Response to Referee #2

Critical Review of the Study: "Groundwater Head Responses to Droughts Across Germany"

The study 'Groundwater Head Responses to Droughts Across Germany' by Pia Ebeling et al. (2024) provides an extensive analysis of groundwater level variations in response to meteorological anomalies over the past three decades. Using a data-driven approach, the study successfully classifies regional groundwater responses and highlights the spatial heterogeneity of aquifer dynamics. While this research is a significant contribution to hydroinformatics and water resource management, there are several areas where its accessibility and interpretation could be improved.

Response: Dear reviewer, we thank you for your assessment of our work and constructive feedback to further improve our study's impact. We address all your comments in detail below.

R2.C1 1. Scientific Significance

The manuscript represents a substantial contribution to scientific progress within the field of hydrology, particularly in the context of groundwater response to climate variability. Several aspects of the study highlight its scientific impact:

- Innovative Use of Large-Scale Groundwater Monitoring Data: By analyzing groundwater head responses from 6,626 wells across Germany, this study provides an unprecedented spatially comprehensive assessment of groundwater dynamics.

- Application of Machine Learning Techniques: The clustering approach offers a novel perspective on groundwater system behavior, moving beyond traditional statistical approaches to classify aquifer responses.

- Integration of Hydrometeorological Indices: The study effectively combines groundwater data with standardized precipitation and evapotranspiration indices (SPI, SPEI, and SGI), providing a robust framework for assessing groundwater vulnerability.

- Relevance to Climate Change Adaptation: The study identifies long-term trends in groundwater behavior, contributing to the understanding of how future climate extremes may affect groundwater resources.

Overall, the research presents substantial new data and methodological advancements that contribute significantly to the broader field of hydrology and environmental science.

Response: We appreciate your positive assessment of our work and summary of identified highlights.

R2.C2 2. Scientific Quality

The scientific approach and applied methods used in this study are largely valid and appropriate for the research objectives. The methodology is well-structured, employing statistical and machine learning techniques to analyze groundwater responses to meteorological anomalies. However, some aspects could be improved to strengthen the scientific rigor and clarity of the study:

Response: Thank you very much for the overall positive feedback and constructive comments for further improvements which we address below.

R2.C2.1 2.1. Validity of Methods

The study effectively integrates large-scale groundwater monitoring data with meteorological variables using machine learning techniques. While this approach provides valuable insights, a comparison with traditional statistical methods (e.g., Bayesian modeling) could enhance the robustness of the findings. Additionally, justifying the selection of specific clustering algorithms and their interpretability in hydrological contexts would improve transparency.

Response: Bayesian statistics are commonly used to quantify uncertainties and is able to integrate prior knowledge. Thus, it is applied to estimate model uncertainties related to parameter estimation (e.g. Yin et al. 2021, https://doi.org/10.1016/j.jhydrol.2021.126682) or for probabilistic forecasting (e.g. for groundwater contamination Yan et al. 2019, https://doi.org/10.1016/j.jhydrol.2019.124160), and also, more recently, Bayesian optimization is applied for hyperparameter tuning of machine learning models to forecast groundwater levels (e.g., Zhu et al. 2025, https://doi.org/10.1016/j.jhydrol.2024.132567). We are not aware of any applications of Bayesian statistics to model distinct groundwater response patterns to hydroclimatic inputs and link those to spatial controls to understand underlying dominant processes. For this purpose, we used two machine learning methods: (1) the kmeans clustering as an unsupervised method to group the time series of groundwater responses into similar groups and (2) random forest models to link the spatial controls to the observed characteristics.

(1) The kmeans clustering approach groups individuals (here groundwater wells) based on their (dis)similarity in an unsupervised form, i.e. no prior knowledge and assumptions are required. This dissimilarity between two time series is quantified based on Euclidean distance, calculated as the square root of the sum of the squared differences between the corresponding SGI values at each time step. This implies that time series that are fully aligned in their dynamics (and exhibit a similar amplitude) are more similar than, for example, temporarily shifted time series. As Euclidean distance takes squared differences, it is sensitive to noise and single extreme differences. However, it is efficient and thus well suited for large-sample data sets and in particular applicable for time series of the same length.

    We will add this information in the corresponding method Section 2.3:
    "The Euclidean distance measures (dis-)similarity based on the squared differences of two SGI time series, making it sensitive to extreme differences and temporal shifts but also computationally efficient."

(2) The random forest models were used to identify controls of observed characteristics due several advantages regarding data structures and efficiency. Please, refer to our response to R1.C10 from reviewer 1 for more details.

    We will add reasons for the use of the RF approach in the manuscript. However, further methodological comparisons like Bayesian modeling although interesting are beyond the scope of this study.

> *"RFs are particularly well-suited for efficiently handling large datasets, managing collinearity among descriptors through random feature selection, and identifying complex non-linear relationships without a priori assumptions. Moreover, they are robust to outliers and noise due to their ensemble approach averaging across trees."*

R2.C2.2 2.2. Discussion and Consideration of Related Work

The discussion is generally well-balanced and considers the implications of the findings within the broader field of groundwater hydrology. However, the study could benefit from a more comprehensive comparison with existing research on groundwater responses to droughts in other regions. Citing and discussing more recent studies that have used similar data-driven approaches would strengthen the context of the results.

Response: Thank you for this comment, we will add more references to discuss our work within existing data-driven studies on groundwater droughts. More specifically, we will add

References:

Brakkee, E., van Huijgevoort, M. H. J., and Bartholomeus, R. P.: Improved understanding of regional groundwater drought development through time series modelling: the 2018–2019 drought in the Netherlands, Hydrol. Earth Syst. Sci., 26, 551–569, https://doi.org/10.5194/hess-26-551-2022, 2022.

Chávez García Silva, R., Reinecke, R., Copty, N.K. et al. Multi-decadal groundwater observations reveal surprisingly stable levels in southwestern Europe. Commun Earth Environ 5, 387 (2024). https://doi.org/10.1038/s43247-024-01554-w

Schreiner-McGraw, A. P. and Ajami, H.: Delayed response of groundwater to multi-year meteorological droughts in the absence of anthropogenic management, Journal of Hydrology, 603, 126917, https://doi.org/10.1016/j.jhydrol.2021.126917, 2021

R2.C2.3 2.3. Appropriateness of References

The references used in the study are relevant and appropriate, but a few additional sources on groundwater modeling and long-term hydrological trends could be included to further support the study's conclusions. Additionally, ensuring that all citations are up to date would improve the scientific quality.

Response: We plan to add additional recent references regarding long-term trends in groundwater, e.g. Chavez et al. 2024 and Jasechko et al. 2024, also in response to your previous comment. We will also add modelling studies investigating changes in groundwater recharge in Europe and Germany to the discussion, for example

Hellwig, J., Stoelzle, M., and Stahl, K.: Groundwater and baseflow drought responses to synthetic recharge stress tests, Hydrol. Earth Syst. Sci., 25, 1053–1068, https://doi.org/10.5194/hess-25-1053-2021, 2021.

Kumar, R., Samaniego, L., Thober, S., Rakovec, O., Marx, A., Wanders, N., et al. (2025). Multi-model assessment of groundwater recharge across Europe under warming climate. Earth's Future, 13, e2024EF005020. https://doi.org/10.1029/2024ef005020

Wunsch, A., Liesch, T. & Broda, S. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. Nat Commun 13, 1221 (2022). https://doi.org/10.1038/s41467-022-28770-2

R2.C3 3. Strengths of the Study

The study demonstrates several strengths that make it a valuable contribution to groundwater research:

- Extensive Dataset: The study utilizes data from 6,626 monitoring wells, ensuring a comprehensive representation of groundwater variability across Germany.

- Advanced Analytical Techniques: The use of machine learning clustering provides a robust classification of groundwater responses, enabling a more detailed understanding of regional differences.

- Temporal and Spatial Analysis: The research effectively captures the variability in response times, showing how groundwater systems react to droughts over different timescales.

- Relevance to Climate Change Adaptation: The study offers crucial insights for policymakers and water managers, helping to design more effective drought mitigation strategies.

Response: Thank you very much for recognizing these strengths of our study.

R2.C4 4. Recommendations for Enhancement

To maximize the impact and accessibility of this research, the following recommendations are suggested:

Response: Please, see our responses below.

R2.C4.1 1. Explicitly link cluster classifications to hydrological processes rather than treating them as purely data-driven groupings.

Response: We will improve the accessibility of the clusters through reducing use of acronyms and referring to their meanings, which we will also add in a new Table highlighting the findings to make them more easily accessible at a glance and in connection to the cluster names and properties. Throughout the text, we will expand several formulations to more clearly address the processes represented by the clusters. However, we would like to note that although we identified characteristics more prevalent in certain clusters that reflect underlying dominant processes, these linkages are non-unique, thus naming clusters after the mechanisms would oversimplify the complexity of interacting processes and the ambiguity in the linkages. For this reason, the clusters are named after the observed patterns. Please, also refer to Comment R1.C1.1 from reviewer 1 and R2.C4.2.

R2.C4.2 2. Ensure consistency in terminology and reduce the use of acronyms to make the study more accessible.

Response: We will carefully go through the manuscript and improve the accessibility through reducing the use of acronyms while including their meanings and checking for consistency.

Please also refer to your previous comment R2.C4.1 and R1.C1.

R2.C4.3 3. Provide a clear justification for the aggregation of groundwater data to a monthly scale, discussing any potential limitations.

Response: We will extend the paragraph describing the data selection. We apologize for any confusion that the description of the data may have caused. We used monthly data as this is the aggregated data provided by CORRECTIV while higher resolved time series were not available. They made this aggregation from the original data with different temporal resolutions, partly daily. At the same time, we think that this temporal resolution is highly suitable to investigate groundwater processes that are temporarily buffered, in particular for event time scales relevant for droughts and their propagation into the groundwater component, e.g. months to years (e.g., Van Loon, 2015). Please, also refer to our response to R1.C2.

R2.C4.4 4. Include comparisons with traditional hydrological models to provide additional validation for the machine learning results.

Response: We will add more references to the discussion setting our results into context of existing studies using either hydrological modelling or data-driven approaches (please, refer to Comment R2.C2.2 and R2.C2.3). One of the challenges to compare the large-sample data-driven results to modelling is the that groundwater models across these large scales are rare and hard to parameterize and base on several assumptions and simplifications of the subsurface. We thus think that the different methods and findings complement across studies, but covering hydrological modelling is beyond the scope of this study.

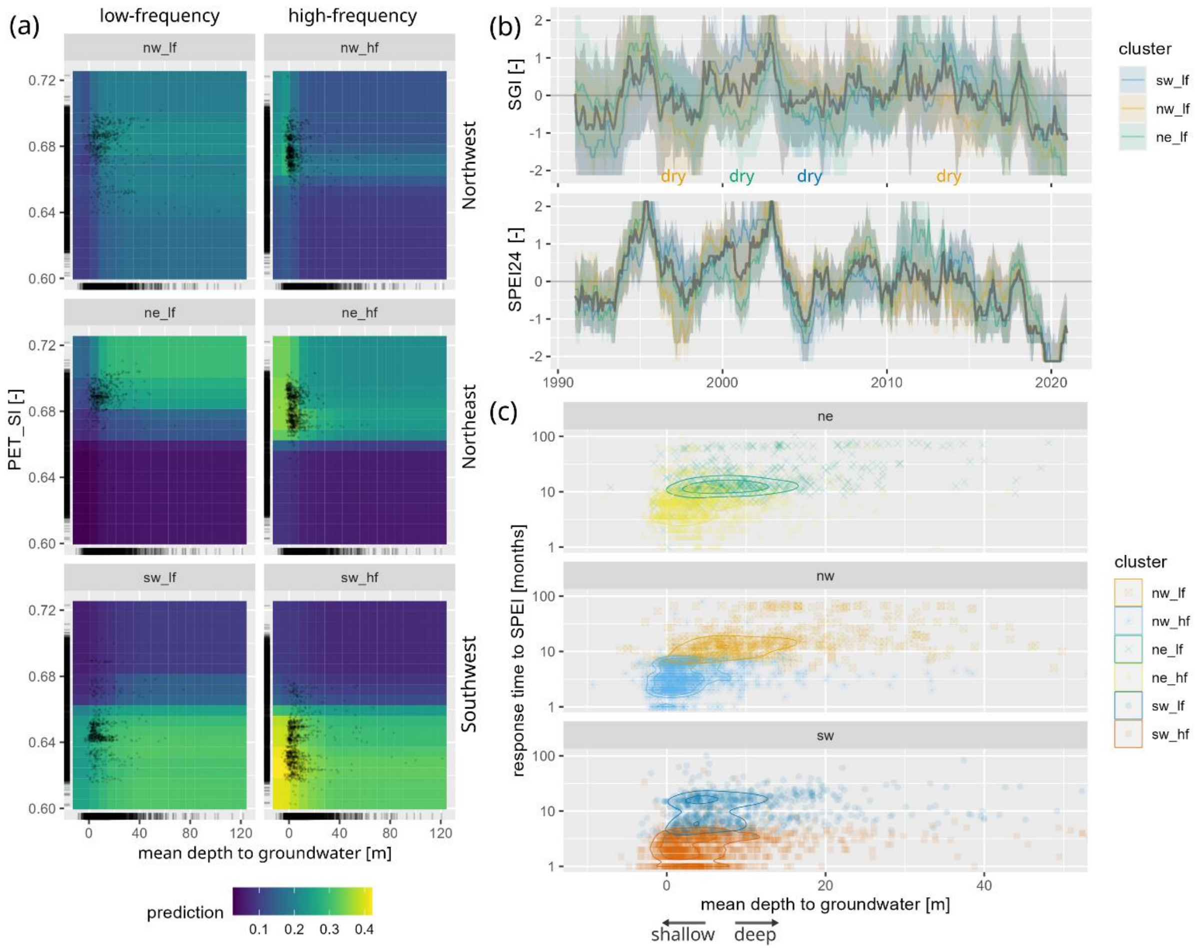
R2.C4.5 5. Clarify the significance of R^2 values and explain their implications for model reliability.

Response: This is a good point, for which we will add reflections on the model performance to the manuscript. The model performances that were reached are in a similar range of a comparable study by Schuler et al. 2022 predicting autocorrelation lengths in groundwater levels in Ireland. Moreover, we think they are not surprisingly low as the predictors used are simple metrics/proxies which are not able to represent the whole complexity of subsurface processes, however, the tendency they show is clear and consistent across the different RF models. This provides confidence in the reliability of model results. Please also refer to the response to R1.C12.

"The similar rankings provide confidence in the robustness of results even if model performances are not high in the regression models with R2<0.5. This range of performance is, however, not surprising given the heterogeneity of subsurface conditions and complexity of processes, which cannot be fully represented by the simple characteristics used as predictors."

R2.C4.6 6.Summarize key findings in tables and figures to facilitate quick interpretation of results for a broader audience.

Response: Thank you for this suggestion, we will add a new Table providing a summary of key results across clusters such as medians and majorities of selected characteristics and the connection between cluster names and characteristics (refer also to comments R1.C11, R2.C4.1, R1.C1.1). In addition, we will modify Figure 4 adding labels to guide the reader:



R2.C5 6. Conclusion and Final Decision

Overall, the study by Ebeling et al. (2024) is a highly valuable contribution to groundwater research, offering a detailed analysis of aquifer responses to drought events across Germany. However, refining the interpretation of results, improving accessibility, and providing stronger contextualization in traditional hydrological frameworks would enhance the study's impact.

Final Decision: Accepted subject to minor revisions.

The study is well-conceived and presents significant scientific advancements. However, minor revisions are required to improve clarity, ensure methodological transparency, and strengthen the discussion by incorporating additional references and comparisons with previous work. Addressing these aspects will enhance the study's accessibility and scientific rigor.

Response: Thank you for the overall positive feedback. We will carefully go through the manuscript to improve clarity, transparency, and accessibility as well as to enhance the comparison to other works and context. Please, refer to the responses to the specific comments.