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.

We kindly thank the reviewer for their review of our manuscript. We have addressed your comments in detail below.

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).

Thank you for pointing this out. We will emphasize the novelty of study more clearly in the introduction, discussion and conclusions. We believe that all three of the points you have mentioned make this study unique (the hyper-resolution dataset, the focus on multi-year droughts and the proposed framework).

We will make this clearer at the end of the Introduction by editing L83-86 to read:

In this paper, we use a new global hyper-resolution (~1 km) groundwater dataset (Jaarsveld et al., 2026) produced by the global groundwater model GLOBGM to investigate the global patterns and drivers of groundwater drought from 1960-2019, with a specific focus on multi-year drought events. We start by exploring the propagation from meteorological to groundwater drought, evaluating how and to what extent the subsurface plays a role in modulating the meteorological drought signal. Subsequently, we define three global groundwater response types based on the relationship between meteorological and groundwater drought, providing a new framework for understanding the processes and geo-physical drivers of normal versus multi-year groundwater drought events within each type. By disentangling this relationship at the global scale, our study is the first of its kind to focus specifically on multi-year groundwater drought at the global scale. Our analysis identifies regions that are most prone to multi-year groundwater drought under *present-day conditions, and uses this knowledge to infer which areas may be more vulnerable in the future. By using a global hyper-resolution groundwater dataset, our results uniquely capture the finer-scale interactions between river systems and groundwater whilst also resolving the spatial variability in global groundwater depth. This added detail allows us to generate new insights into the spatial patterns of global drought propagation at actionable resolution, which will help inform strategies for managing, mitigating, and predicting future water scarcity risks.*

We will re-structure the first paragraph of the discussion to read:

To the authors knowledge, this global analysis is the first to use hyper-resolution data to focus specifically on the drivers and patterns of multi-year groundwater drought. The lack of comprehensive global groundwater data has in the past meant that groundwater has been excluded from global drought propagation analysis (Fuentes et al., 2022; Wang et al., 2025; Han et al., 2019), or has been represented by coarse and temporally limited data (Akl et al., 2025). Although drought propagation has received significant attention at local, regional and national scales, the variability in the global patterns and drivers of groundwater drought mean that these findings cannot be easily extrapolated for actionable use.

The conclusion will be edited to read:

This study presents the first hyper-resolution, global-scale analysis of the drivers and patterns of multi-year groundwater drought. We find that in many parts of the world the subsurface significantly modulates the meteorological drought signal, resulting in frequent multi-year groundwater drought. By categorising the global groundwater data into types which describe the responsiveness of the groundwater to the meteorology, we provide a new framework for understanding the global vulnerability to multi-year drought. Our findings offer novel insights into drought planning and mitigation, whilst emphasising the differences in groundwater response across the world.

The study heavily relies on GLOBGM outputs. Although validation is presented, the model evaluation remains relatively limited.

We agree that the model evaluation present in our manuscript is reasonably limited, but rather than duplicate lots of analysis we would like to refer readers to Jaarsveld et al. (2026) for an extensive model evaluation. We will make this clearer, but will also add some additional information on this topic to our manuscript.

L97-104 in the methods have been supplemented such that the text now reads:

The validation of global groundwater models is notoriously difficult due to mismatches in spatial resolution and representation between observed groundwater data and model simulations of groundwater. Nonetheless, the outputs from GLOBGM have been validated in previous work (Verkaik et al., 2024; Jaarsveld et al., 2026). Jaarsveld et al. (2026) found that the GLOBGM non-parametric KGE skill score was more than 0 (suggesting that the model outperforms the mean flow benchmark) in 75% of the simulated area. This proportion increased to 81% after excluding locations where poor skill scores were attributed to metric artifacts associated with weak groundwater trends and very shallow water tables. The model performance is best where water table depths are shallow (0-20 m), but there is also a tendency for it to underestimate groundwater heads between 0-5 m. Contrastingly, groundwater heads were found to be overestimated at deeper water table depths, perhaps due to delays that occur in deeper unsaturated zones which the model does not accurately account for. Notably though, the ability *of the model to reproduce groundwater variability does not appear to be linked to water table depth. GLOBGM has also been evaluated against other, similar, groundwater models. Compared to alternative 5 arc-minute dynamic model simulations (e.g. de Graaf et al. 2017), GLOBGM displays a 19% increase in the number of locations which have a non-parametric KGE skill score more than 0. For more information on the model evaluation we refer to Jaarsveld et al. (2026).*

Since this study focuses on drought, and the prior validation efforts have not considered this aspect of model performance, an extra validation was carried out to assess how effectively GLOBGM simulates patterns of groundwater drought against two sources of observed groundwater data which are introduced below.

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.

The reviewers suggestion for a full variance decomposition that separates static from dynamic contributions is interesting and we understand their rationale for wanting to do this. However. this would require sensitivity experiments which are beyond the scope of the present study. The reviewer is correct in stating that the recharge and abstraction are derived from the coarser 5 arcmin PCRGLOBWB2. However, as described in Section 2.1.2 of Jaarsveld et al. 2026, groundwater recharge is downscaled to 30 arc-seconds using a GAM trained on observed recharge data and 30 arc-second environmental covariates, and statistically corrected. In addition, river discharge is dynamically routed through the 30 arc-second HydroSHEDS network, ensuring that surface water dynamics are represented at high spatial resolutions and capture another important source of groundwater surface water interactions through hydraulically connected river beds. Therefore, although the inputs are derived from 5 arc-minutes resolution inputs, the data fed to GLOBGM is represented as spatially variable field at the resolution it operates. This is analogous to the common practice in climate models, where coarse-resolution GCM output is statistically downscaled and bias corrected to higher resolution (e.g. GSWP3-W5E5). GLOBGM uses the 5-arcminute abstraction fields, which are regirded to 30-arcsecond resolution using bilinear interpolation. This choice reflects both the large uncertainties associated with representing groundwater abstraction at ~1 km resolution and the practical difficulties of doing so reliably with the currently available data.

The reviewer is correct in highlighting the importance of recharge as a driver of drought. However, the conclusion that sub-10 km variability in groundwater drought is solely an artifact of static parameter fields does not reflect how GLOBGM represents the groundwater system. The hydrogeological static properties such as aquifer geometry, hydraulic conductivity, and topography are primary physical controls on groundwater storage and drainage and therefore on strongly control drought duration and intensity. Their spatial variability is not ancillary, but fundamental to shaping groundwater dynamics. For example. even if two adjacent 1 km cells experience identical recharge histories differences in aquifer thickness, hydraulic conductivity and boundary conditions lead to different groundwater recession coefficients, and thus different characteristic drainage and recovery.

Together, the hyper-resolution variance in simulated groundwater drought emerges from the combined effect of localized hydrogeological structure and high-resolution representations of climate-driven recharge and river-aquifer interactions, rather than from static parameterization alone.

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.

We agree with the reviewer that we should have justified this in the first version of our manuscript, we will do so in the subsequent version.

Our analysis on this topic has concluded that there is no statistical distribution that fits the full global groundwater dataset, and thus we use Z-scores as a compromise between good overall fit and computational efficiency. We acknowledge that our assumption of normality is challenged in many areas, particularly those with shallow water tables, or near to draining streams, but further investigations have concluded that alternative distributions such as the log-normal or gaussian distributions do not have a significantly better fit. Furthermore, due to the size and resolution of our dataset, we cannot use an empirical distribution, since it would not be compatible with computational memory constraints at this spatial and temporal resolution. We would also prefer not to fit a gamma or log-normal distribution since these cannot work with negative numbers and therefore would not be compatible when groundwater heads are above the surface (or any other 0-level).

Our approach is in line with many others in the literature, for example Schreiner-McGraw and Ajami (2021), Li and Rodell (2015), Wang et al. (2020), Liu et al. (2023), Pokhrel et al. (2021) who all base their groundwater standardization on a standardized anomaly or Z-score normalization for drought analysis at regional to global scales. But we do recognize that others have preferred to fit log-normal or gamma distributions (Guo et al. 2021; Mukhawana et al. 2024), and that where it is possible, empirical distributions are often can have the best fit (Bloomfield and Marchant 2013).

We will make this limitation clearer in the manuscript by adding the following text and information to the supplement:

To check which statistical distribution best fit our modelled groundwater data, we performed a global, monthly-scale statistical distribution fitting analysis to determine which probability distribution (Normal, Gamma, or Log-normal) best represents the groundwater timeseries (Figure S6). We use the Akaike Information Criterion (AIC) to determine which distribution fits best. If no distribution achieves an AIC value below two then we assume no distribution fits well. We find that in 54% of the world, the normal distribution fits our groundwater data best, and thus justify our use of the standardized monthly anomaly (or Z-score) for calculating drought events. For consistency we used the normal-distribution also in the other regions to obtain consistent geographical drought patterns.

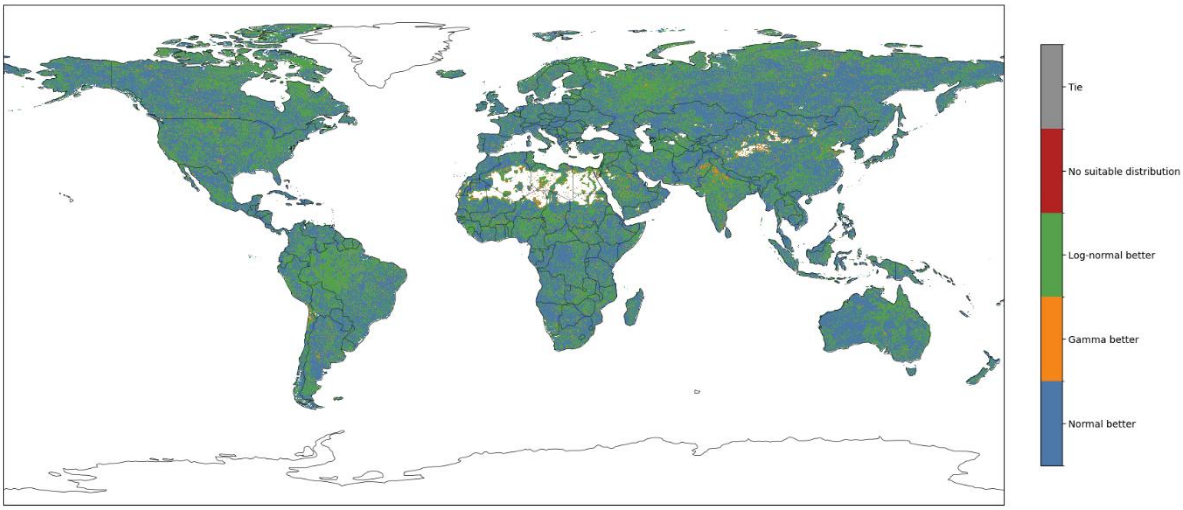

Figure S6. Global pattern of best fitting distribution according to an AIC test. We classify a location as having no suitable distribution when the AIC score is consistently above 2.

We will reference this information in the main text on L135 by saying:

We acknowledge that by using the monthly standardized anomaly we assume that the groundwater data follows a normal distribution, and whilst this is not true everywhere, we found that this distribution fit better than the alternatives (log-normal, gaussian). See supplementary material section S6 for more detail.

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.

We agree that this imbalance in validation data is far from ideal. Unfortunately once we had subset our dataset of groundwater point observations to ensure there was a sufficiently long timeseries this is what we were left with. However, it is for this reason that we also validated the model against the global GRACE data. We agree that as a groundwater community we need to make more progress towards collating suitable evaluation data so that this problem may be overcome in future work. If you have suggestions for additional data we ought to incorporate into our evaluation then we would be happy to do so.

One of the motivations for creating this global hyper-resolution groundwater dataset was to extend groundwater simulations into regions without groundwater monitoring, but unfortunately this does mean analyzing model simulations in regions where the model cannot be properly validated and we have to assume that the model performance is similar to that in well gauged regions.

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.

Thank you for pointing this out. We should have given more context to these anomaly correlations.

We will add a section to the discussion which discusses the implications of the relatively low anomaly correlations on the groundwater response types and the regional reliability. The section paragraph addresses several of the reviewers comments and so has been inserted at the end of this response to reviewers document.

We will also add the following context to the results section at L239:

In several of the areas where the anomaly correlations are lower, we find nuances in the GRACE data which can begin to explain these patterns. Since GRACE is a measurement of total water storage, in tropical regions such as the Amazon or central Africa, the GRACE signal can be dominated by changes in surface water and root zone soil moisture rather than changes in groundwater storage. In more arid regions, the lower anomaly correlations can be partially explained by the time it takes for surface water to reach the groundwater system, meaning that low correlations are not unexpected (Akl et al. 2025; Hohensinn et al. 2026). Previous studies using the IGRAC groundwater point observations have also found challenges when using this data for model evaluation. Jaarsveld et al. (2026) identified that in many of the slow moving systems (e.g. the Sahel, Namibia), the observed and simulated data can show opposing trends, but note that in many of these cases the magnitude of change is very small. As a result in many locations the correlation can over exaggerate the differences in groundwater levels. Finally, we are also aware of flaws in the model input data, for example in sub-Saharan Africa where the forcing data is not always representative (Liu et al. 2024). Nonetheless, there are several regions where it is likely that the model does not have a strong anomaly correlation, and in these places the results should be interpreted with caution.

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.

This is a good point and we should have justified our decision to use SPEI-12 better in the manuscript. We agree that there is clearly huge variation in groundwater response times across the world, and we saw this in our analysis.

We have performed a sensitivity analysis investigating how the distribution of data in each groundwater response type changes using different SPEI accumulation periods and we will add the following findings to the supplementary material to contextualize our results (see below).

In the end, we decided that in order to focus our manuscript on multi-year droughts, we had to maintain a consistent multi-year meteorological drought definition. If we were to vary the accumulation period, the length of a drought which is classified as multi-year would also need to vary to account for the accumulated precipitation and PET in the months leading up to the event. We might also be left with a scenario where neighboring cells exhibit non-natural patterns of multi-year vs normal drought if there is a difference in the optimal accumulation period. We chose SPEI-12 since this is widely used in the literature and thus facilitates easy comparison with other studies (e.g. van Mourik et al. 2025; van de Wiel et al. 2022; Ruijsch et al.

2025; Tarafdar and Dutta 2023; Pascale and Ragona 2025; Theron et al. 2021). A literature review we are compiling on multi-year drought also finds the 12 month accumulation period to be the most common amongst meteorological drought indices (e.g. SPEI and SPI), further increasing the comparability of our study. A further advantage of a 12 month accumulation period is the removal of the seasonal cycle, meaning that the metric is more representative of the long-term water balance.

We will add the following findings to the supplementary material to contextualize our results:

After conducting a sensitivity analysis to evaluate how the SPEI accumulation period impacts the distribution of data across our three groundwater response types, we find that as the SPEI accumulation period increases, more data falls into the meteo>GW category (Figure S7). A longer accumulation period facilitates longer but less frequent meteorological drought, whilst a shorter accumulation period facilitates shorter but more frequent droughts. Whilst we acknowledge that there is variation in groundwater response times across the world, for this study we chose to maintain one constant SPEI accumulation period. We did this in order to maintain a constant multi-year drought definition and to ensure we used a meteorological multi-year drought index which was in line with others used in the wider literature (e.g. van Mourik et al. 2025; van de Wiel et al. 2022; Ruijsch et al. 2025; Tarafdar and Dutta 2023; Pascale and Ragona 2025; Theron et al. 2021).

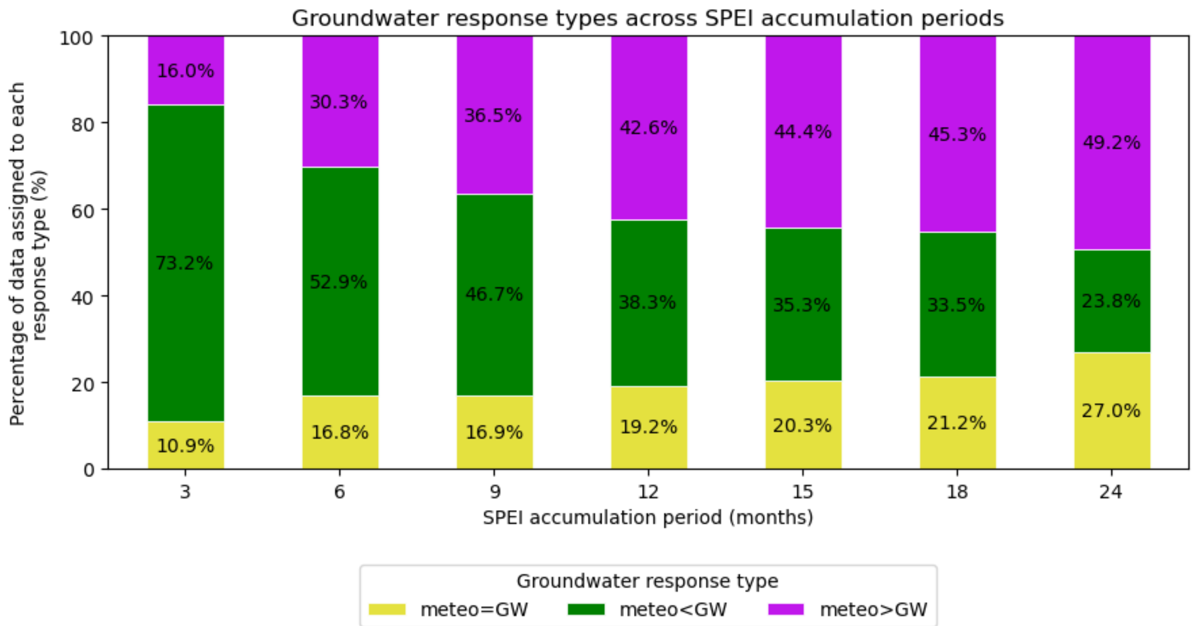

Figure S7. Proportion of data in each groundwater response type for SPEI accumulation periods from 3-24.

We will also add the following text to L415-420 in the discussion:

However, mismatches in SPEI-12 and groundwater accumulation still result in a portion of data falling into the meteo>GW type. In these locations the groundwater responds to the meteorology faster than the 12 month accumulation inherent in SPEI-12 and this results in shorter groundwater drought events. Since the main focus of this study was to identify the drivers of multi-year groundwater droughts, and they are not common in these fast responding regions (there are no places with a multi-year average groundwater drought duration in the meteo>GW type), this has not hindered our analysis, but should be considered if this analysis was to focus *on the drivers of shorter events. For a sensitivity analysis which demonstrates how the SPEI accumulation period impacts the distribution of data across our typology see supplementary material section S7.*

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.

We will change this language to correct for this.

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.

Thank you for pointing this out. We will add the following text to the Difficulties associated with drought definition section in the discussion.

Furthermore, in many locations it is also not possible to distinguish the climatological controls on drought from the anthropogenic controls. Our study assumes that drought (and drought duration) are a reflection of the meteorological signal, when in some cases the drought and its persistence are actually a function of human activity (e.g. groundwater pumping). Therefore in regions with a high level of human interference, groundwater drought should not be considered as solely the manipulation of the meteorological signal.

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.

Thank you for pointing this out. We should have given this more attention and will make sure to do this in the revised version. We will add the following text to the methods to give more context surrounding how abstractions are represented:

The groundwater abstractions in GLOBGM are based on data from PCR-GLOBWB2 (Sutanudjaja et al. 2018). In PCR-GLOBWB2, the water withdrawal is equal to the gross water demand, which made up of both irrigation and non-irrigation sources. The demand from irrigation is based on the crop composition and the irrigated area and is derived from MIRCA2000 (Portmann et al., 2010) and the Global Crop Water Model (Siebert and Döll, 2010). The non-irrigation demand covers three sectors (industry, households and livestock). For more information on the calculation of this demand we refer to Wada et al. (2014). In PCR-GLOBWB2, water demand can be fulfilled from three sources: surface water, groundwater and desalinated water. The fraction of water which is abstracted from groundwater and surface water is calculated based on their relative abundance, where the abundance is calculated based on a 2-year running mean of recharge and discharge. Based on this, PCR-GLOBWB2 calculates how much water is abstracted from each source each month. The total groundwater abstraction is then given as an input to GLOBGM. Groundwater abstractions are highest in many of the recognized water scarcity hot-spots, with the largest average abstractions across the Indo-Gangetic Plain, northern China and western and central United States (see supplementary material Figure S8). Declining water table depth trends in the GLOBGM simulations align well with known regions of excessive irrigation, groundwater abstraction and depletion e.g. the United States, the Arabian

Peninsula and the Indo-Gangetic plain (Jaarsveld et al. 2026). For more information on the groundwater abstraction data we refer users to Sutanudjaja et al. (2018).

We will also add the following figure to the supplement to show the global pattern of modelled groundwater abstractions:

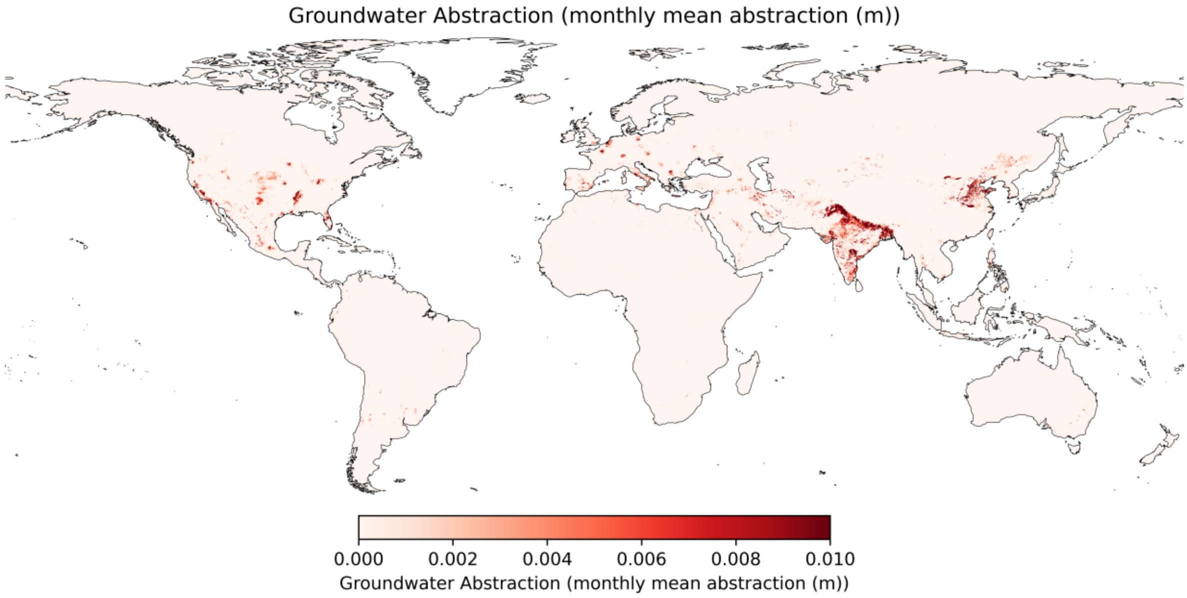

Figure S8. Global groundwater abstractions present in GLOBGM.

We will discuss the model performance in abstraction regions in more detail by adding the following text to the results:

L239: *The model performs slightly better than average in the regions with abstractions in the 95th percentile. In these areas the mean anomaly correlation with the GRACE data is 0.36 (compared to the global average of 0.23) and the median is 0.39 (compared to the global average of 0.27).*

L266: *Multi-year droughts are also common amongst regions with high groundwater abstractions, where areas which have modelled abstractions in or above the 95th percentile have a mean drought duration of 14.7 months. The median, however, is only 5.8 months, suggesting that the mean is skewed up by long but less frequent drought events. In general, the worlds water scarcity hot spots (e.g. the Murray-Darling, Central Valley in California, Spain and the North China Plain) tend to have higher DRR's and longer groundwater drought durations, but there is some variation, for example in the Ganges where areas with lower DRR are still present.*

L323: *The anomaly correlations in regions with high groundwater abstractions (e.g. areas in the 95th percentile) are comparable to the global mean (mean anomaly correlation = 0.36 and median = 0.39).*

We will also elaborate on how our approach to modelled groundwater abstractions impacts our findings in the discussion. We will add a subsection entitled 'Model Limitations' which addresses several of the reviewers comments and has been inserted at the end of this response to reviewers document.

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.

Thank you for pointing this out. We have added some additional analysis which we will refer to in the main manuscript and present in more detail in the supplement.

L287: *Areas with deeper groundwater have a higher proportion of data falling into this category (55% of the groundwater tables deeper than 53 m falls in the meteo<GW type), as do regions classified as arid (56% of data which has a Koppen climate classification of arid falls into the meteo<GW type).*

L296: *A large amount of data still falls into the meteo>GW type (42%), but almost all of these locations have a normal groundwater drought duration. For these regions we see that anomalies in the groundwater system typically have a faster response time than the 12 months inherent in SPEI-12. Although these fast responsing systems make up a large portion of the meteo>GW type, there are also locations with very shallow groundwater or strong connectivity to the surface water that exhibit similar behaviour. We will discuss this in more detail in the Discussion. In general, regions that fall into the meteo>GW response type tend to be shallower (58% of groundwater tables less than 5m deep fall into this response type) and are more likely to be in Tropical and Temperate regions. See supplementary material section S9 for more information on how the climate regimes and groundwater depth relate to the response types.*

We will add the following to a new section in the supplementary material, section S9:

To link the groundwater response types to local characteristics, Figures S9 and S10 display the distribution of data across groundwater response types for different classes of water table depth and for climate zones according to the Koppen climate classification. Shallower water tables (0-15m) tend to be more associated with the meteo>GW response type, whilst deeper water tables (15m +) are more likely to be in the meteo<GW type (Figure S9). When considering the climate types, arid regions have the highest proportion of data in the meteo<GW type, and tropical regions have the largest proportion of data in the meteo>GW type, closely followed by temperate regions (Figure S10). These findings are in line with others in the literature, where arid climate zones have been found to have the weakest relationship with precipitation (Liu et al. 2023), and where deeper groundwater is known to be more disconnected from the meteorology (Hare et al. 2021, Ebeling et al. 2025).

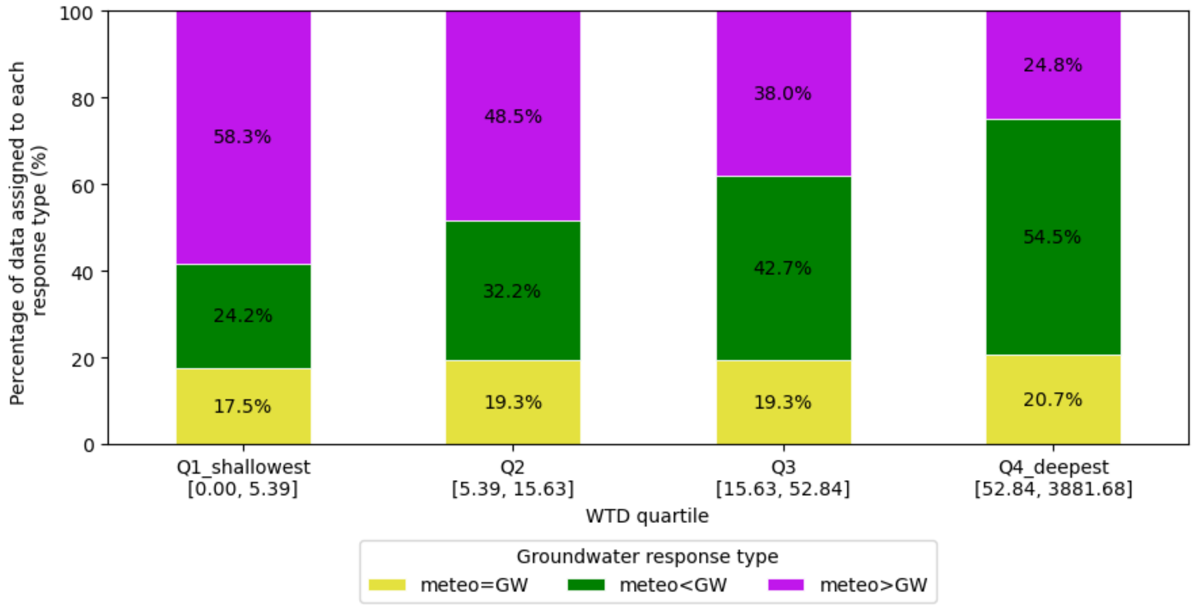

Figure S9. Distribution of data across groundwater response types for four quantiles of water table depth.

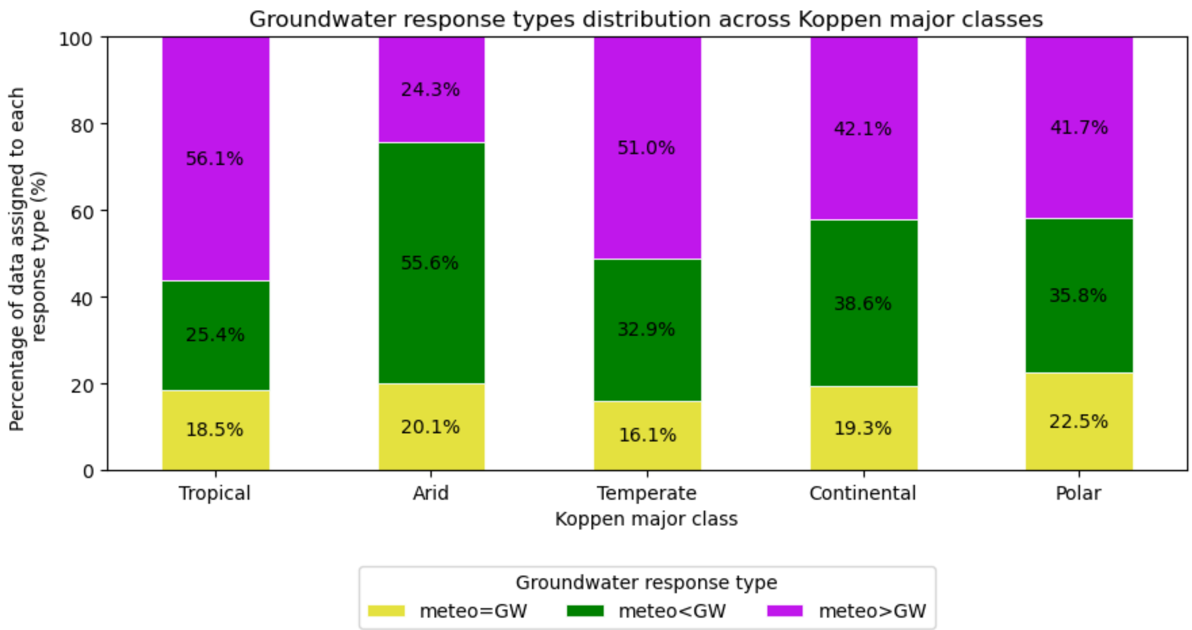

Figure S10. Distribution of data across groundwater response types for four quantiles for Koppen climate types (Beck et al. 2023).

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.

We will add this to the discussion. The additional section addresses several of the reviewers comments and has been inserted at the end of this response to reviewers document.

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?

We will clarify this. It is based on the simulated 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 ⍰_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 ⍰_min will increase the mean. Therefore, ⍰_min represents the upper bound and ⍰_max represents the lower bound. However, the condition for Type 1 (meteo < GW) is written as ⍰_(meteo-min) < ⍰_(GW-max). This formulation is inverted based on the textual definition. The authors must clarify.

Our method was designed to examine whether or not the classification depends on one anomalous long or short drought event. This approach most closely resembles a leave on out validation where we recompute the means leaving out one of the samples. As the shortest and longest droughts have the most impact we reduced the methodology to just leave these out and estimate the associated changes in the mean. The objective was to evaluate robustness to rare events rather than estimate sampling uncertainty. We did consider using bootstrapping, but since this method was substantially more computationally demanding when applied to our global, hyper-resolution dataset, we instead adopted this leave-one-extreme-out approach as a computationally efficient robustness diagnostic.

We will clarify the textual definition to make this easier to interpret. Type 1 represents places where on average the meteorological droughts are shorter than groundwater droughts, so by checking that the mean meteorological drought duration minus the shortest event (the upper bound) is still less than the mean groundwater drought duration minus the longest event (lower bound), we can more confidently say that meteorological drought events are on average shorter than groundwater drought events.

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

We will adjust this language to be more consistent.

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

Yes, we decided not to go into this in detail in this manuscript. We will emphasize that this is speculative and should be looked into more deeply in future work.

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

As part of the additional text describing the impacts of groundwater abstractions on drought we will highlight this.

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

We will comment on this in the limitations section of the discussion and also have added the additional clarification with regard to the general groundwater model performance as obtained from Jaarsveld et al. (2026). This will give the reader a better understanding of the major assumptions and uncertainties within the model simulations.

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.

We have added additional text to L239 which addresses this. The new text has been quoted earlier in our response in reply to a similar comment.

We will also comment on this in a new section of the manuscript which has been inserted at the end of this response to reviewers.

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.

We will add a discuss this in our new Model Limitations section which has been inserted at the end of this response to reviewers document.

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

We will adjust this language to avoid ambiguity simply mention that the pooling metric is larger than two.

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.

We will discuss this as part of the discussion. We will add the following text:

While the ROC AUC scores show a skillful representation of the model, we do also observe that the magnitude of the anomalies in groundwater simulations could be better represented, as shown by the lower anomaly correlations. This indicates that while the groundwater anomalies are represented in the correct direction (either too wet or too dry) their magnitude is not always correctly estimated.

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.

This will be incorporated into the new section of our discussion which we will insert at the end of this response to reviewers.

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).

We agree with the reviewer that a reduction in the temporal variability could in part lead to an overestimation of the groundwater drought durations. We will add this into the discussion of the manuscript.

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

Yes, we decided not to go into this in detail in this manuscript. We will emphasize that this is speculative and should be looked into more deeply in future work.

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

We will change the lines to:

Each meteorological drought can be considered a function of the combined effects of the meteorological anomalies of the last year. This is not the case for a groundwater drought, which by definition integrates the meteorological anomalies. However the groundwater drought in each month is defined only based on the current month assuming that the current groundwater state is already an accumulation of the anomalies in the incoming groundwater recharge over the last 12 months.

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

We will add further discussion on this topic to the existing section. The text will read as follows:

Understanding the dynamics of multi-year drought propagation can inform local drought mitigation and help to predict and prepare for drought impacts (Minea and Albulescu, 2025; Parry et al., 2018). This is particularly important in the context of multi-year groundwater drought, since groundwater is often used to supplement surface water and sustain vegetation (Mu et al., 2021), so a prolonged absence of this resource can have wide-reaching implications. The typology that we implement in this study provides a simple baseline for understanding global patterns of groundwater responsiveness, which in turn can inform drought mitigation. For example, in regions where the meteorology and the groundwater are connected (which in our analysis makes up 19% of the world), the magnitude and timing of a drought can be anticipated based on the observed anomalies in the meteorological conditions. This has significant advantages for early-warning systems, which can be particularly beneficial for managing groundwater abstractions and the conjunctive use of groundwater and surface water. If we know the systems to be closely connected, then the end of a meteorological drought (which is much easier to measure and observe) is likely to also signal the end of a groundwater drought. It is also beneficial to understand whether or not a location falls into the meteo<GW type, since in these cases even a short meteorological drought might signal the start of a multi-year groundwater drought for which society needs to prepare. Contrastingly, in places where the

meteorological droughts tend to be longer than the groundwater droughts, a multi-year groundwater drought is extremely unlikely since multi-year meteorological droughts are rare.

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.

We will add references to van Dijk et al. (2013), Cai et al. (2014), Sousa et al. (2018) here which describe the severity of these events.

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.

We chose this term to encompass many of the factors you have mentioned. We will clarify what this term includes.

It is estimated that 71% of global groundwater aquifers are already declining, with approximately 30% demonstrating an acceleration in decline over recent decades. This trend is only projected to intensify under continued climate change and the resultant increases in human water demands (e.g. demands from irrigation, population growth, drinking water etc.) (Jasechko et al., 2024; Wunsch et al., 2022; Nazari et al., 2025; Bierkens et al. 2019). Future groundwater droughts are likely to be more frequent, but might also last longer, leading to an increase in multi-year droughts (Wunsch et al., 2022).

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.

We will correct this sentence to read:

Our analysis identifies regions that are most prone to multi-year groundwater drought under present-day conditions, and uses this knowledge to infer which areas may be more vulnerable in the future.

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

We have added more information to this paragraph to clarify this.

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.

We have added more information to this paragraph to clarify this.

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.

We have opted not to do this since re-computing the drought events with a new threshold at the global scale would use a large amount of compute resources and the -1 threshold is widely used in the literature.

"DRR" and "DDR" appear inconsistently in parts of the manuscript.

We will correct this. It should be DDR.

Please define all acronyms in figure captions.

We will do this in the revised manuscript.

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

We have tested this with our colorblind co-author who has suggested it is okay, but if you can specify the plot where this is an issue we are happy to change it. We also saw no problems running the figures through a colorblind filter.

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

We will correct this.

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

We will correct this.

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.

We will correct this.

Correct the grammatical error "fast responsing systems" to "fast-responding systems", Line 295.

We will correct this.

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.

This is a mistake, thank you for pointing it out. We have fixed it in the pre-print and will carry this fix over to the revised manuscript.

The following new section will be added to the discussion, entitled Model Limitations.

The validation of global groundwater models is notoriously difficult due to mismatches in spatial resolution and representation between observed groundwater data and model simulations of groundwater. The relatively low anomaly correlations calculated between observed and simulated data should influence the confidence we have in the classification of groundwater response types and interpretation of the data in regions where the model does not perform well or can not be evaluated. In regions where the anomaly correlation is low, the model may not be well suited for regional use and the results must be interpreted with caution. There are several reasons why the anomaly correlations may be low which are unrelated to GLOBGM. Firstly, because of errors in the observational dataset. As discussed in Results section 3.1, in many places the GRACE data is not suitable for direct comparison to the groundwater data, and the signal should not be compared to groundwater table depth. This has been observed by many other studies in the literature who have identified problems with using this data for drought-based evaluation (Akl et al. 2025; van Loon et al. 2017; Long et al. 2013). Although we tried to couple the GRACE evaluation with perhaps more reliable point observations, these are reasonably scarce. As a groundwater community, we need to push for more global groundwater data to aid model evaluation, but in its absence, we suggest that there is still value in running models in regions where they cannot be validated. Since GLOBGM has been proven to perform as well as, or better than many alternative global groundwater models (Jaarsveld et al. 2026), we believe there is still value in using the results for global analysis. Currently the model's performance is on par or better than other global dynamic groundwater model simulations. Choosing suitable metrics for model evaluation is also tricky, and in this case, in many areas the anomaly correlation could be over exaggerating a mis-match in simulated and observed data in regions with little water table variability where small trends oppose each other. It is also likely that the meteorological forcing input to GLOBGM could be decreasing the anomaly correlations in some regions such as sub-Saharan Africa and north-eastern Siberian (Jaarsveld et al. 2026; Liu et al. 2024). Uncertainties remain large for the sub-surface parameterization, however in the absence of better parameterization we have used the state-of-the-art available parameters to perform GLOBGM groundwater simulations.

It is possible that the groundwater abstractions implemented in GLOBGM contribute to poor model performance in some regions. However, in general, GLOBGM simulates large scale groundwater depletion signals well, identifying many of the established regions of known depletion e.g. the Arabian Peninsula and California's Central Valley. It also recreates some of the regions where groundwater supplies are recovering e.g. the Guarani Aquifer System in South America (Jaarsveld et al. 2026).

References

Akl, M., Thomas, B.F. and Clarke, P.J., 2025. Global groundwater drought assessment revisited: A holistic re-evaluation of the GRACE-groundwater drought index across major aquifers. Water Resources Research, 61(12), p.e2025WR040389.

Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A.I. and Miralles, D.G., 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. Scientific Data, 10(1), p.724.

Bloomfield, J.P. and Marchant, B.P., 2013. Analysis of groundwater drought building on the standardised precipitation index approach. Hydrology and Earth System Sciences, 17(12), pp.4769–4787.

Cai, W., Purich, A., Cowan, T., van Rensch, P. and Weller, E., 2014. Did climate change-induced rainfall trends contribute to the Australian Millennium Drought? Journal of Climate, 27(9), pp.3145–3168.

de Graaf, I.E.M., van Beek, R.L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E.H. and Bierkens, M.F.P., 2017. A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. Advances in Water Resources, 102, pp.53–67.

Ebeling, P., Musolff, A., Kumar, R., Hartmann, A. and Fleckenstein, J.H., 2025. Groundwater head responses to droughts across Germany. Hydrology and Earth System Sciences, 29(13), pp.2925–2950.

Fuentes, I., Padarian, J. and Vervoort, R.W., 2022. Spatial and temporal global patterns of drought propagation. Frontiers in Environmental Science, 10, p.788248.

Guo, M., Yue, W., Wang, T., Zheng, N. and Wu, L., 2021. Assessing the use of standardized groundwater index for quantifying groundwater drought over the conterminous US. Journal of Hydrology, 598, p.126227.

Han, Z., Huang, S., Huang, Q., Leng, G., Wang, H., Bai, Q., Zhao, J., Ma, L., Wang, L. and Du, M., 2019. Propagation dynamics from meteorological to groundwater drought and their possible influence factors. Journal of Hydrology, 578, p.124102.

Hare, D.K., Helton, A.M., Johnson, Z.C., Lane, J.W. and Briggs, M.A., 2021. Continental-scale analysis of shallow and deep groundwater contributions to streams. Nature Communications, 12(1), p.1450.

Hohensinn, R., Gou, J., Meyer, U., Humphrey, V., Dorigo, W., Boergens, E., Soja, B., Gruber, A., Jensen, L., Eicker, A. and Rast, M., 2026. Water storage trends derived from the GRACE/-FO global gravity-based groundwater product (G3P). EGUsphere, pp.1–42.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R.G., Fallatah, O. and Kirchner, J.W., 2024. Rapid groundwater decline and some cases of recovery in aquifers globally. Nature, 625(7996), pp.715–721.

Knoben, W.J., Freer, J.E. and Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), pp.4323–4331.

Li, B. and Rodell, M., 2015. Evaluation of a model-based groundwater drought indicator in the conterminous US. Journal of Hydrology, 526, pp.78–88.

Liu, R., Zhang, X., Wang, W., Wang, Y., Liu, H., Ma, M. and Tang, G., 2024. Global-scale ERA5 product precipitation and temperature evaluation. Ecological Indicators, 166, p.112481.

Liu, Y., Shan, F., Yue, H., Wang, X. and Fan, Y., 2023. Global analysis of the correlation and propagation among meteorological, agricultural, surface water, and groundwater droughts. Journal of Environmental Management, 333, p.117460.

Long, D., Scanlon, B.R., Longuevergne, L., Sun, A.Y., Fernando, D.N. and Save, H., 2013. GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. Geophysical Research Letters, 40(13), pp.3395–3401.

Minea, I. and Albulescu, A.-C., 2025. Drought over time: An investigation of the impact of meteorological drought on groundwater and surface water in the East of Romania. Earth Systems and Environment, 9, pp.1959–1981.

Mu, M., De Kauwe, M.G., Ukkola, A.M., Pitman, A.J., Guo, W., Hobeichi, S. and Briggs, P.R., 2021. Exploring how groundwater buffers the influence of heatwaves on vegetation function during multi-year droughts. Earth System Dynamics Discussions, pp.1–29.

Mukhawana, M.B., Kanyerere, T., Kahler, D., Masilela, N.S., Lalumbe, L. and Umunezero, A.A., 2024. Hydrological drought assessment using the standardized groundwater index and the standardized precipitation index in the Berg River Catchment, South Africa. Journal of Hydrology: Regional Studies, 53, p.101779.

Nazari, S., Reinecke, R. and Moosdorf, N., 2025. Global estimates of groundwater withdrawal trends and uncertainties. Environmental Research Letters.

Parry, S., Wilby, R., Prudhomme, C., Wood, P. and McKenzie, A., 2018. Demonstrating the utility of a drought termination framework: Prospects for groundwater level recovery in England and Wales in 2018 or beyond. Environmental Research Letters, 13, p.064040.

Pascale, S. and Ragone, F., 2025. Widespread Multi-Year Droughts in Italy: Identification and Causes of Development. *International Journal of Climatology*, *45*(8), p.e8827.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S.N., Grillakis, M., Gudmundsson, L. and Hanasaki, N., 2021. Global terrestrial water storage and drought severity under climate change. Nature Climate Change, 11(3), pp.226–233.

Portmann, F.T., Siebert, S. and Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global biogeochemical cycles*, *24*(1).

Ruijsch, D., van Mourik, J., Biemans, H., Hauswirth, S.M. and Wanders, N., 2025. Thrive or wither: Exploring the impacts of multi-year droughts on vegetation. Journal of Geophysical Research: Biogeosciences, 130, p.e2025JG008992.

Schreiner-McGraw, A.P. and Ajami, H., 2021. Delayed response of groundwater to multi-year meteorological droughts in the absence of anthropogenic management. Journal of Hydrology, 603, p.126917.

Siebert, S. and Döll, P., 2010. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *Journal of Hydrology*, *384*(3-4), pp.198-217.

Sousa, P.M., Blamey, R.C., Reason, C.J., Ramos, A.M. and Trigo, R.M., 2018. The 'Day Zero' Cape Town drought and the poleward migration of moisture corridors. Environmental Research Letters, 13(12), p.124025.

Sutanudjaja, E.H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H.C., Drost, N., van der Ent, R.J., de Graaf, I.E.M., Hoch, J.M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M.W., Vannametee, E., Wisser, D. and Bierkens, M.F.P., 2018. PCR-

GLOBWB 2: A 5 arcmin global hydrological and water resources model. Geoscientific Model Development, 11, pp.2429–2453.

Tarafdar, S. and Dutta, S., 2023. Long-term decline in rainfall causing depletion in groundwater aquifer storage sustaining the springflow in the middle-Himalayan headwaters. *Journal of Earth System Science*, *132*(3), p.124.

Theron, S.N., Archer, E., Midgley, S.J.E. and Walker, S., 2021. Agricultural perspectives on the 2015-2018 Western Cape drought, South Africa: Characteristics and spatial variability in the core wheat growing regions. *Agricultural and Forest Meteorology*, *304*, p.108405.

van der Wiel, K., Batelaan, T.J. and Wanders, N., 2023. Large increases of multi-year droughts in north-western Europe in a warmer climate. Climate Dynamics, 60, pp.1781–1800.

van Dijk, A.I., Beck, H.E., Crosbie, R.S., de Jeu, R.A., Liu, Y.Y., Podger, G.M., Timbal, B. and Viney, N.R., 2013. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resources Research, 49(2), pp.1040–1057.

van Jaarsveld, B., Wanders, N., Otoo, N.G., Sutanudjaja, E.H., Verkaik, J., Zamrsky, D. and Bierkens, M.F.P., 2026. Global hyper-resolution modeling of historical and future groundwater dynamics. EGUsphere, pp.1–37.

van Loon, A.F., Kumar, R. and Mishra, V., 2017. Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time. Hydrology and Earth System Sciences, 21(4), pp.1947–1971.

van Mourik, J., Ruijsch, D., van der Wiel, K., Hazeleger, W. and Wanders, N., 2025. Regional drivers and characteristics of multi-year droughts. Weather and Climate Extremes, 48, p.100748.

Verkaik, J., Sutanudjaja, E.H., Oude Essink, G.H.P., Lin, H.X. and Bierkens, M.F.P., 2024. GLOBGM v1.0: A parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW global-scale groundwater model. Geoscientific Model Development, 17, pp.275–300.

Wada, Y. and Bierkens, M.F.P., 2014. Sustainability of global water use: Past reconstruction and future projections. Environmental Research Letters, 9, p.104003.

Wang, F., Wang, Z., Yang, H., Di, D., Zhao, Y. and Liang, Q., 2020. Utilizing GRACE-based groundwater drought index for drought characterization and teleconnection factors analysis in the North China Plain. Journal of Hydrology, 585, p.124849.

Wang, S., Huang, S., Wang, C., Zhang, X., Wu, J., Gulakhmadov, A., Niyogi, D. and Chen, N., 2025. Global anthropogenic effects on meteorological–hydrological–soil moisture drought propagation: Historical analysis and future projection. Journal of Hydrology, 653, p.132755.

Wunsch, A., Liesch, T. and Broda, S., 2022. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. Nature Communications, 13(1), p.1221.