

Revision of the manuscript “Global-scale drought propagation and the drivers and patterns of multi-year groundwater drought” by Saskia Salwey, Sandra Hauswirth, Denise Ruijsch, Barry van Jaarsveld, Jonna van Mourik, and Niko Wanders

General Comments

In this study, the authors present a relevant global-scale assessment of groundwater drought propagation using a hyper-resolution groundwater model dataset (GLOBGM). The study addresses an important gap in understanding how meteorological drought propagates into groundwater drought and introduces a useful typology of groundwater response behavior. The topic is of strong scientific and societal importance, particularly under increasing climate variability and groundwater stress. The integration of groundwater drought duration metrics, propagation characteristics, and response typologies at the global scale represents a significant contribution to drought science. However, several methodological assumptions and interpretational aspects require clarification and strengthening. In particular, the drought definitions, uncertainty considerations, implications of model limitations, and interpretation of causality need further refinement. A significant revision is necessary.

My major concerns are listed below.

The manuscript makes an important contribution by introducing a global framework for understanding multi-year groundwater drought dynamics using hyper-resolution simulations. The distinction between meteorological and groundwater drought response types is particularly valuable and has strong potential for operational drought monitoring and adaptation planning. However, the novelty claim should be moderated slightly. Several previous studies have investigated drought propagation globally or regionally, and the manuscript would benefit from a clearer explanation of what is fundamentally new here (e.g., hyper-resolution groundwater representation, focus on multi-year groundwater droughts, or the proposed typology framework). The study heavily relies on GLOBGM outputs. Although validation is presented, the model evaluation remains relatively limited.

GLOBGM v1.1 simulates groundwater heads at a hyper-resolution of ~ 1 km, yet its critical boundary conditions, specifically groundwater recharge and abstraction, are derived from PCR-GLOBWB2 at a much coarser 5 arc-minute resolution (~ 10 km). Because recharge is the primary driver of drought propagation into the unconfined saturated zone, any spatial variability in groundwater drought duration at scales smaller than 10 km is structurally forced by static local hydrogeological parameters (e.g., topographically driven boundary conditions or aquifer geometry) rather than meteorological heterogeneity. The authors should explicitly quantify how much of the simulated hyper-resolution variance is a physical reflection of localized climate forcing versus an artifact of spatial downscaling through static model parameterization.

Equation 1 calculates a basic standard $Z(t)$ to define drought conditions. This mathematical formulation implicitly assumes that monthly water table depths follow a Gaussian (normal) distribution. However, groundwater hydrographs, particularly in shallow systems with rapid surface connectivity or in highly arid regions with episodic recharge, are inherently skewed and bounded. Using a linear standard score on non-Gaussian hydrographs severely distorts the target 16% theoretical drought frequency. The authors should justify the normality of their dataset or fit the water table depth time series to an appropriate distribution.

The validation against the IGRAC database relies on 16,639 locations, but a staggering 96.5% of these data points are tightly clustered within North America, Europe, and Australia. This creates a

severe geographic bias, meaning the model's capacity to simulate drought propagation is virtually untested across the Global South, including highly critical semi-arid regions in Africa, Asia, and South America. The authors need to explicitly address how this extreme validation imbalance compromises the reliability of their global-scale response typologies.

The authors acknowledge moderate anomaly correlations and biases in drought duration, but the implications of these limitations on the derived groundwater response types are not sufficiently discussed. The relatively low anomaly correlation values between the model and observational data (0.23 for GRACE, 0.32 for IGRAC) raise concerns about regional model reliability. The authors should expand their discussion on how these weak correlations affect the confidence levels of the proposed global classification types.

The universal use of SPEI-12 deserves further justification. Groundwater systems exhibit widely varying response times, and a fixed accumulation period may artificially create mismatches in some regions. A sensitivity analysis using alternative SPEI accumulation periods (e.g., 3-, 6-, 24-month) would substantially strengthen the manuscript.

At several points, the manuscript interprets groundwater drought duration as being “driven” by subsurface processes or recharge dynamics. However, the analyses are primarily correlational/statistical rather than mechanistic.

One of the largest limitations is the relatively limited treatment of anthropogenic groundwater use. The manuscript acknowledges that abstractions are included in GLOBGM, but the analyses do not separate natural versus anthropogenic controls on drought persistence. This is a critical omission because some of the identified multi-year groundwater drought hotspots may primarily reflect pumping-induced depletion rather than drought propagation alone.

The manuscript would benefit from a clearer explanation of how abstractions are represented, a discussion of regions where pumping likely dominates, and caution when interpreting prolonged droughts as climate-driven.

The proposed response typology is interesting and potentially impactful. However, the physical meaning of each category could be strengthened. Currently, the categories are primarily statistical constructs based on duration comparisons. The manuscript would benefit from linking them more explicitly to aquifer storage properties, recharge seasonality, groundwater depth, or climate regime. The validation section needs further discussion regarding the relatively modest correlation coefficients. The manuscript should better justify why the model performance is sufficient to support strong global-scale conclusions.

Specific Comments

The statement that “35% of the world has an average groundwater drought duration which is multi-year” requires clarification. Is this based on land area?

The uncertainty bounds used for defining groundwater response types are unconventional. Please justify why excluding the shortest and longest event is preferable to bootstrap confidence intervals or nonparametric uncertainty estimates, Lines 185–189. The notation used to define the uncertainty bounds is highly confusing and mathematically counterintuitive. The text states that μ_{\max} represents the mean drought duration excluding the longest event. Mathematically, removing the maximum value from a sample will always decrease the calculated mean. Conversely, excluding the shortest event μ_{\min} will increase the mean. Therefore, μ_{\min} represents the upper bound and μ_{\max} represents the lower bound. However, the condition for Type 1 (meteo < GW) is written

as $\mu_{\text{(meteo-min)}} < \mu_{\text{(GW-max)}}$. This formulation is inverted based on the textual definition. The authors must clarify.

The distinction between “average duration” and “time spent in drought” should be clarified more consistently.

The interpretation of bimodal drought behavior in shallow groundwater systems is intriguing but speculative.

Some regions identified as highly vulnerable may strongly reflect groundwater abstraction impacts rather than natural drought propagation.

The manuscript should more explicitly discuss model structural uncertainty and parameter uncertainty.

Elaborate on why the anomaly correlation coefficients are structurally low across large portions of the globe. Address whether this stems from grid-mismatch errors, missing localized pumping data, or structural limitations of GLOBGM v1.1.

The model includes groundwater abstractions, but they are excluded from the drought typology analysis. Given that abstractions heavily mimic or exacerbate multi-year drought signals, explain how their omission might distort the purely climate-driven classification.

Provide a clearer physical justification or statistical reference for defining "significant pooling" as a mean meteorological overlap > 2 in line 305.

Reconcile the low anomaly correlations with the highly optimistic ROC AUC scores (84% exceeding 0.5). Discuss if the binary nature of ROC thresholds oversimplifies the evaluation compared to continuous time series anomalies. The authors must explicitly discuss this limitation to prevent readers from overestimating the model's continuous predictive capability based solely on the optimistic ROC metrics.

The authors must provide a dedicated section discussing why the temporal synchronicity is so poor and explain how a model with such low phase alignment can reliably classify global drought response types.

The text notes that GLOBGM exhibits reduced variability in monthly water table depths and performs best specifically where water tables are between 5m and 60m, Lines 100–102. Since drought intensity and duration are fundamentally controlled by the amplitude of water table fluctuations, how does this systematically muted variability impact the calculated Drought Duration Ratio (DDR)? A model that dampens water table variability will artificially extend drought duration because it takes longer to recover to the mean, which likely explains why the simulated mean drought duration (9.9 months) is nearly double the observed GRACE duration (5.5 months).

The concept of “tipping points” in shallow groundwater systems is interesting but currently unsupported by direct analysis.

Consider revising lines 151-154 to clarify the temporal integration behavior inherent to groundwater dynamics.

The discussion could benefit from stronger linkage to drought management applications and groundwater early-warning systems.

Minor issues

Typographical and grammatical issues should be corrected.

Please add appropriate references for the examples mentioned, particularly the Millennium Drought in Australia and South Africa's “Day Zero” drought, lines 38-39, to support this statement.

The term “human water demand” is somewhat broad and ambiguous. Consider specifying whether this refers to increased groundwater withdrawals, water consumption, irrigation demand, population growth, industrial use, or overall anthropogenic pressure on groundwater resources under climate change, lines 54-58.

The reference to “future” vulnerability, line 84, may be misleading, as the analysis appears to be limited to the period 1960-2019 and does not include future projections or scenario-based assessments. Consider revising the sentence to clarify that the identified vulnerabilities are based on present-day or historical conditions rather than future changes.

The statement indicates that approximately 85% of the evaluated points showed a positive correlation with local observations (line 100); however, the strength and statistical significance of these correlations are unclear. Please provide additional information on the magnitude of the correlations and whether they were statistically significant

The manuscript states that model performance is best when water table depths are between 5 and 60 m (line 102); however, the reason for this performance range is not explained. A brief discussion of the hydrological and modelling factors contributing to performance range and to reduced performance in very shallow and very deep groundwater systems would strengthen the interpretation.

Since values below -1 generally correspond to near-normal to mild drought conditions, lines 134 and 151, consider evaluating the sensitivity of the results to a more severe threshold (below -1.5) to assess the robustness of the identified drought characteristics.

“DRR” and “DDR” appear inconsistently in parts of the manuscript.

Please define all acronyms in figure captions.

Figure color palettes may be difficult for colorblind readers; consider accessibility-friendly alternatives.

Correct the spacing typo in "Murray- Darling Basin" to "Murray-Darling Basin", Line 44.

Harmonize the terminology between "5 arc-minutes" and "5 arcmin" for consistency, Line 94 & 140.

The text states, “A drought is initiated when this value falls below -1”, Line 134. For strict clarity and adherence to standard run-theory terminology, specify if the threshold is less than or equal to -1.

Correct the grammatical error “fast responding systems” to “fast-responding systems”, Line 295.

In the conceptual schematic of Figure 1, the color assigned to meteo < GW is green. However, in the global spatial maps in Figure 3a, a purple-to-green gradient is used. Ensure that the color anchors used in the conceptual models perfectly match the final data visualizations to improve cross-figure scan ability.