Response to Referee 3

General response

We thank the reviewer 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.

> ***Reviewer 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 reviewer for identifying this inconsistency. The propagation probabilities reported in the original Table 4 were not consistent with the equation given in Table 2. We therefore recalculated the metric from the revised SGI event series and now define propagation probability exclusively as $P_{3 \rightarrow 12} = N_{(SGI3 \rightarrow SGI12)} / N_{(SGI3)}$, 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, $P_{3 \rightarrow 12} = 4/5 = 0.80$.

The event-matching table and worked example are included in the revised reproducibility package so that the numerator and denominator can be verified directly.

> ***Reviewer 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 $CRT = 1 - N_{(\text{drought months})} / N_{(\text{valid months})}$, 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.

> ***Reviewer 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 reviewer 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.

> ***Reviewer 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 reviewer 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 duration, severity, CRT, CRS, vulnerability, recovery, and propagation metrics consistently with Table 4 and Figures 5–6. Table 4 is generated directly from the reproducible output.

> ***Reviewer 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.

> ***Reviewer 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 reviewer. 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 through leave-one-out cross-validation rather than through 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.

> ***Reviewer 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 differ among piezometers in both start and end dates, and several series contain gaps. We removed this statement and now report the actual observation period and 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 changing number of wells through time, whereas both deep wells were required for the deep-aquifer series.

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

> ***Reviewer 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 Φ^{-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.

> ***Reviewer 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 provide a uniquely determined 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 based on moving 10-year windows shifted by one 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.

> ***Reviewer 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.

Reviewer 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 behaviour and may therefore already contain 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 described only 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 ($-0.1 \leq E-P \leq 0.1$), 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.

Reviewer 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 as 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 formulation and the equal-weight formulation is now described as a limited sensitivity check of 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.

Reviewer 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 reviewer 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 now been replaced and archived as **version v1.3**. It includes the 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 reproducing 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 weighted and equal-weight DIPI calculations. 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.

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

Response:

We thank the reviewer for identifying this inconsistency. We corrected the terminology throughout the manuscript and now use “Drought Impact Potential Index” consistently in the title, 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.

> ***Reviewer 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 reviewer 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.

> ***Reviewer comment 16:*** *Lines 144–145 list four aquifer categories, whereas the analysis uses three groups.*

Response:

We thank the reviewer 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.

> ***Reviewer 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.