of groundwater levels to climatic anomalies is not detectable by the aggregated annual scale of SPEI 12 compared with SGI.

Due to the importance of this statement in discussion and conclusion section, to avoid misinterpreting of results and findings, a montly based analisys using SPEI 1 would be recommended to eventualy detect and quantify lags or strongly confirm the "immediate response of GW to climate variability" statements of the whole manusccript.

Authors' response

We thank the reviewer for this important and constructive comment. We agree that the previous wording could be misinterpreted. We have changed the wording to avoid misinterpretation, clarifying that the absence of a detectable lag at the SPEI-12 scale does not demonstrate an immediate physical recharge response. Rather, it indicates that the response time of groundwater levels to climatic anomalies is not resolved by the 12-month accumulated SPEI timescale. Because SPEI-12 already integrates the climatic water balance over the preceding year, a zero-lag correlation means that SGI co-varies with annual-scale climatic anomalies, not that groundwater storage responds instantaneously.

Following the reviewer's suggestion, we performed a monthly lag-correlation analysis using SPEI-1, SPEI-3, SPEI-6, and SPEI-12 against SGI for lags of 0–24 months. The SPEI-1 analysis showed only weak correlations with SGI ($r = 0.108$ at 3-month lag), whereas correlations increased for longer accumulation periods and were strongest for SPEI-12 at zero lag ($r = 0.632$, $r^2 = 0.40$). This supports the interpretation that the groundwater system primarily reflects accumulated climatic anomalies over multi-month to annual timescales rather than individual monthly anomalies (please see new Table S2 in Supplementary Materials).

We have also added a hydrogeological explanation specific to the confined aquifer setting. In this system, SGI represents piezometric head variations. These may respond rapidly to pressure propagation and changes in boundary conditions, but this mechanism differs from direct recharge or soil-moisture storage responses typical of unconfined aquifers. We therefore revised the Discussion and Conclusion to distinguish between piezometric response and direct groundwater recharge/storage response.

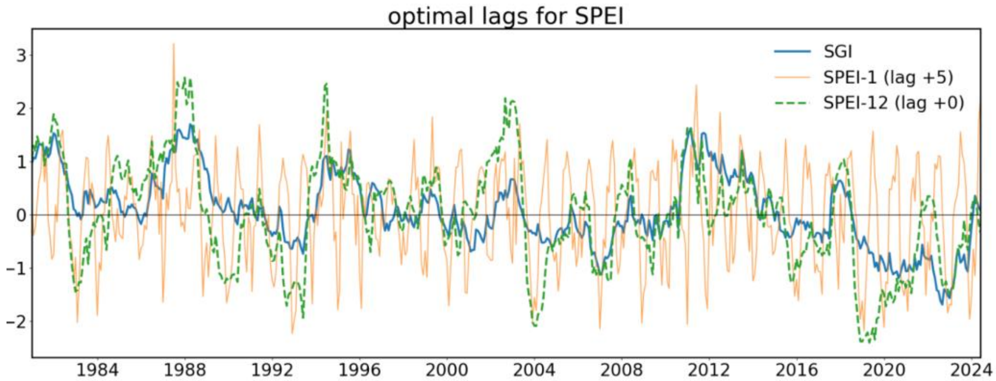
Changes made in the manuscript:

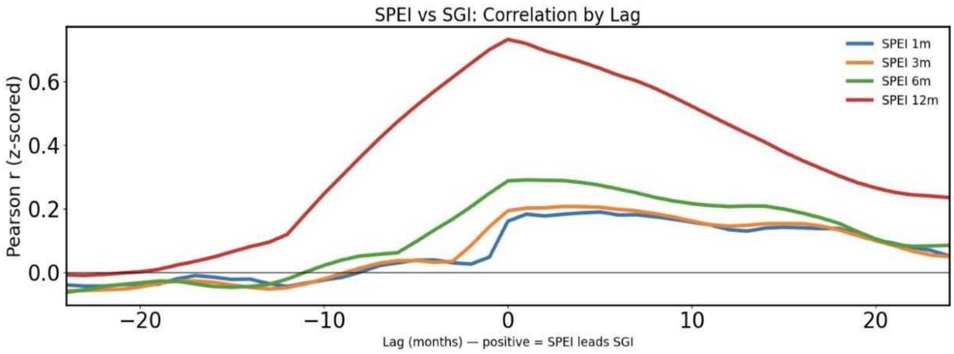
Discussion (Lines 487-488)

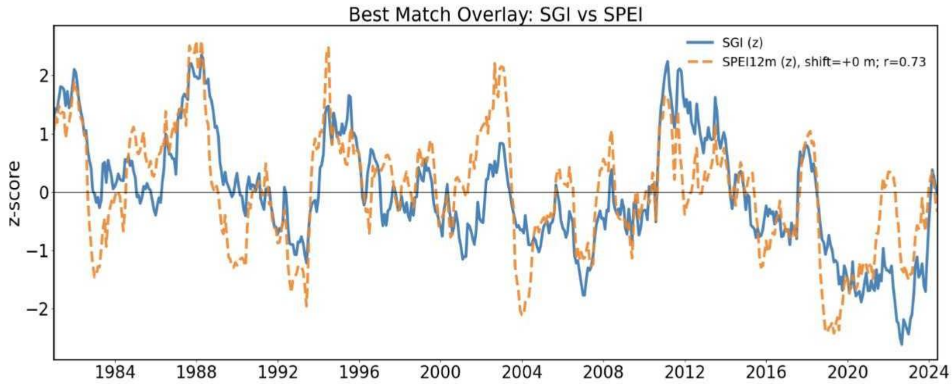
The maximum SPEI-12–SGI correlation at zero lag does not indicate instantaneous recharge. Because SPEI-12 already represents the accumulated climatic water balance of the previous 12 months, the zero-lag correlation shows that groundwater anomalies co-vary with annual-scale climatic conditions. The weak SGI correlations with SPEI-1, which increased with longer accumulation periods and peaked for SPEI-12 (Table S2), further indicate that the confined multi-layer aquifer responds mainly to accumulated climatic deficits rather than individual monthly anomalies. Moreover, SGI in this confined aquifer primarily reflects piezometric head changes, which can respond rapidly to pressure propagation and changing boundary conditions without immediate movement of newly recharged water.

Table S2. Lag-correlation analysis for SPEI-1, SPEI-3, SPEI-6, and SPEI-12 vs SGI

Index	Best lag	Pearson r	r^2	n	Interpretation
SPEI-1 vs SGI	5 months	0.191	0.036	533	Very weak
SPEI-3 vs SGI	3 months	0.208	0.043	531	Very weak
SPEI-6 vs SGI	1 month	0.291	0.085	528	Weak
SPEI-12 vs SGI	0 month	0.734	0.539	522	Strongest







Conclusion (Lines 533-535)

First, hydroclimatic and hydrogeologic indicators (SPEI-12 and SGI) show that groundwater-level anomalies in the confined multi-layer system are strongly coupled to annual-scale climatic water-balance anomalies, although the actual response time cannot be resolved from the SPEI-12 zero-lag relationship alone.

2) Section 3.2, Line 200-206: Application of SGI to this context should be explained properly and with more details. Some parts of this paragraph are unclear, for example you state that "a common preprocessing step is to remove non-stationarity" but it's not clear if you proceeded this way or not.

Authors' response:

We thank the reviewer for pointing out that the original description of the SGI calculation was not sufficiently clear. We have revised Section 3.2 to explain the SGI method in more detail.

In the revised manuscript, we now specify that SGI was calculated following the non-parametric monthly method of Bloomfield and Marchant (2013). Groundwater levels were ranked separately for each calendar month and transformed to standard-normal scores using the inverse normal distribution. This month-wise standardization removes the recurring seasonal cycle from the groundwater-level records.

We also clarified the issue of non-stationarity. No additional detrending was applied. Long-term groundwater-level trends were intentionally retained because they form part of the groundwater drought and depletion signal relevant to this study. The previous sentence referring generally to "removing non-stationarity" has therefore

Changes made in the manuscript:

Section 3.2 (Lines 200-206)

Groundwater drought was quantified using the non-parametric monthly Standardized Groundwater Index (SGI) of Bloomfield and Marchant (2013). This approach is suitable for groundwater-level data because it does not assume a specific probability distribution and accounts for seasonal variability. For each well, monthly groundwater-head records were arranged as a continuous time series, with missing values excluded only from ranking. SGI was calculated separately for each calendar month, so January values were compared only with other January values, February values with February values, and so on. Within each month, groundwater levels were ranked, converted to plotting positions, and transformed to standard-normal scores:

For a given calendar month, the SGI was calculated as:

$$SGI = \Phi^{-1}(p_i)$$

$$p_i = \frac{i - 0.5}{n}$$

where Φ^{-1} is the inverse standard-normal cumulative distribution function and p_i is the plotting position *i* is the rank, and *n* is the number of observations for that month. Negative SGI values indicate below-normal groundwater levels, whereas positive values indicate above-normal conditions. Month-wise standardization removes the seasonal cycle, while no additional detrending was applied because long-term groundwater decline was considered part of the drought and depletion signal. SGI was first calculated for individual wells and then spatially aggregated for the entire catchment and its 21 sub-catchments

3) Globally, a flowchart image of the methodology would enrich methodological section.

Authors' response:

We thank the reviewer for this valuable suggestion. We agree that a flowchart significantly improves the clarity of our complex methodology. We have added a comprehensive methodological flowchart (Fig. 4) that illustrates the sequential structure of our framework, starting from data processing, through the four main analytical components (drought assessment, water balance, GIS-MCDA, and flow-threshold analysis), culminating in the final ASTR suitability and economic outputs. A reference to this figure has been added to Section 3.

Changes made in the manuscript:

Section 3.2 (Lines 165-170)

The overall structure and sequential progression of the methodological framework are illustrated in the flowchart (Fig. 4), which integrates hydroclimatic drought assessment, GIS-MCDA site suitability, surface water availability, and economic screening.

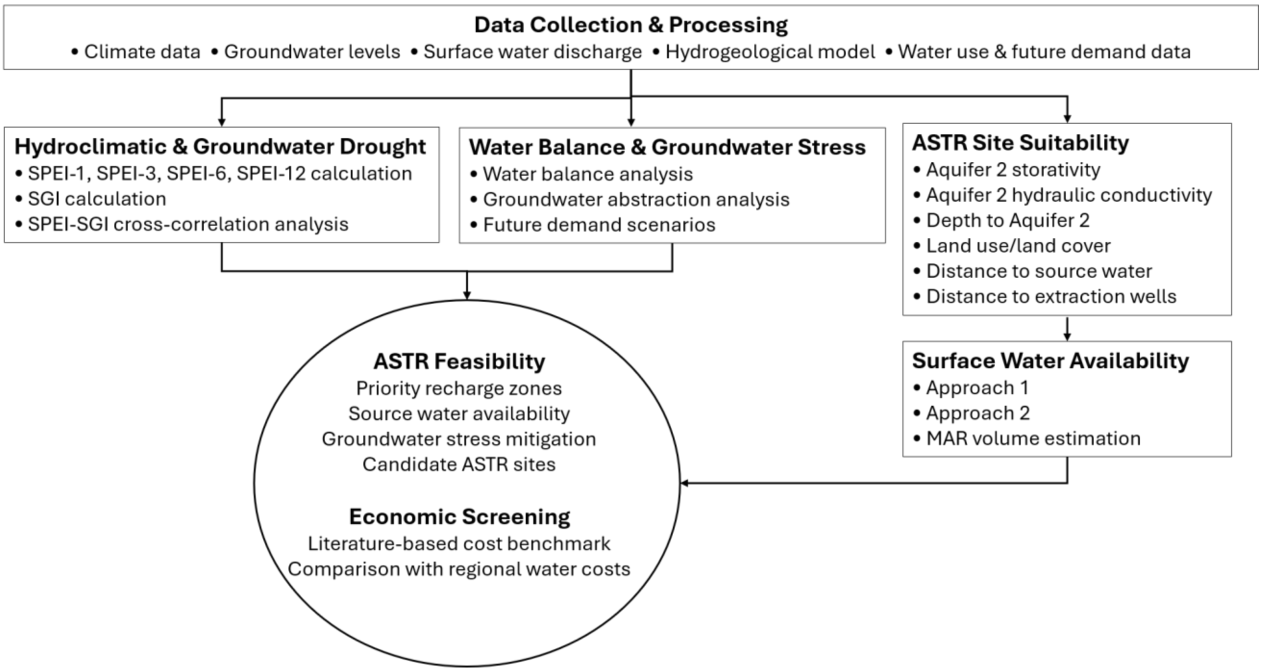


Figure 4. Methodological flowchart illustrating the sequential structure of the integrated ASTR feasibility assessment framework, from data processing and analytical components to final suitability and economic outputs.

4) In section 3.4, regarding the MCDA procedure for selecting site suitabilities for ASTR and MAR, you selected six site-selection criteria and classify their suitability scores. MCDA applications in groundwater reletad Issues, such as groundwater potential recharge evaluation and mapping, are very common in scientific literature of recent Years; often exploring and selecting different ranges of criteria due to case specific needs or constrains. In your case, criteria selection seems to be arbitrary or author-knowledge based, as well as the attributed scores in table 1 and their range values, which can be acceptable only if properly justifyied for the characteristics of your case study and strongly linked to scientific literature and to the purpose of your sudy. In any case, you choises have to be properly framed in the context of contemporary scientific literature on the topic. From both aspects, this methodological section of the manuscript is weak and needs improvements.

Authors' response:

We thank the reviewer for this valuable feedback. We have substantially revised Section 3.4 to explicitly ground all six criteria and their scoring thresholds in recent peer-reviewed MAR literature (e.g., Russo et al., 2015; Sallwey et al., 2019; Dillon et al., 2019; Seidl et al., 2024; Sitek et al., 2026), adding 13 new references.

To address concerns about arbitrary weighting, we performed a sensitivity analysis using three distinct weighting scenarios (equal weight, hydro-heavy, and distance-heavy). Results demonstrate high robustness: the "Very Low" suitability class remained stable at ~29% across all scenarios (variation <0.3%), and high-suitability zones stayed in the same geographic locations. This confirms that identified priority zones are controlled by actual physical constraints rather than subjective weight assignments (Hani et al., 2023; Vandala and Mahed, 2025). We have updated Sections 3.4 and 4.3, added Tables 2–3, and revised Figure 6 to present these findings.

Changes made in the manuscript:

Section 3.4 (revised text)

We applied a GIS-based multi-criteria decision analysis (MCDA) to map the suitability of the study catchment for ASTR implementation, following established MAR/ASTR frameworks (Khalil et al., 2022; Sallwey et al., 2019; Russo et al., 2015). We defined six key site-selection criteria (Table 1) based on their demonstrated influence on recharge efficiency in comparable hydrogeological settings (Gibson et al., 2018; Dillon et al., 2019; Seidl et al., 2024). These criteria included Aquifer 2 storativity (specific storage × thickness), hydraulic conductivity (K) derived from the 3D hydrogeological model, depth to Aquifer 2, land use, and the locations of rivers and extraction wells. We assigned suitability scores ranging from 1 (unsuitable) to 5 (highly suitable) using thresholds widely applied in MAR studies (Russo et al., 2015; Gibson et al., 2018; Abdo et al., 2024). We also applied exclusion masks to account for absolute constraints, including surface water bodies and Well Protection Zones 1-2.

3.4 ASTR/MAR Site Suitability (MCDA)

Aquifer 2 storativity: We calculated Aquifer 2 storativity as the product of specific storage (Ss) and layer thickness (b) ($S = Ss \times b$) to represent actual storage capacity, following methods established in MAR site assessments (Russo et al., 2015; Gibson et al., 2018). We derived specific storage values from from lithology-based estimates within the 3D hydrogeological model and multiplied by the actual Aquifer 2 thickness for each cell. Suitability scores ranged from 5 ($>5\times10^{-3}$) to 1 ($<1\times10^{-4}$), consistent with storage capacity thresholds identified in successful MAR applications (Seidl et al., 2024; Vandala and Mahed, 2025).

Hydraulic conductivity (K): We prioritized areas with higher K values, as they allow water to infiltrate and spread more effectively, a critical factor for ASTR efficiency (Dillon et al., 2019; Meng et al., 2024). Suitability scores ranged from 5 (>30 m/d) to 1 (<5 m/d), aligning with thresholds commonly applied in MAR feasibility studies (Russo et al., 2015; Sitek et al., 2026).

Depth to Aquifer 2: We assessed the depth from land surface to the aquifer top, as greater depths increase injection costs and reduce economic feasibility (Gibson et al., 2018; Abdo et al., 2024). Scores ranged from 5 (highly suitable, <20 m) to 1 (unsuitable, >50 m), following recommendations from MAR implementation guidelines (Stefan and Ansems, 2018; Seidl et al., 2024).

Land Use/Land Cover: We excluded densely urban areas, water bodies, wetlands, and protected natural reserves, favoring forests, open land, and agricultural areas, as they are more likely to accommodate recharge facilities with minimal siting conflicts (Sallwey et al., 2019; Vandala and Mahed, 2025). Forest scored 5, shrubs/herbaceous 4, cropland 2, consistent with land-use suitability classifications in recent MAR studies (Sitek et al., 2026; Meng et al., 2024).

Distance to source water: We considered the distance to the river and its tributaries to minimize pipelines, canals, and infrastructure costs, while excluding flood-prone areas (Gibson et al., 2018; Russo et al., 2015). We estimated groundwater travel time from recharge areas to source water using particle tracking simulations within the hydrogeological model to prevent rapid return flow that could reduce recovery efficiency or affect the river ecosystem (Pollock, 1994; Harbaugh, 2005; Wang et al., 2020). Sites within 500-800 m scored 5, while those >1500 m scored 2, following distance-decay principles established in MAR site selection literature (Dillon et al., 2019; Seidl et al., 2024).

Distance to extraction wells: We assessed the proximity of groundwater pumping wells, excluding highly sensitive Protection Zones 1 and 2 to avoid well interference while ensuring effective recovery (Gibson et al., 2018). We relied on official LfU data on groundwater travel times to pumping wells, which are based on hydrogeological assessments and the delineation of water protection areas. For the remaining areas, we scored suitability according to travel time: <10 years = 5 (highly suitable), 10-30 years = 4, >30 years within the catchment = 3, and outside the catchment = 1 (unsuitable), consistent with well protection criteria in ASR applications (Russo et al., 2015; Stefan and Ansems, 2018).

Weighting and Sensitivity Analysis: To address potential subjectivity in criterion weighting (Hani et al., 2023; Abdo et al., 2024), we performed a sensitivity analysis comparing three scenarios: (S1) equal weighting (all criteria = 0.167), (S2) hydro-heavy weighting (storativity and K = 0.25 each), and (S3) distance-heavy weighting (distance to source and wells = 0.275 each) (Table 2). This follows best practices for validating MCDA robustness in groundwater studies (Hani et al., 2023; Vandala and Mahed, 2025).

Table 2: MCDA weighting scenarios for sensitivity analysis

Scenarios	Storativity	Hydraulic conductivity	Depth	Land Use	Distance to source	Distance to wells
S1 (equal weight)	0.167	0.167	0.167	0.167	0.167	0.167
S2 (hydro-heavy)	0.25	0.25	0.20	0.10	0.10	0.10
S3 (distance-heavy)	0.10	0.10	0.10	0.15	0.275	0.275

Section 4.3 (revised text)

We developed an ASTR suitability map using GIS-MCDA (Fig. 6). The analysis integrated six criteria (Fig. 4a), classified into five classes: very low (0-20), low (21-40), moderate (41-60), suitable (61-80), and highly suitable (81-100).

Sensitivity Analysis Results: We tested three weighting scenarios (Table 2). Results demonstrate high robustness (Fig. 6; Table 3). The "Very Low" suitability class remained stable at ~29% across all scenarios (variation <0.3%), indicating physical constraints consistently define unsuitable areas. The "Suitable" class showed moderate variation (36.7-43.3%; Δ = 6.6%), while "Highly Suitable" areas varied between 17.2% (S3) and 27.7% (S2) (Δ = 10.5%).

Despite percentage variations, the spatial distribution of high-suitability zones remained geographically consistent across all scenarios (Fig. 6a-c), confirming that site suitability is primarily controlled by underlying physical characteristics rather than subjective weight assignment (Hani et al., 2023; Vandala and Mahed, 2025).

Baseline Scenario (S1) Results: Under the equal-weight baseline scenario (S1), 19.5% of the area (670 km²) was highly suitable, and 43.3% (1,484 km²) suitable. Combined, 62.8% of the catchment (2,154 km²) is viable for ASTR. Highly suitable zones concentrate where Aquifer 2 has higher storativity ($>5\times10^{-3}$) and K (>30 m/d), optimal distances to rivers (500–800 m) and wells (<10 years travel time), and minimal land-use conflicts.

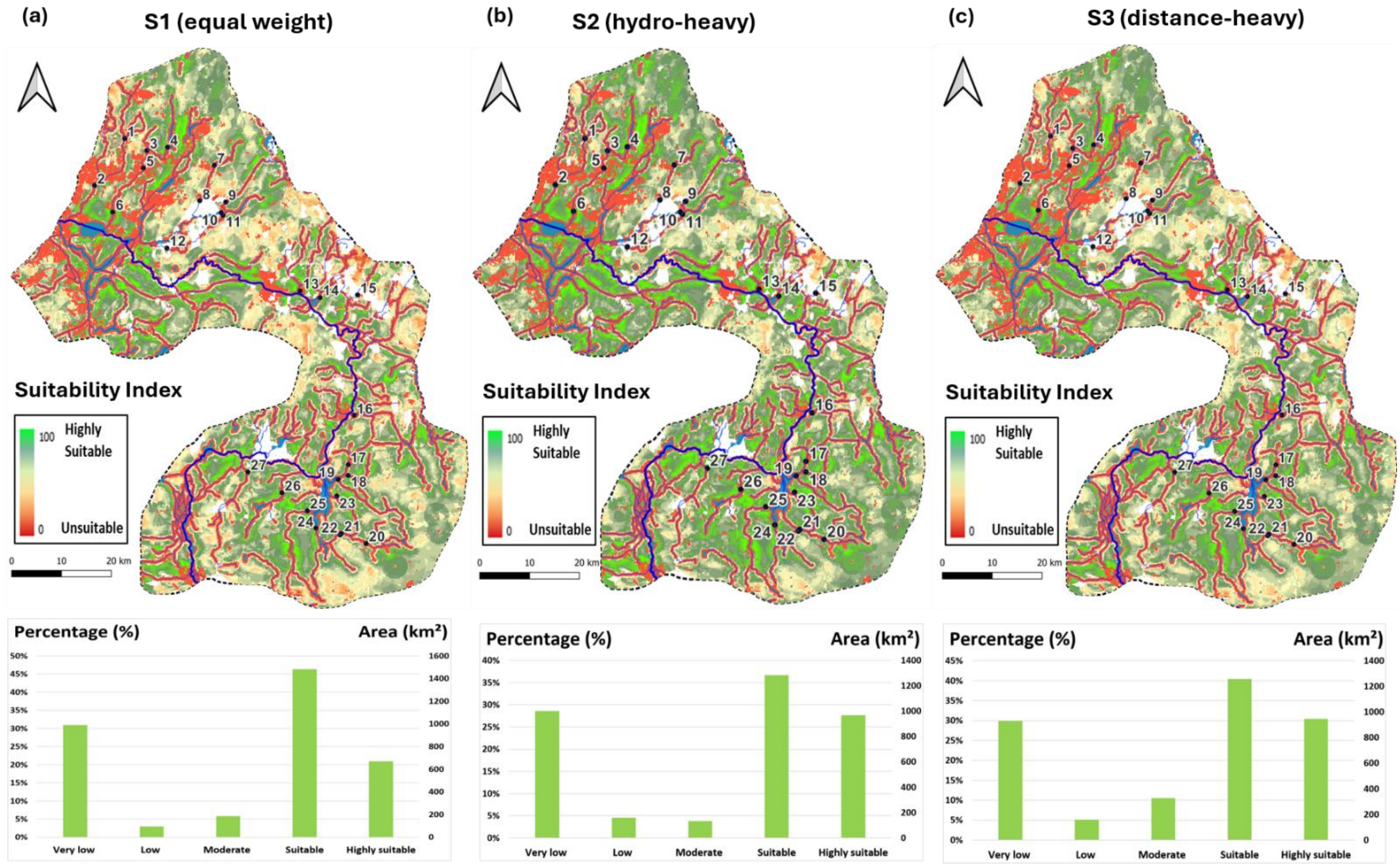


Figure 6. Sensitivity analysis of ASTR site suitability under three MCDA weighting scenarios: (a) S1 (equal weight), (b) S2 (hydro-heavy), and (c) S3 (distance-heavy). Bar charts show the percentage (%) and area (km²) per suitability class. The consistent spatial distribution of high-suitability zones across scenarios demonstrates the robustness of the MCDA results to criterion weighting. See Tables 2–3 for detailed weights and distributions.

Table 3: Suitability class distribution across weighting scenarios

Suitability class	S1 (%)	S2 (%)	S3 (%)	S1- S2	S1- S3
Very low	28.92	28.63	28.84	0.29	0.08
Low	2.82	3.21	2.92	-0.39	-0.1
Moderate	5.4	3.78	10.56	1.62	-5.16
Suitable	43.31	36.72	40.46	6.59	2.85
Highly suitable	19.56	27.66	17.22	-8.1	2.34

Discussion (Lines 507)

…., suggesting the framework itself is highly transferable, moving the study beyond a single case towards a generalizable methodology for regional water security planning. Furthermore, a sensitivity analysis of the MCDA weighting schemes confirmed that these identified priority zones are driven by robust physical constraints (e.g., aquifer properties and depth) rather than subjective weight assignments, reinforcing the reliability and transferability of the framework for regional planning.

References

Abdo, H. G., Vishwakarma, D. K., Alsafadi, K., Bindajam, A. A., Mallick, J., Mallick, S. K., Arun Kumar, K. C., Albanai, J. A., Kuriqi, A., & Hysa, A. 2024. GIS-based multi-criteria decision making for delineation of potential groundwater recharge zones for sustainable resource management in the Eastern Mediterranean: A case study. Applied Water Science, 14, Article 160. https://doi.org/10.1007/s13201-024-02217-z

Dillon, P., Stuyfzand, P., Grischek, T., Lluria, M., Pyne, R. D. G., Jain, R. C., Bear, J., Schwarz, J., Wang, W., Fernandez, E., Stefan, C., Pettenati, M., van der Gun, J., Sprenger, C., Massmann, G., Scanlon, B. R., Xanke, J., Jokela, P., Zheng, Y., ... Sapiano, M. 2019. Sixty years of global progress in managed aquifer recharge. Hydrogeology Journal, 27(1), 1–30. https://doi.org/10.1007/s10040-018-1841-z

Gibson, M.T., Campana, M.E., Nazy, D., 2018. Estimating Aquifer Storage and Recovery (ASR) Regional and Local Suitability: A Case Study in Washington State, USA. Hydrology 5(1), 7. https://doi.org/10.3390/hydrology5010007

Hani, H.M., Nour El Din, M.M., Khalifa, A., Elalfy, E., 2023. Sensitivity Analysis for Multi-Criteria Decision Analysis Framework for Site Selection of Aquifer Recharge with Reclaimed Water. Sustainability 15(6), 5399. https://doi.org/10.3390/su15065399

Lange, S., 2019. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). Geoscientific Model Development 12(7), 3055–3070. https://doi.org/10.5194/gmd-12-3055-2019

Meng, F., Khan, M. I., Naqvi, S. A. A., Sarwar, A., Islam, F., Ali, M., Tariq, A., Ullah, S., Soufan, W., & Faraj, T. K. 2024. Identification and mapping of groundwater recharge zones using multi influencing factor and analytical hierarchy process. Scientific Reports, 14, Article 19240. https://doi.org/10.1038/s41598-024-70324-7

Russo, T.A., Fisher, A.T., Lockwood, B.S., 2015. Assessment of Managed Aquifer Recharge Site Suitability Using a GIS and Modeling. Ground Water 53(2), 269–287. https://doi.org/10.1111/gwat.12213

Sallwey, J., Bonilla Valverde, J.P., Vásquez López, F., Junghanns, R., Stefan, C., 2019. Suitability maps for managed aquifer recharge: a review of multi-criteria decision analysis studies. Environmental Reviews 27(2), 138–150. https://doi.org/10.1139/er-2018-0069

Seidl, C., Sprenger, C., Wang, W., 2024. Understanding the global success criteria for managed aquifer recharge. Journal of Hydrology 628, 130411. https://doi.org/10.1016/j.jhydrol.2023.130469

Sitek, S., Janik, K., Piechota, A., Rubin, H., & Witkowski, A. J. 2026. Application of GIS-MCDA methodology for managed aquifer recharge suitability mapping in Poland. Water, 18(2), Article 219. https://doi.org/10.3390/w18020219

Stefan, C., Ansems, N., 2018. Web-based global inventory of managed aquifer recharge applications. Sustainable Water Resources Management 4, 153–162. https://doi.org/10.1007/s40899-017-0212-6

Vandala, B., Mahed, G., 2025. Managed aquifer recharge (MAR) site suitability in the Nelson Mandela Bay: the application of multi-criteria decision analysis techniques. Sustainable Water Resources Management 11, 78. https://doi.org/10.1007/s40899-025-01241-4

5) Moreover, regarding MCDA procedure, you apply equal weighting to all six MCDA criteria. It is well-known that weighting choices strongly influence MCDA outcomes in MAR suitability studies. The manuscript acknowledges this only in passing ("based on literature and expert judgment"). A sensitivity analysis testing alternative weighting schemes is strongly recommended.

Authors' response:

We fully agree that weighting significantly influences MCDA outcomes. As detailed in our response to Comment #4, We have performed a comprehensive sensitivity analysis testing three alternative weighting schemes: (S1) equal weighting, (S2) hydro-heavy weighting (prioritizing storativity and hydraulic conductivity), and (S3) distance-heavy weighting (prioritizing proximity to source water and wells). The results (new Figure 6, Tables 2–3) confirm that while class proportions shift slightly, the spatial location of high-suitability zones remains consistent across all scenarios. This demonstrates the robustness of our findings and confirms that identified priority zones are controlled by actual physical constraints rather than subjective weight assignments. This sensitivity analysis has been explicitly added to Section 4.3.

Changes made in the manuscript:

Section 4.3 (revised text)

Text is in our response to Comment #4

6) Section 3.6 on Cost Benchmarking from Literature, the cost-analysis is just mentioned as evaluated "from a review of recent european MAR literature". An economic cost analysis would be very interesting and enrich your work, but it hasn't been investigated further since you state it's beyond the scope of your study. At least, you should properly report a summary table of specific informations (if available) for your case study from metioned literature in this paragraph.

It's important to underline that the economic analysis in your study uses a single literature-derived cost range of €0.37–0.51/m³. This range is derived from Ross and Hasnain (2018) and Sprenger et al. (2017), which are based on heterogeneous international datasets. You acknowledge that site-specific analysis was beyond scope, but the manuscript presents potential savings of €23–111 million/year with a level of confidence that is not warranted given the uncertainty in both cost and volume estimates. At minimum, a clear uncertainty paragraph should be introduced, and potential savings claims in the manuscript and in the abstract should be scaled back and framed explicitly only as order-of-magnitude estimates.

Authors' response:

We thank the reviewer for this critical feedback. We agree that presenting specific savings figures with high precision was unwarranted given the heterogeneity of the source data. We have: (1) added Table 4 summarizing four European cost sources with explicit uncertainty notes; (2) scaled back all savings claims to "order-of-magnitude estimates" (e.g., "potential for substantial cost savings" or "tens of millions of euros") in the Abstract, Discussion, and Conclusion; and (3) removed all specific monetary figures (€23–111M) from the manuscript to avoid false precision.

Changes made in the manuscript:

Abstract (Lines 26-28)

...Literature-based cost estimates (€0.30–0.60/m³) are substantially lower than regional drinking-water production costs (~€1.80/m³), indicating potential for substantial cost savings relative to conventional water supply, though precise economic benefits require site-specific validation.

Section 3.6

To provide initial economic context, we derived indicative ASTR cost benchmarks from recent European MAR literature. While site-specific analysis was beyond the scope of this study, reported total costs for well-based MAR typically range from €0.30–0.60/m³ (CAPEX + OPEX) (Ross and Hasnain, 2018; Stefan and Ansems, 2018; Ward et al., 2022; Dillon et al., 2019). For comparison, regional drinking water production costs are ~€1.80/m³ (BWB, 2019). Table 4 summarizes specific benchmarks extracted from reviewed studies, highlighting their inherent limitations.

Table 4: Indicative ASTR/MAR Cost Benchmarks from European Literature

Study	Country	Technology	Total Cost (€/m³)	Key Uncertainty & Limitations	Reference

Ross & Hasnain (2018)	Germany (NRW)	ASTR (well-based)	0.37–0.51	Levelized cost only; ±30% site-specific variation due to geology. No CAPEX/OPEX split.	Ross and Hasnain, 2018
Stefan & Ansems (2018)	Germany	ASTR (well-based)	0.35–0.55	CAPEX amortized over 20–30 yrs; OPEX excludes major well rehabilitation.	Stefan and Ansems, 2018
Ward et al. (2022)	UK/Europe	Well-based MAR	0.30–0.48	Costs scale inversely with capacity; deeper aquifers increase CAPEX by 40–60%. Energy prices based on 2020 tariffs.	Ward et al., 2022
Dillon et al. (2019)	Europe (multi-country)	ASR/ASTR	0.20–0.60	Aggregated from 12 heterogeneous case studies; not directly comparable without normalization.	Dillon et al., 2019
BWB (2019)	Germany (Berlin-Brandenburg)	Drinking Water Production	~1.80	Conventional treatment + distribution; comparative benchmark only.	BWB, 2019

Note: All values are indicative order-of-magnitude estimates. Actual costs depend on local hydrogeology, energy tariffs, and regulatory requirements.

Discussion (Lines 516-520)

Based on the indicative cost benchmarks in Table 4 and the technically viable recharge volumes identified, ASTR implementation at the candidate sites could yield annual cost savings on the order of tens of millions of euros relative to conventional drinking water production (~€1.80/m³). However, given the ±30% uncertainty in literature-derived costs and variability in actual recharge performance, this figure should be treated as an order-of-magnitude estimate for strategic planning purposes only. Site-specific engineering and economic feasibility studies are necessary to refine these projections.

References

Ward, J.D., Simmons, C.T., Barnett, G.B., 2022. Economics of managed aquifer recharge: A European perspective. Journal of Hydrology 612, 128156. https://doi.org/10.1016/j.jhydrol.2022.128156

7) You note that the study covers only 33% of the total surface-water network. While this limitation is acknowledged in the Discussion, the scale of this gap is substantial and is used to argue that actual regional MAR potential is likely substantially higher. This extrapolation is speculative without further analysis since the unmonitored 67% could include tributaries with lower flow regimes, higher ecological sensitivity, or no overlap with suitable ASTR zones. This claim should be removed or framed as future research direction to investigate and verify.

Authors' response:

We agree with the reviewer that extrapolating the regional MAR potential to the unmonitored 67% is speculative. We have removed this claim and reframed it as a future research direction in the Discussion section (Lines 479-481):

Changes made in the manuscript:

Discussion (Lines 479-481)

Finally, we note that the 27 gauges analysed represent only about 33% of the surface-water network. While this monitoring coverage provides a robust basis for the catchment-scale assessment, the remaining 67% of the network was not included in this analysis. Future research should investigate the hydrological characteristics, ecological constraints, and ASTR suitability of these unmonitored tributaries to verify the full extent of the regional MAR potential.

8) Approach 2 relies on hydroecological thresholds that are described generically. The manuscript does not clarify whether these thresholds were derived from local ecological assessments or transferred from literature.

Authors' response:

We thank the reviewer for highlighting this important clarification. We revised Section 3.5 to explicitly distinguish the sources of the thresholds applied in Approach 2. We clarified that we obtained the base hydroecological limit directly from the local guidelines of the State Office for the Environment Brandenburg (LfU), while we derived the additional hydrological thresholds (e.g., median flow and flood peak percentiles) from established scientific literature (e.g., Yarnell et al., 2020; CEFWG, 2021).

Changes made in the manuscript:

Section 3.5 (Lines 286-295)

Approach 2: A more restrictive method that limits abstraction to extreme flood peaks and prioritizes the preservation of environmental flows. We combined several hydrological and hydroecological thresholds to refine the conditions for abstraction. We obtained the base hydroecological limit directly from the local ecological flow guidelines provided by LfU, ensuring its applicability to the regional context. We derived the additional hydrological thresholds (e.g., median daily flow, 2-year and 5-year flood levels) from established scientific literature (e.g., Yarnell et al., 2020; CEFWG, 2021) to preserve channel maintenance flows and associated habitat functions. Under this approach, we do not divert water for ASTR when flows are at or below the hydroecological limit, at or below the median daily flow, or within the range defined by the 2-year and 5-year flood levels. Abstraction is only permitted when flows exceed the median daily flow but remain below the 2-year flood threshold, provided that post-abstraction discharge remains at or above the median daily flow, or when flows exceed the 5-year flood level up to the physical capacity of the intake, pretreatment, wells, and pond storage infrastructure (Schmidt et al., 2004; Vanham et al., 2022; Alley et al., 2022; Richter et al., 2011).

9)The abstract states results clearly but uses the phrase "offsetting 6–79% of local extractions" without specifying which sites this applies to. Readers may misinterpret this as a catchment-wide statement. Clarify that this refers to specific downstream sites.

Authors' response:

We agree with the reviewer that this phrase could be misinterpreted as a catchment-wide statement. We have revised the Abstract to explicitly clarify that this percentage applies only to specific high-potential downstream sites, preventing any overgeneralization.

Changes made in the manuscript:

Abstract (Lines 24-25)

Flow-threshold analysis at 27 gauges showed that specific high-potential downstream sites could provide mean annual recharge volumes of 1.6–4.3 Mm^3, offsetting 6–79% of local extractions at those specific locations.

10) The study focuses on ASTR (Aquifer Storage, Transfer and Recovery), but the framework description and some results sections frequently use ASR and ASTR interchangeably. Since ASTR involves an additional spatial separation between injection and recovery wells compared to ASR, this distinction carries operational significance and should be used consistently throughout.

Authors' response:

We thank the reviewer for this important clarification. We agree that ASTR differs from ASR due to the spatial separation between injection and recovery wells, which has operational significance. We have standardized the terminology throughout the manuscript by replacing ASR/ASTR and standalone ASR terms with ASTR where appropriate. We also added a clarification in the Introduction (Lines 60–65) to distinguish ASTR from ASR.

Changes made in the manuscript:

Introduction (Lines 61-65)

The adoption of Aquifer Storage and Recovery (ASR) and Aquifer Storage, Transfer and Recovery (ASTR) in these settings is limited by data scarcity and operational uncertainty. While ASR typically involves co-located injection and recovery wells, this study focuses specifically on ASTR, which utilizes spatially separated injection and recovery sites. This spatial separation carries distinct operational, hydrogeological, and regulatory significance, which our framework explicitly addresses.

Throughout the manuscript

We replaced any standalone use of "ASR" or "ASR/ASTR" with "ASTR" to ensure consistent terminology, except when referring to general concepts or citing studies that specifically address ASR.