**Response to the Editor and Referees**
*Integrating propagation and recovery dynamics into groundwater drought vulnerability assessment through exposure, pressure, and aquifer system response*
**Katarzyna Sawicka and Klaudia Jurzyk**
Manuscript: egusphere-2026-1904

## Response to the Editor

We thank the Editor for carefully assessing the discussion-stage responses and for the opportunity to revise the manuscript. We have implemented all changes described in our public responses and conducted an additional consistency audit of the manuscript, tables, figures, data package, and analytical workflow. In particular, we have (1) separated empirical statistical associations from causal or mechanistic interpretations, (2) strengthened the treatment of monitoring-network limitations, including the two-well representation of the deep confined aquifer, and (3) added leave-one-out cross-validation and an explicit discussion of spatial interpolation uncertainty. The revised manuscript now presents DIPI consistently as a hydrogeologically informed exploratory screening index rather than as a calibrated, predictive, or causally resolved model.
Because the revision involved extensive restructuring and substantial changes throughout the manuscript, we refer to the relevant sections, tables, figures, and appendices rather than to specific line numbers. These locations are provided after each response for the convenience of the individual referees and are also summarized in a consolidated table for the Editor.

**Summary of major revisions**

- Recalculated SGI-3, SGI-6, and SGI-12 using trailing complete-window aggregation followed by a calendar-month non-parametric normal-score transformation; all dependent drought metrics, tables, figures, and spatial layers were regenerated.
- Corrected and fully documented propagation probability, CRT, CRS, vulnerability, recovery, and Exposure calculations, including a worked propagation example and reproducible event-matching outputs.
- Separated aquifer-group temporal analyses from piezometer-level spatial Exposure inputs and released the complete well-level interpolation input table.
- Added moving-window lag-stability analysis and clarified that SPI–SGI maxima represent empirical temporal association rather than fixed propagation time or causal mechanism.
- Added leave-one-out cross-validation of the three Exposure subcomponents and interpreted the maps and area summaries as exploratory screening outputs subject to substantial spatial uncertainty.
- Clarified that SGI may contain combined climatic and anthropogenic signals, that Exposure and Pressure are not independent, and that E−P is a descriptive mapped contrast only.
- Replaced the incomplete repository release with a complete reproducibility package and provided direct GitHub and archived Zenodo access.

## Response to Referee #1

**General assessment by Referee #1 :**
The manuscript investigates groundwater drought by combining meteorological drought information, represented by SPI, with groundwater response, represented by SGI. It quantifies drought propagation, lag, recovery, and vulnerability across multiple timescales and compares these processes among shallow, intermediate, and deep aquifer systems. The study further attempts to link these temporal drought-response metrics with spatial vulnerability assessment through the composite DIPI framework.
While the topic is relevant and the SPI-SGI analysis has potential value for understanding groundwater drought propagation. The main limitation is that the hydrometeorological process mechanisms are not developed in enough depth. Instead, the paper places substantial emphasis on hydrogeological interpretation, vulnerability mapping, and the composite DIPI index.
**Response:** We sincerely thank Referee 1 for the careful and constructive assessment. The revised manuscript now foregrounds the SPI–SGI propagation and recovery analysis as the primary temporal component, while

DIPI is presented as an index-based spatial extension that integrates observed groundwater-drought behavior with Pressure and Sensitivity proxies. We have removed or qualified claims that implied causal separation or mechanistic simulation.
**Changes in the revised manuscript:** Abstract; Introduction; Sections 4.4–4.5, 5.2–5.6, 6.1–6.5, and Conclusions.

**Referee #1 comment 1:** The main strength of the current manuscript lies in the SPI-SGI relationship and the analysis of multi-timescale drought propagation and recovery. However, the latter part of the paper, including DIPI, the Exposure–Pressure–Sensitivity framework, GDEs, groundwater quality, and mining pressure, is more closely related to groundwater vulnerability or water resources management. The manuscript needs to address hydrometeorological questions more explicitly. For example: How do precipitation anomalies propagate into groundwater drought through recharge, storage memory, lag, and attenuation? How much explanatory power does SPI have for SGI at different timescales? How can meteorological signals be distinguished from anthropogenic abstraction signals?
**Response:** We agree that the temporal SPI–SGI analysis is the primary hydrometeorological contribution. We rebalanced the manuscript accordingly. The revised Introduction and Discussion distinguish between empirical temporal response analysis and the exploratory DIPI screening framework. DIPI is no longer presented as a process model or as a tool that causally separates natural and anthropogenic controls.
**Changes in the revised manuscript:** Introduction; Sections 4.4–4.5, 6.2–6.4, and 7.

**Referee #1 comment 2:** The manuscript uses monthly precipitation data from five meteorological stations for 1986–2024 to calculate SPI and groundwater-level data to calculate SGI. However, the current SPI–SGI analysis remains relatively limited in terms of process interpretation and relies mainly on correlation analysis and qualitative explanation. I recommend that the authors expand this analysis.
**Response:** We expanded the interpretation of SPI–SGI relationships to include attenuation, lag, storage memory, recharge connectivity, and recovery asymmetry. We also added a moving 10-year-window stability analysis and repeated the full-period analysis using detrended series. The maximum-correlation lag is now described as an empirical indicator within the tested 0–12-month range, not as a uniquely determined physical response time.
**Changes in the revised manuscript:** Section 4.4; Section 5.2; Sections 6.1–6.2; Appendix B, Fig. B1; Appendix C, Fig. C1.

**Referee #1 comment 3**: Hydrometeorological forcing should not be limited to precipitation alone. Groundwater drought development and recovery are often also influenced by evapotranspiration, temperature, potential evapotranspiration, actual evapotranspiration, soil moisture, recharge timing, and, where relevant, snow or frozen-ground processes. The current manuscript relies only on SPI and does not include SPEI, PET, actual ET, soil moisture, recharge estimates, or water-balance indicators. This weakens the interpretation of the results. The authors should include at least one supplementary analysis. For example, SPEI or precipitation minus PET could be used instead of SPI to test the robustness of the main conclusions. Alternatively, the authors should clearly justify why precipitation is the dominant meteorological forcing in the study area and why temperature or evapotranspiration changes are not key controls.
**Response:** We retained SPI because the study specifically examines the transmission of precipitation-deficit anomalies into groundwater-level anomalies. The revised Methods now state explicitly that SPI does not represent PET, actual evapotranspiration, soil moisture, snow processes, or effective recharge. These omissions are also acknowledged as limitations, and SPEI, soil-moisture observations, and independent recharge estimates are identified as priorities for future work.
**Changes in the revised manuscript:** Section 4.1; Section 6.5.

**Referee #1 comment 4:** The study uses only 16 groundwater monitoring wells, including 10 shallow wells, 4 intermediate confined wells, and only 2 deep confined wells. The authors also acknowledge that the aquifer grouping is expert-based rather than statistically derived, and that the limited number of confined-aquifer wells means that the results should be interpreted as characteristic of typical settings rather than as population-wide statistics for the entire aquifer system. Therefore, the conclusion that deeper aquifer systems have higher propagation probability, longer recovery time, and greater vulnerability may be hydrogeologically plausible, but it is not yet sufficiently supported statistically.
**Response:** We agree that two deep confined piezometers do not support broad population-wide generalisation. Statements concerning depth and confinement were softened throughout. The revised manuscript reports the deep-system behaviour as characteristic of the monitored locations and explicitly notes confounding by geographic location, record length, hydrogeological setting, and anthropogenic pressure.
**Changes in the revised manuscript:** Abstract; Section 3.2; Sections 5.2–5.5; Sections 6.1–6.5; Conclusions; captions to Figs. 4–7.

**Referee #1 comment 5**: DIPI is one of the main innovations of the manuscript, but it is more closely related to risk mapping and vulnerability assessment. The authors combine Exposure, Pressure, and Sensitivity into DIPI and argue that the similarity between the weighted and equal-weighted versions demonstrates the robustness of the spatial pattern. First, the DIPI weights are subjectively assigned, and the manuscript lacks a systematic sensitivity analysis. Comparing only one weighted version with an equal-weighted version is not sufficient to demonstrate robustness. Second, the Exposure component of DIPI is derived from SGI, but SGI itself may already be affected by groundwater abstraction and mining-related dewatering. At the same time, Pressure separately includes abstraction and mining effects. Therefore, E and P are not necessarily independent. The E−P spatial contrast should not be interpreted directly as a causal separation between "intrinsic system response" and "human pressure." Third, the spatial interpolation of DIPI is strongly constrained by the limited number of monitoring wells. Referees are likely to expect spatial uncertainty analysis, interpolation cross-validation, justification of the interpolation method, and error maps. Without these additions, statements such as "the hotspot configuration is robust" are too strong.
**Response:** DIPI is now described as an exploratory screening index with expert-based weights that have not been calibrated against independent impact observations. The weighted/equal-weight comparison is presented only as a limited sensitivity check. Exposure and Pressure are explicitly described as neither statistically nor causally independent, and E−P is interpreted only as a descriptive contrast of mapped component magnitudes. We also performed leave-one-out cross-validation and expanded the discussion of uncertainty.
**Changes in the revised manuscript:** Sections 4.5, 5.5–5.6, 6.3–6.5, and 7; Figs. 7–8; Appendix A.

**Minor comment 1**: The name of DIPI is inconsistent. The abstract and methods refer to the Drought Impact Potential Index, whereas around line 135 the manuscript refers to the Drought Impact and Pressure Index. The terminology should be made consistent throughout the manuscript.
**Minor comment 2:** Line 115 states that the pressure component is shown in Fig. 6, but Fig. 6 actually presents the propagation–recovery trajectories. The DIPI spatial map appears to correspond to Fig. 7 or Appendix A. This figure reference should be corrected.
**Response:** The name "Drought Impact Potential Index" is now used consistently, and the incorrect reference to Fig. 6 in the pressure description has been corrected to Fig. 7 and Appendix A, Fig. A1.
**Changes in the revised manuscript:** Section 2.3; Section 4.5; Table 3; figure captions; appendices; Code and data availability.

## Response to Referee #2 / Community Comment

**General assessment by Referee #2 / Community Comment:**
This paper presents an interesting approach to groundwater drought by introducing the Drought Impact Potential Index. The overall writing is clear, the study area is well described, and the general idea of moving from static maps to dynamic time-series metrics is a step in the right direction.
**Response:** We thank the commenter for the careful hydrogeological assessment. The revised manuscript now clearly distinguishes between an indicator-based statistical framework and a numerical process model, and treats SGI signals in this mining-affected groundwater body as potentially reflecting combined climatic and anthropogenic forcing.
**Changes in the revised manuscript:** Abstract; Introduction; Sections 4.2, 6.3–6.5, and Conclusions.

**Referee #2 comment 1:** However, there are serious concerns regarding how the methods and results are interpreted from a hydrogeological perspective. The primary issue lies in designating this framework as a process-based approach. In groundwater science, a process-based approach implies solving physical equations of water flow using specific parameters such as hydraulic conductivity and storativity. This study relies entirely on statistical indices like the Standardized Groundwater Index. While using time lags and propagation metrics represents an advancement over simple static maps, it remains a strictly statistical approach. Labeling it as process-based might mislead readers into believing that actual physical flow dynamics were calculated.
**Response:** We replaced or qualified "process-based" and "mechanistic" terminology throughout. The framework is now described as hydrogeologically informed, statistical, process-informed, or indicator-based, depending on context, and the manuscript explicitly states that it does not solve groundwater-flow equations or simulate recharge, storage, or hydraulic parameters.
**Changes in the revised manuscript:** Abstract; Introduction; Sections 4.5, 6.3, and 7.

**Referee #2 comment 2:** This terminology issue leads directly to the most critical misinterpretation in the paper. The authors note that the deep aquifer system shows no recovery and exhibits very long drought durations, attributing this to a long system memory. However, the study area is heavily influenced by open-pit mines. When a mine continuously pumps water to keep the pit dry, the water table drops and remains depressed. This phenomenon constitutes human-made drawdown rather than a natural drought. The groundwater index will naturally remain negative because the system is artificially drained, not because the climate exhibits a long memory. Concluding that deep aquifers have poor drought recovery based on wells affected by mine dewatering is physically incorrect. The paper conflates natural climate deficits with permanent human-induced hydraulic changes.
**Response:** We clarified that SGI cannot independently distinguish meteorological drought from abstraction, mining-related drawdown, aquifer memory, or induced leakage. Negative SGI values are therefore interpreted as signals of groundwater-level drought under combined climatic and anthropogenic forcing. Mining-affected wells were retained because the study addresses drought behavior in a human-modified groundwater system rather than attempting to reconstruct a purely natural signal.
**Changes in the revised manuscript:** Section 4.2; Sections 5.2–5.4; Sections 6.3–6.5.

**Referee #2 comment 3:** Furthermore, the authors divide the aquifers into three clusters, yet the deep cluster relies on only two monitoring wells. In a variable glacial geological setting characterized by interbedded layers of clay and sand, two points cannot adequately represent the behavior of an entire regional aquifer system covering over 3600 km2. Interpolating data from just two wells to generate smooth continuous maps for the whole region creates a false sense of precision. The conclusions regarding deep aquifer behavior should be strictly limited to these two specific locations and not generalized across the entire region.
**Response:** Conclusions from the two deep wells are limited to monitored conditions. The spatial surfaces are described as exploratory screening products, and leave-one-out cross-validation quantifies the interpolation's inability to reliably reproduce local piezometer-scale variability.
**Changes in the revised manuscript:** Sections 4.5, 5.5, 6.3–6.5, and 7.

**Referee #2 comment 4:** The methodology used to quantify human pressure also requires improvement. The authors employ water well density as a proxy for pressure, which does not reflect hydrogeological reality. A single household well pumping minimal volumes and a large mine dewatering system are treated identically as single points on a map. True hydraulic pressure depends on the actual volume of water being extracted, rather than the mere number of boreholes. Utilizing well density creates an output that appears precise but fails to reflect the actual physical stress exerted on the hydrodynamic system.
**Response:** Well density is now presented as a proxy for the spatial distribution of abstraction pressure rather than a measure of withdrawal volume or hydraulic stress. The mining layer is likewise described as an indicator of potential influence and does not quantify pumping rates, temporal variability, or dewatering intensity.
Changes in **the revised manuscript:** Section 4.5; Table 3; Sections 6.3 and 6.5; Fig. 7 caption.

**Referee #2 comment 5:** There is also a significant missed opportunity regarding vertical flow between layers, and addressing this is crucial for the paper to be scientifically sound. In a multi-layered system, intensive pumping from the deep aquifer for mining purposes induces a downward hydraulic gradient. Consequently, water from the shallow aquifer leaks downward into the deeper zone. The authors might be observing a decline in shallow groundwater levels and misinterpreting it as meteorological drought propagation, whereas it may actually represent water being physically drawn down into the mining drainage system. The authors do not necessarily need to run complex new calculations to fix this, but they must address this mechanism directly in their text to show they understand the physical limits of their data. First, they should add a dedicated paragraph in the Discussion section explaining that intensive pumping reverses natural vertical gradients, creating a downward pull. They must explicitly state that this physical drainage could be recorded by shallow monitoring wells as a drought signal, admitting that what looks like drought propagation might actually be water drawn down into the mining system. Second, the authors need to address their data limitations by explaining that separating natural propagation from mining-induced leakage requires calculating vertical hydraulic gradients using pairs of shallow and deep piezometers at the exact same site. Since their network consists of individual scattered wells rather than purpose-built multi-level nests, they cannot calculate these exact gradients, and being honest about this limitation is much better than ignoring the physics. Third, the authors should adjust the methodology section to acknowledge this upfront by noting that standardized indices aggregate all water level declines into a single signal and cannot independently distinguish between a drop caused by a lack of rainfall and a drop caused by vertical leakage. Finally, the authors must soften their conclusions regarding the shallow aquifer. Right now, the paper strongly links shallow aquifer declines to meteorological propagation. The relevant sentences should be revised to state that the observed declines are likely a combined effect of meteorological drought propagation and potential vertical leakage driven by the artificial drawdown of the deep aquifers. Making this change shows scientific maturity and proves that while their tool is statistical, their understanding of the actual hydrogeological system is grounded in physical reality.
**Response:** The Discussion now explains that mine dewatering may alter vertical gradients and induce downward leakage from shallower units into deeper drained zones. We also state that quantifying these mechanisms would require nested or paired shallow and deep piezometers, which are not available in the present network. The interpretation of shallow groundwater declines is now more tentative. The revised text recognizes the possible combined influence of precipitation deficit, local abstraction, mining-induced hydraulic redistribution, and vertical leakage.
**Changes in the revised manuscript:** Section 6.4; Section 6.5; Section 4.2; Section 6.4.

**Referee #2 comment 6:** The justification for the expert-based clustering and the chosen index weights remains somewhat weak, although testing the unweighted version was a commendable step to verify the robustness of the spatial patterns. In summary, the authors do not need to construct a complex numerical groundwater model to address these shortcomings. They simply need to isolate the wells affected by mining from those reflecting natural conditions in their drought analysis, avoid interpolating two deep wells over an expansive area, and remove the claim that this constitutes a process-based physical approach. If they refine their conclusions to explicitly acknowledge the limitations of statistical indices in explaining physical groundwater processes, the paper will represent a valuable contribution to the journal.
**Response:** We now explicitly state that the aquifer grouping and DIPI weights are expert-based and introduce uncertainty. The equal-weight formulation is treated as a limited sensitivity check, and future calibration and denser monitoring are identified as requirements for application elsewhere.
**Changes in the revised manuscript:** Section 3.2; Section 4.5; Sections 6.3, 6.5, and 7.

## Response to Referee 3

We thank the Referee for the careful and technically detailed assessment of the manuscript. The comments led us to verify and substantially revise the SGI workflow, groundwater-drought metrics, spatial Exposure analysis, lag interpretation, interpolation uncertainty assessment, and the scope of the Drought Impact Potential Index (DIPI). We have recalculated all dependent results and regenerated the affected tables and figures. We have also revised the terminology and interpretation throughout the manuscript to present the framework as a hydrogeologically informed, indicator-based exploratory screening approach rather than a mechanistic or predictive model.
Revisions are detailed point-by-point below.

**Referee #3 comment 1**: Table 2 defines propagation probability as the fraction of SGI-3 events developing into SGI-12 events, yet Table 4 reports a probability for every SGI scale. Some values are impossible under the stated formula. The authors should provide the implemented equation and a worked example, and then correct the affected results.
**Response:** We thank the Referee for identifying this inconsistency. The propagation probabilities reported in the original Table 4 were inconsistent with the equation in Table 2. We therefore recalculated the metric from the revised SGI event series and now define propagation probability exclusively as
$P_{3\rightarrow12} = N_{(SGI3\rightarrow SGI12)} / N_{(SGI3)}$, where $N_{(SGI3)}$ is the number of SGI-3 drought events, and $N_{(SGI3\rightarrow SGI12)}$ is the number of SGI-3 events matched to an overlapping or subsequent SGI-12 event beginning within 12 months of the SGI-3 onset. Each SGI-3 event is counted only once.
Propagation probability is now reported as a single SGI-3-to-SGI-12 metric for each aquifer system, rather than separately for SGI-3, SGI-6, and SGI-12. The recalculated probabilities are 1.00 for the shallow unconfined and intermediate confined aquifers and 0.80 for the deep confined aquifer. Table 4, Sections 4.3, 5.3, and 5.4, and Figures 5 and 6 have been revised accordingly.
For example, the deep confined aquifer contained five SGI-3 drought events, four of which were matched to SGI-12 events; therefore, $P_{3\rightarrow12} = 4/5 = 0.80$.
The event-matching table and worked example are included in the revised reproducibility package to enable direct verification of the numerator and denominator.
**Changes in the revised manuscript:** Section 4.3; Table 2; Sections 5.3–5.4; Table 4; Figs. 5–6; reproducibility package.

**Referee #3 comment 2:** The reported CRT values cannot be reconciled with the equation in Table 2. Values of 0.90–0.98 would imply that negative SGI occurs in only about 2–9% of all months, which conflicts with Fig. 4. The authors should verify the formula, indicator definitions, and reported values.
**Response:** We agree that the original definition and reported values were inconsistent. We recalculated CRT using the explicit definition $CRT = 1 - N_{(drought\ months)} / N_{(valid\ months)}$, where groundwater-drought months are

defined as months with SGI < −1. CRT therefore represents the fraction of valid observations outside groundwater-drought conditions.
The equation, variable definitions, and Table 4 values have been corrected. The recalculated CRT values range from 0.815 to 0.918 and are consistent with the frequency of groundwater-drought months in the revised SGI series. To avoid ambiguity, CRT is now described as a resistance/reliability metric rather than as a direct measure of aquifer recovery.
**Changes in the revised manuscript:** Section 4.3; Table 2; Section 5.3; Table 4.

**Referee #3 comment 3:** Table 3 and Fig. A1 define exposure as E = 0.4VUL′ + 0.3CRS_inv′ + 0.3MaxDur′, whereas Fig. 7a uses E = 0.6VUL′ + 0.4CRS_inv′. The authors should identify the equation actually used and regenerate all dependent maps and statistics.
**Response:** We thank the Referee for identifying this inconsistency. The intended Exposure formulation is E = 0.4VUL_N + 0.3CRS_INV + 0.3MaxDur_N, where VUL_N, CRS_INV, and MaxDur_N are the normalized piezometer-level SGI-12 vulnerability, inverse resilience, and maximum drought-duration metrics, respectively. The alternative expression shown in the original caption of Fig. 7a was inconsistent with the methodology and has been removed.
Because the SGI series and dependent drought metrics were recalculated during the revision, we regenerated the Exposure component using the stated three-metric formulation. All dependent DIPI and E−P maps, summary statistics, area proportions, and relevant figure panels were recalculated. Figures 7, 8, and A1 were regenerated, and the Exposure equation is now reported consistently in Table 3, Section 4.5, the figure captions, and the reproducibility package.
**Changes in the revised manuscript:** Section 4.5; Table 3; Figs. 7–8; Appendix A (Fig. A1); reproducibility package.

**Referee #3 comment 4:** Section 5.3 assigns values of 11 months, 23 months, and −17.07 to Cluster 3 at SGI-3, whereas Table 4 reports different values. The values in the text instead match another group and aggregation scale. This section should be corrected against the analysis output.
**Response:** We thank the Referee for identifying this transcription error. The values in the original Section 5.3 were incorrectly assigned and did not correspond to the analysis output.
Because the SGI procedure and all event metrics were recalculated, we replaced the entire Section 5.3 using the revised results rather than correcting only the isolated values. The revised section now reports the recalculated number of events, mean and maximum durations, severity, CRT, CRS, vulnerability, recovery, and propagation metrics, consistent with Table 4 and Figures 5–6. Table 4 is generated directly from the reproducible output.
**Changes in the revised manuscript:** Section 5.3; Table 4; Figs. 5–6.

**Referee #3 comment 5:** The spatial interpolation inputs are unclear. The Methods state that drought metrics were derived from aquifer-group mean SGI series, but the maps are described as being interpolated from piezometers. If group-level values were assigned to individual wells, the interpolation contains only three distinct drought-metric values. If well-level metrics were used, they should be reported and made available.
**Response:** We agree that the original description did not clearly distinguish between the aquifer-group and piezometer-level analyses. Aquifer-group mean SGI series were used only to compare temporal groundwater-drought behaviour among the three monitored aquifer systems and to derive the group-level results presented in Table 4 and Figures 4–5.
The spatial Exposure analysis was based on piezometer-level SGI-12 drought metrics calculated separately for each of the 16 monitoring wells. Group-level values were not assigned to individual piezometers. For each well, vulnerability (VUL), inverse resilience (CRS_INV), and maximum drought duration (MaxDur) were calculated from its individual SGI-12 series. Each metric was min–max normalized across the monitoring network, interpolated independently, and combined using E = 0.4VUL_N + 0.3CRS_INV + 0.3MaxDur_N.

Section 4.3 now distinguishes explicitly between group-level metrics used for temporal comparison and well-level metrics used for spatial analysis. Section 4.5 now specifies the exact piezometer-level inputs to interpolation. The Exposure subcomponent maps, Exposure surface, DIPI maps, and E−P contrast map were regenerated from the corrected well-level inputs.

The complete well-level input table is provided as Exposure_SGI12_piezometers.csv and includes PIEZ_ID, aquifer-system assignment, VUL, normalized VUL, CRS, inverse normalized CRS, MaxDur, normalized MaxDur, and Exposure.

**Changes in the revised manuscript:** Sections 4.3 and 4.5; Fig. 6; piezometer-level Exposure input table in the reproducibility package.

**Referee #3 comment 6:** The monitoring network contains only 16 wells, including two deep wells. Without cross-validation and spatial uncertainty estimates, this network provides insufficient support for precise regional area percentages. The maps should therefore be presented as exploratory screening products.

**Response:** We agree with the Referee. The limited and spatially uneven monitoring network does not support treating the interpolated surfaces as precise regional predictions. We therefore revised both the analysis and its interpretation.

Leave-one-out cross-validation was performed separately for the three piezometer-level SGI-12 Exposure subcomponents: normalized vulnerability (VUL_N), inverse resilience (CRS_INV), and normalized maximum drought duration (MaxDur_N). In each iteration, one of the 16 piezometers was omitted, the surface was regenerated from the remaining 15 wells using the same interpolation method, spatial extent, cell size (approximately 211.8 m), and boundary constraints as in the final maps, and the predicted value was extracted at the omitted location.

Predictive performance was evaluated using mean error (ME), mean absolute error (MAE), root mean square error (RMSE), and the Pearson correlation between observed and predicted values. Mean errors were close to zero for all three variables, indicating little overall bias. However, MAE ranged from 0.211 to 0.287 and RMSE from 0.315 to 0.344 on the normalized 0–1 scale. Observed–predicted correlations were weak, ranging from $r = -0.195$ to $r = 0.108$. These results indicate that the interpolation does not reliably reproduce local piezometer-scale variability.

The largest leave-one-out errors occurred at spatially isolated locations and where neighbouring piezometers showed strongly contrasting drought-response metrics. In some cases, nearby wells represent different aquifer systems. For example, II/1285/1 and II/797/1 are spatially close but represent the shallow unconfined and deep confined aquifers, respectively. A two-dimensional interpolation treats such observations as part of a single continuous surface and cannot explicitly represent vertical hydrogeological separation. The uneven monitoring density, with fewer wells in the southern part of the study area, further increases uncertainty in sparsely supported areas.

Because Topo to Raster does not provide a model-based prediction standard-error surface, uncertainty was evaluated empirically using leave-one-out cross-validation rather than a continuous kriging-variance map.

The cross-validation procedure has been added to Section 4.5, the results are reported in Section 5.5, and their implications are discussed in Section 6.5. Detailed observed, predicted, and error values are provided in TopoToRaster_LOO_results.csv, with summary statistics in TopoToRaster_LOO_summary.csv.

Sections 6.3–6.5 and the Conclusions have been revised so that the Exposure, DIPI, and E−P maps are consistently described as exploratory screening products rather than predictive or causally resolved spatial models. Statements concerning robust hotspot boundaries and precise regional spatial patterns have been removed or softened. Area proportions are retained only as descriptive summaries of mapped classes and are explicitly qualified as sensitive to monitoring density, aquifer representation, and interpolation uncertainty.

**Changes in the revised manuscript:** Sections 4.5, 5.5, 6.3–6.5, and 7; LOOCV output files in the reproducibility package.

**Referee #3 comment 7:** The manuscript states that all series were truncated to 2005–2024, whereas Table 1 gives 2005–2021 for the shallow group. The authors should report the actual period used for each well and explain how missing observations were handled.

**Response:** We agree that the statement that all groundwater-level series were truncated to a common 2005–2024 period was inaccurate. Monitoring records vary across piezometers in both start and end dates, and several series contain gaps. We removed this statement and now report the actual observation period and the number of valid monthly observations for each piezometer in the revised monitoring-well completeness information in Table 1.

SGI was calculated separately for each piezometer using all available valid monthly observations. Trailing 3-, 6-, and 12-month aggregation windows were retained only when all constituent calendar months were present; windows crossing a missing month were excluded. Missing groundwater-level values were not interpolated.

Aquifer-system mean SGI series were calculated from the available well-level SGI values for each month. A monthly group value was retained only when observations were available from at least 50% of the wells in the group and from at least two wells. Consequently, the shallow and intermediate series may include a varying number of wells over time, whereas the deep-aquifer series required both deep wells.

Table 1, Sections 3.2 and 4.2, and the group-level analysis description have been revised. The reproducibility package includes a completeness table listing the observation period, the number of valid months, and the number of missing months for each piezometer.

**Changes in the revised manuscript:** Table 1; Sections 3.2 and 4.2; well-completeness information in the reproducibility package.

**Referee #3 comment 8:** The SGI procedure is insufficiently documented. The manuscript should clarify whether the normal-score transformation was performed by calendar month, whether temporal aggregation occurred before or after standardization, and how incomplete windows were treated.

**Response:** We agree. During verification of the workflow, we found that the archived SGI series had been generated using a global z-score transformation of aggregated groundwater levels rather than the calendar-month normal-score transformation described in the manuscript. We therefore recalculated all SGI series and all dependent groundwater-drought metrics.

In the revised workflow, monthly groundwater levels are first aggregated using trailing 3-, 6-, and 12- month arithmetic means. An aggregation window is retained only when all constituent monthly observations are available. The aggregated values are then transformed separately for each calendar month using the non-parametric normal-score transformation $SGI = \Phi^{-1}[(r - 0.5)/n]$, where r is the rank of an aggregated groundwater-level value within the corresponding calendar month, n is the number of available values for that month, and $\Phi^{-1}$ is the inverse standard normal cumulative distribution function. Average ranks are used for tied values.

Missing groundwater-level observations are not interpolated. Aggregation windows containing missing observations are excluded, and a gap in the resulting SGI series terminates an ongoing drought event. These details have been added to Section 4.2 and the reproducibility package.

Because the corrected SGI procedure changed the identification and characteristics of some drought events, Table 4, Sections 5.2–5.4, and Figures 4–6 were regenerated.

**Changes in the revised manuscript:** Section 4.2; Sections 5.2–5.4; Table 4; Figs. 4–6; SGI notebooks in the reproducibility package.

**Referee #3 comment 9:** The lag analysis uses overlapping and autocorrelated series and selects the lag with the maximum correlation. A time-series-aware uncertainty analysis or stability test is needed. The current correlations show temporal association but do not establish a propagation mechanism.

**Response:** We agree that selecting the maximum correlation from overlapping and autocorrelated series does not, by itself, uniquely determine the propagation lag. We therefore revised both the analysis and its interpretation.

The lag corresponding to the maximum SPI–SGI correlation is now described as an empirical indicator of temporal association within the tested 0–12-month range, rather than as direct evidence of a propagation mechanism or a fixed aquifer-response time.
We added a temporal stability analysis using 10-year moving windows, each shifted by 1 year. The broad contrast among aquifer systems is retained, but the exact lag varies among subperiods, particularly for SGI-3. The SGI-12 maximum for the shallow unconfined aquifer was comparatively stable at 7–10 months, while the deep confined aquifer most frequently reached the upper tested limit of 12 months.
The latter suggests that the response delay may exceed the tested range and cannot be resolved precisely.
The moving-window analysis does not remove the effects of serial dependence and is used as a stability diagnostic rather than as a formal confidence interval for lag. The revised methodology is described in Section 4.4, the stability results are summarized in Section 5.2, and the full moving-window results are presented in Appendix C, Figure C1. Statements implying a mechanistic or uniquely determined propagation time have been removed.
**Changes in the revised manuscript:** Section 4.4; Section 5.2; Appendix B (Fig. B1); Appendix C (Fig. C1).

**Referee #3 comment 10:** The framework is described as process-based and mechanistic, although it consists of standardized indices, event statistics, and weighted spatial layers. No groundwater-flow or water-balance calculation is included, and hydraulic parameters do not enter the index. "Hydrogeologically informed statistical framework" would be a more accurate description.
**Response:** We agree that the terms "process-based" and "mechanistic" were too strong for the analytical framework. The framework does not solve groundwater-flow equations, calculate a groundwater balance, or explicitly incorporate hydraulic parameters. Instead, it combines standardized drought indices, event- based response metrics, hydrogeological grouping, and weighted spatial indicators.
We revised the terminology throughout the manuscript. "Process-based" and "mechanistic" have been replaced, depending on context, with "hydrogeologically informed statistical framework," "process-informed framework," or "indicator-based framework incorporating propagation and recovery metrics." Statements suggesting that the method directly represents physical groundwater processes have also been revised. The manuscript now clarifies that the framework uses statistical indicators reflecting observed lag, persistence, vulnerability, and recovery, but does not constitute a numerical groundwater-flow or process model.
**Changes in the revised manuscript:** Abstract; Introduction; Section 4.5; Section 6.3; Section 7.

**Referee #3 comment 11:** E−P cannot separate natural and anthropogenic controls because Exposure is derived from groundwater levels that may already contain climate, pumping, and mine-dewatering signals. Subtracting two normalized indices does not provide causal attribution, and the reported percentages should be described only as mapped E−P classes.
**Response:** We agree that the E−P contrast cannot be interpreted as a causal separation between intrinsic aquifer response and anthropogenic pressure. Exposure is derived from observed groundwater-level behavior and may therefore already include effects of climatic variability, groundwater abstraction, and mining-related dewatering. Subtracting the normalized Pressure component does not isolate independent natural and human controls.
The E−P layer is now defined solely as a spatial contrast between the relative mapped magnitudes of normalized Exposure and Pressure. Positive values indicate locations where the mapped Exposure score exceeds the mapped Pressure score, whereas negative values indicate the opposite. These classes do not identify the cause of groundwater drought or quantify the relative contributions of climatic and anthropogenic drivers.
The revised mapped-class proportions are 65.5% for $E > P$, 31.6% for near-balanced conditions ($-0.1 \le E-P \le 0.1$), and 2.9% for $E < P$. These values are reported in Section 5.6 and discussed in Section 6.4 solely as descriptive areal proportions of mapped contrast classes. Statements referring to disentanglement, causal separation, or attribution of natural and anthropogenic controls have been removed.

**Changes in the revised manuscript:** Sections 4.2 and 4.5; Section 5.6; Sections 6.4–6.5; Section 7; Fig. 8.

**Referee #3 comment 12:** DIPI should be presented as an exploratory screening index. Its weights are not calibrated, well density is only a proxy for water use, and the mining layer does not quantify dewatering intensity. Comparing one weighted scheme with one equal-weight scheme does not establish robustness or transferability.

**Response:** We agree that DIPI should be presented as an exploratory screening index rather than a calibrated or predictive measure of groundwater drought impact. The component weights were selected using expert judgement and were not calibrated against an independent impact dataset. Abstraction-well density is only a spatial proxy for groundwater use, while the mining component delineates areas of potential dewatering influence without quantifying pumping rates, temporal variability, or dewatering intensity.

The manuscript now describes DIPI as a hydrogeologically informed exploratory screening tool for comparing relative spatial patterns of drought Exposure, anthropogenic Pressure, and environmental Sensitivity within GWB 43. It is not presented as a validated estimate of absolute impact, risk, or causal contribution.

The comparison between the weighted and equal-weight formulations is now described as a limited sensitivity check on the broad internal spatial pattern rather than as evidence of robustness, calibration, or transferability. The interpretation has been revised in Sections 5.5, 6.3, 6.5, and 7.

Statements that the hotspot configuration is "robust" or that the framework is already "transferable" have been removed or qualified. The Discussion and Conclusions also identify the need for future calibration using independent impact observations, pumping-rate data, and quantitative mine- dewatering information.

**Changes in the revised manuscript:** Section 4.5; Section 5.5; Sections 6.3 and 6.5; Section 7; Fig. 7; Appendix A.

**Referee #3 comment 13**: The Zenodo v1.1 package available through DOI 10.5281/zenodo.19371351 on 16 July 2026 contained placeholder material and example metadata for only PZ01–PZ03. It did not contain the complete processed data or an executable workflow. The authors should provide the actual reproducibility package or a direct link to any complete GitHub repository.

**Response:** We thank the Referee for identifying this problem. We agree that the Zenodo v1.1 archive available during the review was incomplete and contained placeholder material rather than the full processed datasets and executable workflow.

The reproducibility package has been replaced and archived as version v1.3. It includes processed groundwater-level and precipitation data; revised SGI-3, SGI-6, and SGI-12 series; well- and aquifer-group drought events and metrics; propagation matches; recovery outputs; DIPI component tables; figure source data for the main text and Appendices A–C; and executable Jupyter notebooks that reproduce the analytical workflow.

The updated package documents the corrected calendar-month non-parametric normal-score SGI procedure, complete-window treatment, handling of missing observations, event detection, recovery calculations, SGI-3 to SGI-12 propagation matching, moving-window lag-stability analysis, and DIPI calculations with equal and weighted weights. The corresponding GitHub repository and archived Zenodo release are:

GitHub repository: https://github.com/sawicka-byte/GWB43-groundwater-drought-DIPI

Zenodo archived release v1.3: 10.5281/zenodo.21471221

The Data and Code Availability section has been updated accordingly.

**Changes in the revised manuscript:** Code and data availability; GitHub repository; archived Zenodo release v1.3.

**Referee #3 comment 14:** Use "Drought Impact Potential Index" consistently; "Drought Impact and Pressure Index" is used incorrectly in several places.

**Response:** We thank the Referee for identifying this inconsistency. We corrected the terminology throughout the manuscript and now use "Drought Impact Potential Index" consistently in the abstract, Section 4.5, Table 3, the main text, figure captions, appendices, and the Data and Code Availability section. All

occurrences of the incorrect term “Drought Impact and Pressure Index” have been removed.
**Changes in the revised manuscript:** Abstract; Section 4.5; Table 3; main text; figure captions; appendices; Code and data availability.

**Referee #3 comment 15:** Section 2.3 incorrectly refers to Fig. 6 for the pressure map; the relevant maps are in Fig. 7 or Fig. A1.
**Response:** We thank the Referee for identifying this incorrect cross-reference. The reference in Section 2.3 has been corrected. The spatial distribution of anthropogenic-pressure component layers is now referred to Fig. A1, while the composite Pressure surface used in the DIPI framework is referred to the appropriate panel of Fig. 7.
**Changes in the revised manuscript:** Section 2.3; Fig. 7; Appendix A (Fig. A1).

**Referee #3 comment 16:** Lines 144–145 list four aquifer categories, whereas the analysis uses three groups.
**Response:** We thank the Referee for identifying this inconsistency. The manuscript analyses three aquifer systems, not four. The relevant sentence and all associated descriptions have been corrected to refer consistently to the shallow unconfined, intermediate confined, and deep confined aquifers.
**Changes in the revised manuscript:** Section 3.2 and descriptions throughout the manuscript; Fig. 1 and associated captions.

**Referee #3 comment 17:** The statement that deeper aquifers generally have higher propagation probability is not consistently supported and should be reconsidered after the metric has been verified. Depth is also confounded with location and hydrogeological setting.
**Response:** We agree. After recalculation, propagation probability does not increase systematically with aquifer depth. The revised probabilities are 1.00, 1.00, and 0.80 for the shallow unconfined, intermediate confined, and deep confined aquifers, respectively.
We removed statements suggesting a general increase in propagation probability with depth. The revised interpretation emphasizes that the clearest distinction of the monitored deep confined aquifer concerns drought persistence, cumulative deficit, low resilience, and prolonged or absent recovery rather than propagation probability alone.
We also clarify that aquifer depth is confounded with location, hydrogeological setting, monitoring-period length, and the small number of deep wells. The results are therefore interpreted as characteristic of the monitored aquifer systems rather than as a general depth-dependent relationship.
Figure 6 has been redesigned to show piezometer-level variability. Propagation probability was calculated independently for each piezometer, while SGI-12 vulnerability and recovery status were derived from the corresponding well-level series. The figure therefore illustrates within-system heterogeneity and is not intended to represent dispersion around the aquifer-system-level propagation probabilities.
**Changes in the revised manuscript:** Sections 5.3–5.4, 6.1–6.2, 6.5, and 7; Figs. 5–6.

## Consolidated locations of revisions

| Referee | Comment | Location in the revised manuscript |
|---|---|---|
| R1 | Comment 1 | Introduction; Sections 4.4–4.5, 6.2–6.4, and 7. |
| R1 | Comment 2 | Section 4.4; Section 5.2; Sections 6.1–6.2; Appendix B (Fig. B1); Appendix C (Fig. C1). |
| R1 | Comment 3 | Section 4.1; Section 6.5. |
| R1 | Comment 4 | Abstract; Section 3.2; Sections 5.2–5.5; Sections 6.1–6.5; Section 7; captions to Figs. 4–7. |
| R1 | Comment 5 | Sections 4.5, 5.5–5.6, 6.3–6.5, and 7; Figs. 7–8; Appendix A. |
| R1 | Minor 1 | Title, Abstract, Section 4.5, Table 3, figure captions, appendices, and Code and data . |
| R1 | Minor 2 | Section 2.3; Fig. 7; Appendix A (Fig. A1). |
| R2 | Comment 1 | Abstract; Introduction; Section 4.5; Section 6.3; Section 7. |
| R2 | Comment 2 | Section 4.2; Sections 5.2–5.4; Sections 6.3–6.5; Section 7. |
| R2 | Comment 3 | Section 3.2; Sections 4.5, 5.5, 6.3–6.5, and 7; captions to Figs. 4–7. |
| R2 | Comment 4 | Section 4.5; Table 3; Sections 6.3 and 6.5; Fig. 7 caption. |
| R2 | Comment 5 | Section 4.2; Sections 6.4–6.5; Section 7. |
| R2 | Comment 6 | Section 3.2; Section 4.5; Sections 6.3, 6.5, and 7; Appendix A. |
| R3 | Comment 1 | Section 4.3; Table 2; Sections 5.3–5.4; Table 4; Figs. 5–6; reproducibility package. |
| R3 | Comment 2 | Section 4.3; Table 2; Section 5.3; Table 4. |
| R3 | Comment 3 | Section 4.5; Table 3; Figs. 7–8; Appendix A (Fig. A1); reproducibility package. |
| R3 | Comment 4 | Section 5.3; Table 4; Figs. 5–6. |
| R3 | Comment 5 | Sections 4.3 and 4.5; Fig. 6; piezometer-level Exposure input table in the reproducibility package. |
| R3 | Comment 6 | Sections 4.5, 5.5, 6.3–6.5, and 7; LOOCV output files in the reproducibility package. |
| R3 | Comment 7 | Table 1; Sections 3.2 and 4.2; well-completeness information in the reproducibility package. |
| R3 | Comment 8 | Section 4.2; Sections 5.2–5.4; Table 4; Figs. 4–6; SGI notebooks in the reproducibility package. |
| R3 | Comment 9 | Section 4.4; Section 5.2; Appendix B (Fig. B1); Appendix C (Fig. C1). |
| R3 | Comment 10 | Abstract; Introduction; Section 4.5; Section 6.3; Section 7. |
| R3 | Comment 11 | Sections 4.2 and 4.5; Section 5.6; Sections 6.4–6.5; Section 7; Fig. 8. |
| R3 | Comment 12 | Section 4.5; Section 5.5; Sections 6.3 and 6.5; Section 7; Fig. 7; Appendix A. |
| R3 | Comment 13 | Code and data availability; GitHub repository; archived Zenodo release v1.3. |
| R3 | Comment 14 | Abstract; Section 4.5; Table 3; main text; figure captions; appendices; Code and data. |
| R3 | Comment 15 | Section 2.3; Fig. 7; Appendix A (Fig. A1). |
| R3 | Comment 16 | Section 3.2 and descriptions throughout the manuscript; Fig. 1 and associated captions. |
| R3 | Comment 17 | Sections 5.3–5.4, 6.1–6.2, 6.5, and 7; Figs. 5–6. |

**Closing statement**

We thank the Editor and the Referees for their careful and constructive evaluations. Their comments led to substantial corrections to the analytical workflow, improved reproducibility, a more cautious interpretation of statistical and spatial results, and a clearer definition of the framework's intended scope. We believe the revised manuscript now addresses all substantive and technical concerns raised during the discussion stage.