Author Responses to Referees #1 and #2

Groundwater recharge in Brandenburg is declining – but why?

T. Francke, and M. Heistermann
NHESS Discussions, doi:10.5194/egusphere-2025-222

RC: ***Reviewer Comment***, AR: *Author Response*, ☐ Manuscript text

Dear editor, dear referees,

thanks again for the thoughtful and detailed referee reports, and for the swift editorial decision.

We have revised the manuscript along our replies in the interactive discussion. Following the numerous suggestions and comments made by the referees, we are confident that the comprehensibility of the manuscript has improved, and hope that it can now be considered for publication in NHESS.

Please find below our point-by-point replies to the referees which are closely related to our replies in the interactive discussion.

Thanks again for your feedback, and your support in improving the manuscript.

Kind regards,
Till Francke and Maik Heistermann

Responses to referee #1

RC: ***This is a very interesting paper that analyses the groundwater recharge behavior in 5 catchments in Brandenburg. The paper fits very well within the scope of the special issue.***

AR: Thank you for the positive reception of the manuscript.

RC: ***I think the paper has great potential, but it is somewhat hindered by missing information and lacking methodology descriptions. This makes the paper somewhat difficult to follow, I often had to backtrack quite a lot in the text to see if I missed some detail or information. Therefore I am sending below my recommendations to improve the manuscript.***

AR: We revised the manuscript in order to improve readability and comprehensibility. To that end, the comments of both referees were very helpful.

RC: ***My main question is about the model setup. Are the models set up on a per catchment basis, or over a grid or other spatial sub-units? This was very confusing to me – sometimes I was sure everything is aggregated over catchments, but in some parts further inhomogeneities were resolved. It would be important to include more details about this in the text. Please find my specific comments below.***

AR: A similar comment was made by referee #2. We added a sub-section at the beginning of the methods section in order to give an overview of the modelling approach. Specifically, this sub-section clarifies how the spatial average of the simulated GWR is obtained per catchment, based on the hydrotope concept. For convenience, we reproduce the new subsection 3.1 here:

> The main methodological approach of our study is to simulate series of GWR in each study catchment for the period from 1980 to 2023. From these series, we computed the trends in GWR and compared them to the corresponding trends of observed discharge. A good agreement between both would indicate that the simulation model is able to explain the discharge trends as outlined in Fig. 1 and Tab. 1 (see also section 1 for additional background).
>
> To obtain GWR, the one-dimensional Soil-Water-Atmosphere-Plant model (SWAP, van Dam et al., 2008, section 3.4) was used to simulate the surface water balance and the resulting percolation of water through the unsaturated zone down to the groundwater table. This daily "bottom flux" was aggregated to obtain the annual sum of groundwater recharge.
>
> In order to obtain the spatial average of GWR per catchment, we followed the concept of "hydrotopes", i.e. spatial sub-units that are considered as homogeneous with regard to (i) climate forcing, (ii) soil texture, (iii) land use, and (iv) groundwater depth. For each catchment, climate and soil were assumed to be uniform across the entire catchment (i.e. *one* class each per catchment): the climate forcing was based on the nearest of the four selected climate stations (section 2.3), and soil texture was represented by the dominant soil texture class in the catchment (section 2.6). Land use and groundwater depth, however, were assumed to be heterogeneous across the catchment: land use was represented by *two* classes (forest and grass-/cropland) and groundwater depth by *13* classes (see section 2.6). This resulted into a total of 26 hydrotope classes (1 climate x 1 soil x 2 land uses x 13 groundwater depths). By spatially intersecting all four layers, we quantified the areal fraction of each hydrotope class per catchment. Running the SWAP model for each of the 26 hydrotope classes, the daily GWR per catchment was then obtained as the area-weighted average of the simulated daily bottom flux per hydrotope.
>
> In the following subsections, we further explain the treatment of missing hydro-climatological data (section 3.2), the precipitation correction (section 3.3), the SWAP model and its parameterisation (section 3.4), the specific design of the simulation experiments (section 3.5) and the calculation of trends (section 3.6).

RC: ***L33: These few paragraphs could use some references to support the text.***

AR: We would like to point out that the corresponding paragraph ends with the statement "And, in fact, there is strong evidence for such a decline." This statement essentially sets the scene for the two subsequent paragraphs which introduce two types of evidence including references (first, the decline in groundwater levels, second the decline of discharge). This should be kept in mind because otherwise the above statement would in fact appear to lack support by reference. We emphasized the link to the subsequent paragraphs by adding a colon at the end of the statement instead of a full stop.

Having said that, we agree that the paragraph ll. 33-35 could, in itself, be better supported references. For this purpose, we selected two additional references, so that the entire paragraph became:

> Any long-term shift of Brandenburg's already unfavourable vertical water balance towards lower rates of net groundwater recharge (GWR) would put water resource management in Brandenburg at risk – including the German capital Berlin in its centre (Somogyvári et al. 2024; Pohle et al. 2025; Somogyvári et al. 2025). And, in fact, there is strong evidence for such a decline: first, [...]

RC: ***L36: groundwater***

AR: Was fixed throughout the text.

RC: ***L42: Does the term "combination" here means a gauging station after multiple rivers join together?***

AR: As already pointed out in l. 42 of the preprint, we would like to refer to section 2.2 for further explanation (in order to keep the introduction concise). In section 2.2, we enhanced the original explanation (around ll. 136 ff. of the preprint) in order to improve the comprehensibility (also based on a comment by referee #2). The corresponding sentences in section 2.2 became:

> In three cases, adjacent catchments were aggregated in order to further reduce the effect of uncertain belowground watersheds, and to increase the area over which the water balance is computed. "Aggregating" two (or three) catchments means that their areas were merged to one coherent area, and that the observed runoff at the gauges of each catchment was summed up in order to obtain the total runoff from the merged area. Tab. 1 gives a comprehensive overview of the used discharge gauges and the aggregation of catchment areas and runoff gauge observations for our analysis.

RC: ***L50: can you provide a reference?***

AR: The preprint contained, in line 66, the corresponding reference which establishes that surface runoff is negligible in Brandenburg (LUA Brandenburg, 2001). In order to improve the readability of the manuscript, and avoid the need to jump back and forth while reading, we also added the corresponding reference after the statement in ll. 50-52 of the preprint.

RC: ***L52: what about cross-flow with lower aquifer layers? The recent study of Tsypin et al., 2024 showed that deeper layers interact with the uppermost at certain geological settings, mainly where the Rupelian clay layer is eroded. These flows could act as a water source or sink as they flow upwards or downwards at different locations.***

AR: Thanks for this interesting reference. However, it is difficult for us to assess the potential role of such flows between aquifer layers with regard to long term trends. The authors of the study themselves do not explicitly associate the fluxes between pre-Rupelian and Quaternary aquifers with any long-term trends in groundwater levels. Instead, they hypothesize that these trends "can be explained by (a) changes in water balance; (b) anthropogenic factors; (c) memory effects in the unsaturated zone." In our view, the present manuscript is not a suitable venue to speculate about the relevance of such local cross-fluxes for the water balance of the Quaternary aquifers.

RC: ***Fig. 1: Which wells were used to create this map? Is the data coverage homogeneous, or are there any areas with fewer information?***

AR: As pointed out in the figure caption, the map was adopted from Landesamt für Umwelt Brandenburg (2022). That reference does not provide further information to answer the question. In general, however, there are less gauges in areas with very deep groundwater tables.

RC: ***This figure would be a better place to show the used gauge locations than fig. 5.***

AR: We are not entirely sure what the referee means by ”gauge locations”, but assume that the locations of the climate stations are meant. Still, we are confused because the original Fig. 5 (now Fig. 6) only contains a small inset map to support the interpretation of the figure’s other panels while the locations of the climate stations have already been introduced in Fig. 2. In any case, we would prefer not to introduce additional elements in Fig. 1a because it is only intended to give a spatial overview of groundwater level trends as adopted from Landesamt für Umwelt Brandenburg (2022).

RC: ***L67: ”any long term trend in discharge can thus be interpreted as a long term trend in GWR.” Can you elaborate on this a bit further: what is the long-term criteria, why short term changes cannot be used in such way? Is there a model that can show this behavior/or a past study where this was investigated?***

AR: We are a bit confused by this comment since the criteria (multiple decades) and also the reasoning with regard to the use of long-term trends are elaborated in the subsequent sentence of the preprint:

> Any *long-term* trend in discharge can thus be interpreted as a long-term trend in GWR. The emphasis is on long-term (in the sense of multiple decades) because any short-term dynamics, e.g., across events or seasons, might differ substantially.

In our view, this statement does not require support by modelling results or references, but decided to explain our view a little bit better by modifying and extending the two sentences as follows:

> Any *long-term* trend in discharge can thus be interpreted as a long-term trend in GWR. The emphasis is on long-term, in the sense of multiple decades, since we can assume that over long time periods, any GWR will also end up in the surface water bodies that drain the system. Short-term dynamics of river discharge (e.g., across rainfall events or seasons) are, however, governed by the travel times of water along the different transit paths in the vadose zone and the aquifer.

RC: ***L108: reference to the dataset***

AR: The reference to the dataset, including some background, was provided in section 2.5 in the overall "Data" section. To improve readability, we introduced the reference after ”we use a recently published LAI dataset”, corresponding to l. 107 of the preprint.

RC: ***L125: reference needed***

AR: We added the corresponding reference (LBGR, 2024).

RC: ***L129: Can you support this statement with a reference, or by referring back to the introduction***

AR: There is already a reference to the introduction at the end of the sentence.

RC: ***L136: What does partly mean here?***

AR: Based on this comment and a comment by referee #2, we understand that the explanation should be improved, so we changed the corresponding sentence as follows:

In three cases, adjacent catchments were aggregated in order to further reduce the effect of uncertain belowground watersheds, and to increase the area over which the water balance is computed. "Aggregating" two (or three) catchments means that their areas were merged to one coherent area, and that the observed runoff at the gauges of each catchment was summed up in order to obtain the total runoff from the merged area. Tab. 1 gives a comprehensive overview of the used discharge gauges and the aggregation of catchment areas and runoff gauge observations for our analysis.

RC: ***L143: I am just curious here: what about not interpolated, but dynamically downscaled datasets, such as CERv2 (https://www.tu.berlin/klima/forschung/regionalklimatologie/mitteleuropa/cer) . Would they not be a better choice for gridded data?***

AR: The mentioned dataset is certainly interesting, and the big advantage of such a dataset would be that it is physically consistent in space and time. Furthermore, Bart et al. (2024) in fact demonstrated considerable improvements in comparison to CERv1, particularly with regard to precipitation (with a mean deviation of 0.1 $\mathrm{mm\,d^{-1}}$ from DWD stations). Yet, to our knowledge, it has not yet been systematically assessed how well CERv2 captures the long-term trends in the hydro-climatological variables that are key to the assessment of GWR trends (solar irradiation, temperature, humidity, wind speed, precipitation). Apart from these uncertainties, CERv2 is not suitable for our study because it only starts at 1980 while we require a long and consistent warm-up starting in 1951.

While we consider such a discussion interesting, we think that it will lead too far away from the scope of our study, so we preferred not to go into such depths in the context of the revised manuscript.

RC: ***L161: parenthesis missing***

AR: We removed the superfluous closing parenthesis.

RC: ***Were there any weighting used when multiple gauges were included? I think using a buffer area as a criteria here could easily introduce further uncertainties to the data (as precipitation in Brandenburg could be very heterogeneous, especially during extreme rainfall events) – does including them make a big difference? (does it worth it to use them instead of just the climate station)***

AR: The rain gauge observations were not weighted. The observed series of each selected rain gauge was used individually to force the model. Any weighting would correspond to an interpolation, which we intend to avoid. We are fully aware that, for individual events, single gauges are not able to reproduce the mean areal precipitation of a catchment, specifically in case of convective heavy rainfall during the summer months. Yet, for our analysis, individual precipitation events are not the focus, as we are interested in the long-term behaviour. For us, it was more important to use homogeneous forcing series. And while precipitation can substantially vary in space for individual events, the spatial variability of average (seasonal) precipitation is much lower.

In the revised manuscript, we modified the respective line (l. 82 of the preprint) in section "Design of simulation experiment" to

- ...
- precipitation observed at other precipitation gauges (individually, i.e. without any further weighting) in or close to the study catchments (ref. section 4)
- ...

RC: ***L196: The used dataset is a categorized dataset for groundwater distance. Does its non-continuous nature create any issues for your analysis?***

AR: It is true that the dataset is categorical data. Yet, for the definition of hydrotopes, we would have to create categorical data anyway since each hydrotope is defined by its membership in distinct classes of the relevant variables (climate station, soil texture, land use, depth to groundwater). We think that the definition of categories is reasonable (meter resolution for shallow groundwater, wider class boundaries for deep groundwater). The resulting uncertainties should, on average, not cause any systematic errors: for shallow groundwater tables, the groundwater depth has a strong effect on evapotranspiration (due to the interaction of surface and groundwater); however, for shallow groundwater, the groundwater depth data is sufficiently resolved (1 m increments down to a depth of 5 m). For larger depths, the groundwater depth only affects the transit time of percolating water through the unsaturated zone. Using classes here will not cause a systematic error. In the revised manuscript, we explained this issue in a bit more detail after l. 199 of the preprint.

RC: ***General question for methods: Was the analysis done over a grid, over catchment or any other spatial unit? It is not clear for me from the text. Also, for a better comprehensibility, this section could really use a flowchart on how the different parameters interact, or even a conceptual figure on how the modeling concept looks like.***

AR: As pointed out above, we introduced a new sub-section at the beginning of the methods section in which we outline the overall methodological approach, with specific focus on the spatial model set-up.

RC: ***L214: I really like this concept for filling in missing data. Can you give more details on how this package was used (as it is a specific module for photovoltaic modelling, it is not straightforward how to use it in this setting). Which function was used for modelling with what parameters?***

AR: Beginning in l. 212, we essentially report all relevant information with regard to model training. The fact that the package is geared towards applications in the energy sector is irrelevant as to its application for retrieving clear-sky radiation for any specific date. Clear-sky radiation was modelled for the training location (Potsdam) and for the application locations (other gauges) based on geographic coordinates, altitude, and datetime at 20 minutes intervals. The results were then aggregated to daily values. The random forest model was using 100 trees with a maximum node depth of 10. We added the additional information in the revised version of the manuscript as follows:

> [...] we trained a random forest (100 trees and a maximum node depth of 10) to model solar irradiance from the following predictive features: clear-sky radiation as simulated from geographic coordinates, altitude, and datetime by the Python package pvlib (Holmgren et al., 2018; Anderson et al., 2024; using the function ”get_clearsky” and aggregating from 20 min resolution to a daily sums of radiation), sunshine hours per day, relative and absolute humidity, and seconds elapsed since 1951. The resulting random forest model yielded an R^2 of 98.4 % on independent test data, with clear-sky radiation and sunshine hours being the most important features by far [...]

RC: ***L228: same question: what random forest model was used and how?***

AR: We assume the question refers to L258 of the original manuscript. The random forest model was using 100 trees with a maximum node depth of 10. As predictive features, we used the local daily ERA5-Land wind speed estimates and the wind speed observed at the climate station in Potsdam. The missing information was added to the section.

RC: ***L266: How did you check that this spin up period was enough for the model?***

AR: The required spin-up time corresponds to the time required for any signal (in terms of a flux) to reach the groundwater table for the initial condition being set to a hydraulic equilibrium along the entire profile. The required transit time is longest for forest locations (lowest GWR) and, evidently, for the hydrotopes with the deepest groundwater (up to 70 m).

RC: ***L270: was the modelling run on a grid? – I am really confused about this at this point***

AR: Please see our response above: We added a sub-section at the beginning of the methods section to outline the overall approach.

RC: ***L282: why was this setup needed?***

AR: The background is explained in section 2.4 (ll. 153-156). To make this clear, we had already added a reference to section 2.4 at the end of l. 282 in the preprint.

RC: ***Table 3: this table is a bit complicated and difficult to understand. Please consider revising it.***

AR: We agree that the table is a bit bulky. In our view, though, it is the most straightforward way to document the different model realisations in a transparent way. In order to make the table more comprehensible, we revised the table caption, adding more explanation on how to read the table.

RC: ***L320: maybe: "(small triangles)"***

AR: We changed "triangles" to "small triangles".

RC: ***Figure 4: can you add a legend instead of explaining everything in the captions?***

AR: We deliberately chose to explain the meaning of the symbols in the caption because a corresponding legend would have to been too text-heavy. We would prefer to keep it as is.

RC: ***L318: are the blue/red lines are simply drawn by connecting the points of the corresponding LAI setup?***

AR: Yes, that is correct.

RC: ***L327: This could also be referred to as a linear relation between precipitation and GWR***

AR: We would prefer to not explicitly frame the relationship as "linear" (see ll. 370-71). We understand that l. 327 is a little bit unwieldy, yet we chose the wording deliberately to make it more unambiguous.

RC: ***L338: Can you explain further how the significance was calculated?***

AR: We added the following explanation to section "Trend calculation":

> [...] The significance of the trends is likewise computed according to Sen (1968), and expressed as the p-value.

RC: ***Figure 5: This figure is very complex, but I think the chosen visualizations are really the best way to show the findings. You could make the figure a bit less busy by removing the gauge locations from the map. Also positioning the legend next to the map is not ideal.***

AR: In the revised manuscript, we removed the gauge locations from the map, as suggested. We also switched the positions of map and legend which, in our opinion, further improves the structure of the figure. Please note that the original Fig. 5 is now Fig. 6 because a new figure was inserted as Fig. 5 based on the next comment of the referee.

RC: ***Could you consider adding another figure where (some of – maybe just the main configuration) the modelled GWR timeseries are shown against the discharge timeseries for the catchments? It would help the text in my opinion...***

AR: We added the requested figure to the results section ”Trends in GWR” as Fig. 5, and a brief description of the figure.

RC: ***L351: rephrase ”While this is plausible. . . ”***

AR: We rephrased as follows:

> This is plausible. Still, trends from HYRAS-DE-PR need to be considered with care (as repeatedly mentioned before) because they are based on temporally varying sets of gauges [...]

RC: ***L361: what are the blue and red lines, are they trendlines or just connectors between the points?***

AR: The lines simply are connectors between the grey circles, showing which circles are the result of simulations with or without an LAI trend. They are intended to help the visual differentiation of these two groups, which is otherwise difficult, as the colour of the dots is already used for encoding their significance. We had also tested adding a coloured outline to the dots, but this proved to be much less legible.

RC: ***L365: which offset? Could you be more exact?***

AR: As ”offset”, we refer to the ordinate value at an abscissa value of 0. We added this information in the revised manuscript, and use the term ”intercept” instead of ”offset” to avoid ambiguity.

RC: ***Can one state here that there is an unknown water loss from the system as the discharge trend is steeper?***

AR: This would be one explanation. Uncertainty in the precipitation would be another. Overall, we would like to reserve section 5 for the discussion of such hypotheses.

RC: ***L370: Does it mean more humid conditions or a faster change towards humid conditions?***

AR: Both. The Stepenitz-Löcknitz catchment is an example for a more humid catchment where the slope of the lines is the steepest among all five study catchments. However, for the remaining catchments (most pronounced for the Ucker-Welse), we observe that the lines become a little steeper for higher positive precipitation trends which effectively implies an increase of humidity over time.

RC: ***L373: point out whiskers on the legend***

AR: We would prefer to explain the whiskers in the figure caption as it is difficult to unambiguously distinguish them in the legend from the lines denoting the two simulation groups (with/without LAI trend, see also reply to comment L361 above).

RC: ***L380: where was this pointed out?***

AR: This was pointed out in section 1. In the revised version, we added a corresponding cross-reference.

RC: ***How could the model have GWR estimations for different depths – was it over a grid?***

AR: Please see our response above: we added a brief overview of the modelling concept at the beginning of the methods section, which hopefully clarifies this issue.

RC: ***5 limitation and uncertainties: I really like this section, but it could have a more exact title like “Explaining***

the gap between trends"

AR: While the section has a focus on explaining the gap between observed discharge trends and simulated GWR trends, it is not entirely limited to this issue. Instead, we would like to cover a wider range of potential error sources included in the data and modelling chain and their interpretation. We would hence prefer to keep the section title more generic.

RC: ***L437: a similar behavior was pointed out locally by Somogyvari et al., 2024 for a lake system in the region. They also used a combination of different factors as an explanation for system water loss, together with an environmental tipping point. Did you consider such explanations?***

AR: In our study, we only accounted for the effect of climatic drivers and trends in LAI, while other factors are only touched in the discussion. Still, it is encouraging to see that the results from the two studies are consistent as to the effect of vegetation (in terms of NDVI or LAI). We added a reference to Somogyvári et al. (2024) in the corresponding discussion, corresponding to l. 437 of the preprint, including a brief explanation of the context.

RC: ***L511: Could you rank the different explanations based on plausibility?***

AR: We understand this comment in reference to all sub-sections of section 5. As of now, we would prefer not to provide an overall assessment of relative importance, since governing processes might also vary regionally while other mentioned processes are not yet fully understood. In the different sub-sections, we stated if we considered a certain explanation to be less relevant (e.g., section 5.4 with regard to water consumption by irrigation). However, we have no rigorous way of ranking the explanations.

Responses to referee #2

RC: ***The manuscript by Francke and Heistermann is giving an approach to explain dropping groundwater recharge in Brandenburg. Overall, the study is impressive and very interesting and will doubtless find its place in the journal, where it fits well into. The manuscript is appropriately designed and consistent, figures are of high quality and in principle well prepared. Few minor things should be modified, which I marked below.***

AR: Thank you for the overall positive assessment.

RC: ***More specifically I was wondering, why the authors mainly base their arguments on discharge measurements only rather than including direct groundwater observations to justify dropping groundwater recharge.***

AR: This issue is raised again by the referee below in the comment reg. ll. 69-73 of the preprint. We will address it in response to the below comment.

RC: ***During reading, I also had the problem, that I often had to jump back in the text to be sure I understand what is meant and don't miss things. Sometimes expressions are difficult to read and to follow, but when authors carefully inspect the manuscript again, I'm confident it will be improved and become published in NHESS.***

AR: This impression is consistent with the report by referee #1. In the revised manuscript, we tried to implement the referees' numerous suggestions to improve the comprehensibility of the manuscript. Please see our below responses with respect to specific comments.

RC: ***L30 f. I would rather use deep and shallow instead of distant and close water tables.***

AR: We revised the wording accordingly.

RC: ***L38. In an intensely cultivated and urbanized region as Brandenburg, there might be increased abstraction by man different actors (farmers, water suppliers, private wells, etc.). How comes you generally state: "there is no basis to assume such a widespread abstraction of groundwater"? That statement is a fundamental prerequisite of the coming explanation for dropping gwr, but there is no evidence presented.***

AR: We agree that this statement should be better justified. It is also imprecise in its original form and hence needs improvement.

We agree with the referee that there are abstractions from the groundwater across the state of Brandenburg (although we would not qualify Brandenburg as "intensely cultivated and urbanized" in comparison to other federal states in Germany). However, these abstractions are unlikely to explain the widespread decline of groundwater levels in the recharge areas because there is no evidence that these abstractions were *increasing*. Quite the contrary, water abstractions have been *decreasing* considerably for the past decades - at least since 1991 (see Landesamt für Umwelt Brandenburg, 2022).

We revised the sentence around ll. 37-38 of the preprint to make clear that there is no basis to assume that a *change* in water abstraction led to the observed decline in groundwater levels.

> [...] ground water levels decreased significantly between 1976 and 2020 in the recharge areas (i.e. in the elevated areas with large distances to the ground water table, see Fig. 1). The most plausible hypothesis to explain such a state-wide prevalence of negative trends is a declining GWR, while there is no evidence for increasing groundwater abstractions in the past decades. Quite the contrary, water abstractions considerably decreased, at least since 1991 (Landesamt für Umwelt Brandenburg, 2022) [...]

Furthermore, we would like to point out that we show, in section 5.4, that there is no evidence of significantly increasing irrigation water withdrawals over the past four decades. In section 5.4, we focused on irrigation because it is the only type of water use during which a large portion of the water is actually lost to evapotranspiration, while water from most other water uses actually returns to the local catchments so that these water uses will typically not affect the catchment's long-term water balance.

RC: ***L39. How is it possible, an aquifer gets confined by a river? Please think about wording.***

AR: We agree that the word "confined" already has a different meaning in hydrogeology. In the revised version, we rephrased as follows:

> Please note that the lowlands do not show strong trends in either direction because the groundwater level is stabilised by the water level of the heavily-regulated streams and rivers to which the aquifers are connected.

RC: ***L50: is it indeed possible to generalize it as you do: "Given the highly permeable soils in Brandenburg and the resulting insignificance of surface runoff..."? I think the evidence coming from that statement is very serious and actually builds the foundation of the subsequent argumentation: dropping river discharge as mirror of dropping gwr, because baseflow is (in your argumentation) the only source of surface runoff. I***

think that point, in combination with my concerns from line 38 (missing abstraction), should be treated with more caution and seriously justified.

AR: In the preprint, we introduced, in line 66, the corresponding reference which establishes that surface runoff is negligible in Brandenburg (LUA Brandenburg, 2001). In order to improve the readability of the manuscript, and avoid the need to jump back and forth while reading, we also added the corresponding reference after the statement in ll. 50-52 of the preprint.

RC: ***Fig. 1: a) the names given therein are unclear in their meaning: are these cities, counties, catchments?***

AR: Thank you for pointing this out. In fact, these are names of counties. We added this information in the figure caption.

RC: ***b) in each of the figures, 2 numbers are given, but it is not clear what they represent. Also the unit is not clear to me: mm/a/dec, millimeter per year is okay, but what is "Dec"?***

AR: We clarified this information in the figure caption of the revised manuscript , but also use the opportunity to provide some explanation here: as the referee states, the annual average discharge or the average GWR in a catchment can be expressed in the unit $\mathrm{mm\,a^{-1}}$. The temporal trend of such a quantity, i.e. the change of that quantity over a given time period, would then be expressed by the unit of our quantity ($\mathrm{mm\,a^{-1}}$) in relation to (i.e. divided by) the length of that period. For our study, we chose to represent the trend as the change of the quantity over the period of a decade (10 years). We explained this in the original preprint in section 3.6 (l. 291), but we fully agree that it is necessary to explain this earlier.

In the revised version of the manuscript, we also explain the unit after its first occurrence in the introduction, according to l. 50 of the preprint.

RC: ***L69-73: I partly disagree with that statement. The analyses of gw-timeseries gives worthwhile information as for gwr and though the given issues are valid in a certain way, it is mentioned before, the catchments that are investigated are not influenced by abstraction. Of course, storativity is always an uncertain parameter, but I guess there should be few pumping tests available to at least estimate gwr in few areas, based on gw timeseries.***

AR: The estimation of GWR from GW level records is not only limited by the high uncertainty of storativity. A local change in GW level results from the water balance in a control volume which is not only affected by GWR, but also by the aquifer flow, and it is difficult to distinguish these quantities without further assumptions. But while Collenteur et al. (2020) in fact demonstrated how to reconstruct GWR from GW level series under certain conditions, they also outlined serious limitations in case of deep and shallow groundwater tables. This makes it very difficult to derive *robust* quantitative long-term trends in Brandenburg where we have both, very shallow and very deep groundwater, in each study catchment.

RC: ***L75 ff. Please use consistent wording: shallow and deep gw tables and don't switch between deep and distant.***

AR: We agree. The manuscript was revised accordingly.

RC: ***L107: where is the dataset from? Reference?***

AR: The reference to the dataset, including some background, was provided in section 2.5 in the overall "Data" section. To improve readability, we introduced the reference after "we use a recently published LAI dataset", corresponding to l. 107 of the preprint.

RC: ***L136: I do not understand the integration of neighbouring catchments: they do not contribute to the***

discharge of the gauge. And what does it mean to include them "partly"?

AR: Thanks for the comment. We agree that our explanation was prone to misunderstanding. In order to be more explicit, we changed the sentence as follows:

> In three cases, adjacent catchments were aggregated in order to further reduce the effect of uncertain belowground watersheds, and to increase the area over which the water balance is computed. "Aggregating" two (or three) catchments means that their areas were merged to one coherent area, and that the observed runoff at the gauges of each catchment was summed up in order to obtain the total runoff from the merged area. Tab. 1 gives a comprehensive overview of the used discharge gauges and the aggregation of catchment areas and runoff gauge observations for our analysis.

RC: ***L162. There is a spare parenthesis at the end of the sentence.***

AR: Was removed.

RC: ***L179. Where are the field LAI measurements were taken from? Just out of curiosity: during 2019-2024, forest dieback (as observed in other areas in Germany) did not apply to Brandenburg at all?***

AR: The sources of field measurements are comprehensively documented in section 2.8 of the underlying reference by Cao et al. (2023). As to the question of forest degradation as a consequence of the subsequent drought years between 2019 and 2024 (or, rather, between 2018 and 2022), it should be noted that the number provided in ll. 179 is a long-term trend over four decades which is also an average over all of Brandenburg. The increase of 0.1 $m^2\ m^{-2}$ per decade over almost 40 years does not imply that Brandenburg's forests were unaffected by the impacts of the recent drought years, although the most dramatic consequences, Germany-wide, were observed for spruce forests which are rather uncommon in Brandenburg. It should also be noted that the dataset by Cao et al. (2023) only reaches until 2020. Still, we can observe some decline in the most prominent drought years in 2018 and 2019 (see Fig. 3f of the preprint). Overall, we would prefer not to go into further depth here since this issue is, although relevant as such, beyond the immediate scope of the present study.

RC: ***L196 ff. The steady state type of gw-depth information may not fit to the proposed changes in gw depth due to dropping gwr, am I wrong?***

AR: You are right: our model approach does not account for transient changes in ground water depth. The data that would be required to force the model with a dynamic lower boundary condition is, however, not available in sufficient spatial and temporal detail. A transient lower boundary would also make it difficult to apply our hydrotope concept, since the temporal variation would be another dimension that would need to be accounted for when defining the hydrotopes. However, in section 5.6 of the preprint, we discuss possible uncertainties with regard to the simulation of GWR that may arise from the assumption of stationarity. These are expected to be rather low: In areas with deep groundwater, groundwater depth does not affect GWR (only the transit time in the unsatured zone). In areas with shallow groundwater tables, groundwater depth *does* exert a fundamental control on evapotranspiration and hence GWR. Yet, groundwater tables in areas of shallow groundwater appeared to be quite stable in the past decades (as pointed out before, in section 1) as they are mostly coupled to surface water bodies and stabilised by their hydraulic regulation. So here, the assumption of stationarity is in fact more justified.

RC: ***L246: where is soil moisture data taken from?***

AR: The soil moisture simulated by the model itself is used to compute actual evapotranspiration from potential evapotranspiration, essentially based on the Penman-Monteith equation and the Feddes model, as outlined in

Kroes et al. (2017). In the revised manuscript, we tried to clarify this by changing the sentence in ll. 246 ff. of the preprint as follows:

> Evapotranspiration depends on atmospheric conditions as well as on vegetation and soil properties and is estimated using the Penman-Monteith equation in combination with the Feddes model which accounts for the limitation of transpiration by the availability of soil water (Kroes et al., 2017).

RC: ***L251: is a constant head not contradicting the question, whether gwr is dropping and hence gw tables?***

AR: Please see our above response with regard to the comment in ref. to ll. 196 ff.

RC: ***L265 ff.: is the concept of hydrotopes following the idea of HRUs in hydrology? How did you include the depth to groundwater, 13 classes of gw depth each multiplied by land use, soil type etc.? How did you integrate the hydrotopes spatially: as grid or as a mesh of irregular/regular triangles?***

AR: We added a brief sub-section at the very beginning of the methods section (section 3.1) in which we provide an overview of the methodological approach, including the hydrotope concept and area-weighted averaging.

> The main methodological approach of our study is to simulate series of GWR in each study catchment for the period from 1980 to 2023. From these series, we computed the trends in GWR and compared them to the corresponding trends of observed discharge. A good agreement between both would indicate that the simulation model is able to explain the discharge trends as outlined in Fig. 1 and Tab. 1 (see also section 1 for additional background).
>
> To obtain GWR, the one-dimensional Soil-Water-Atmosphere-Plant model (SWAP, van Dam et al., 2008, section 3.4) was used to simulate the surface water balance and the resulting percolation of water through the unsaturated zone down to the groundwater table. This daily "bottom flux" was aggregated to obtain the annual sum of groundwater recharge.
>
> In order to obtain the spatial average of GWR per catchment, we followed the concept of "hydrotopes", i.e. spatial sub-units that are considered as homogeneous with regard to (i) climate forcing, (ii) soil texture, (iii) land use, and (iv) groundwater depth. For each catchment, climate and soil were assumed to be uniform across the entire catchment (i.e. *one* class each per catchment): the climate forcing was based on the nearest of the four selected climate stations (section 2.3), and soil texture was represented by the dominant soil texture class in the catchment (section 2.6). Land use and groundwater depth, however, were assumed to be heterogeneous across the catchment: land use was represented by *two* classes (forest and grass-/cropland) and groundwater depth by *13* classes (see section 2.6). This resulted into a total of 26 hydrotope classes (1 climate x 1 soil x 2 land uses x 13 groundwater depths). By spatially intersecting all four layers, we quantified the areal fraction of each hydrotope class per catchment. Running the SWAP model for each of the 26 hydrotope classes, the daily GWR per catchment was then obtained as the area-weighted average of the simulated daily bottom flux per hydrotope.

> In the following subsections, we further explain the treatment of missing hydro-climatological data (section 3.2), the precipitation correction (section 3.3), the SWAP model and its parameterisation (section 3.4), the specific design of the simulation experiments (section 3.5) and the calculation of trends (section 3.6).

RC: ***L281: what is the unit "dec" standing for?***

AR: Please see our response to the comment of the referee concerning Fig. 1 b).

RC: ***L291: now the explanation comes, much too late…***

AR: Please see our response to comment of the referee concerning Fig. 1 b).

RC: ***In general, it is hard to understand the concept of mm/a/dec. To become able to translate it into a logical unit, it needs a brief explanation at the very beginning.***

AR: Please see our response to the comment of the referee concerning Fig. 1 b).

RC: ***L296: why was just LAI selected to be the controlling parameter of dropping gwr? Increasing ET is obviously a key-parameter as well controlling the soil moisture (fig. 3d).***

AR: We are not sure how the comment refers to l. 296 of the preprint. However, we will try to respond to the comment itself.

We do not quite agree with the notion that we "selected just LAI to be the controlling parameter of dropping GWR". What we did was to compare GWR trends with and without an external LAI trend, given the observed climate forcing.

GWR and its change are governed by the long-term balance between precipitation and (actual) evapotranspiration. But while precipitation is an external forcing, *actual* ET is inferred internally. Although it is driven externally by irradiation, temperature, vapour pressure, and wind speed (which all affect *potential* ET), actual ET is the result of various interlinked internal processes (e.g., root water uptake, bare soil evaporation, percolation below the root zone) and internal state variables, the most important of which is soil water content, representing the actual availability of water for transpiration and evaporation (in combination with the soil hydraulic properties which, in essence, control the matric potential for a given water content). We have clarified in our response to the referee's above comment how the Feddes model is used to account for the effect of soil water content on transpiration.

As a consequence, soil water content cannot be used as a "control" parameter in the sense of an external forcing. In turn, potential evapotranspiration (referee's reference to Fig. 3d) and its trend is the result of a variety of forcing variables (as outlined above), so it is not helpful to consider it as a control variable, either, in the sense that it could be varied (controlled) independently.

Surely, LAI could also be considered as an internal state variable, in the case that vegetation dynamics were computed internally as a function of atmospheric forcing and internal limitations by water and nutrients. However, this is deliberately not how we set up the model, since this introduces a series of other uncertainties.

RC: ***L297: please explain CI***

AR: We apologize, but we cannot relate this comment to the content of l. 297 in which the term "CI" does not occur in the preprint. The first occurrence of the abbreviation "CI" is in l. 323 where the meaning (confidence interval) is also explained.

RC: ***L344: where do I see it in Fig. 5? The figure is very complex but nice. However, I have difficulties to differentiate between various shades of grey in some of the graphs.***

AR: We understand this difficulty. We had tried various ways to differentiate levels of significance for the simulated GWR without introducing too many different colours (which again causes problems in interpretability). Still, we think that three different levels of grey (one actually black, one intermediate, one very light) are an acceptable compromise. One might have to look twice to be sure, but this is still our preferred solution. So regarding the question with regard to l. 344: "where do I see [significance levels of $p \leq 0.1$] in Fig. 5?', the answer is: everywhere where the dots do *not* have the lightest shade of grey. We added an explanation after the corresponding sentence in the revised manuscript in order to support the interpretation of the figure. Please note that the original Fig. 5 is Fig. 6 in the revised manuscript.

RC: ***L362: in fig. 3 there is no real trend in precipitation observable, how comes it though brings the gwr to drop?***

AR: Fig. 3e is maybe not ideal to appreciate the long term precipitation trend because the strong interannual variability of precipitation requires the y-axis scaling to cover quite a wide range, making it difficult to read the precipitation trends of the four selected climate stations directly from the trend lines. This is why we highlight the trends at the four selected climate stations also in Figs. 4 (large dots) and 5 (large triangles, now Fig. 6), showing that the downward trend in precipitation is substantial (though not statistically significant) at the stations in Lindenberg and Marnitz, and, to a lesser extent, in Potsdam, too.

RC: ***Fig. 6: how is it explainable, gwr is almost negligible in 4 out of 5 catchments fro shallow water tables? Contrastingly, the deeper groundwater receives reasonable gwr of 100-150 mm/a.***

AR: The reason behind this is explained in ll. 380-386 of the preprint. In that paragraph, we now explicitly refer to Fig. 7a (formerly Fig. 6a) in order to improve the comprehensibility. At the beginning of the paragraph, we also added another reference to section 1 (where the effect of shallow groundwater tables on net GWR is explained).

RC: ***What does "area weighted" mean here? How is areal fraction calculated?***

AR: We assume that this comment specifically refers to the caption of Fig. 7 (Fig. 6 in the preprint). Let use explain this more explicitly: for each study catchment of area A (see Tab. 1), we define $A_{\leq 3}$ as the total area of hydrotopes with a groundwater depth ≤ 3 m, and $A_{>3}$ as the total area of hydrotopes with a groundwater depth >3 m. The term "areal fraction" (Fig. 7c) hence refers to the fractions $\frac{A_{\leq 3}}{A}$ and $\frac{A_{>3}}{A}$, respectively, summing up to a total value of 1. These fractions also constitute the weights for an area-weighted average of the GWR itself (Fig. 7a) and the area-weighted average of the *trend* of GWR within each of the two groundwater depth classes.

RC: ***L383 ff. I have difficulties to understand why the contribution of the shallow gw is that high, if their (however calculated) areal contribution in the total catchment is low. And why show shallow gw at all different gwr compared to deeper gw bodies? Is it a function of delay? But how was it modelled in 1D?***

AR: Areas with shallow groundwater tables are characterised by very high evapotranspiration rates because of the connectivity of the vegetation's root system to the groundwater table. The vegetation is able to draw water from the groundwater even if the groundwater table is somewhat below the maximum rooting depth. One reason behind this is capillary rise. The second reason is the inversion of the hydraulic gradient under dry summer conditions, leading to an upward water flux that can, within limits, be sustained in case the groundwater table is sufficiently shallow. Both processes are represented by the model. The resulting high values of summerly evapotranspiration lead to very low net fluxes towards the groundwater table, if integrated over the entire year. In total, this phenomenon has nothing to do with transit times, but with the connectivity between plants and groundwater and its effect on net water fluxes.

In the revised manuscript, we provide more explanation in the introductory part (ll. 74-76 of the preprint) where this issue is addressed first, and refer to the introduction in the line that corresponds to l. 380 of the preprint. We also revised the former ll. 382-386 to explain more clearly why the contribution of areas with shallow groundwater tables to the overall GWR trend of a catchment is so high, even though their areal fraction is rather low.

RC: ***L403: there is a "by" too much.***

AR: Was removed.

RC: ***L453f. Where stems the data from? A study from 2009 might not be representative to activities during the last decade.***

AR: There is no evidence of a substantial extension of irrigated areas since 2009. The most recent figure, as provided by the statistical agency, is 309 km² in the year 2022 (Amt für Statistik Berlin-Brandenburg, 2024). This is close to what we referred to as "roughly between 200 and 300 km²" (l. 453 of the preprint) and cannot help to explain any substantial portion of the overall trend. We revised the manuscript accordingly in order to include the information and the additional reference.

References

Amt für Statistik Berlin-Brandenburg: Pressemitteilung Nr. 71, https://www.statistik-berlin-brandenburg.de/071-2024, last accessed: 29 April 2025, 2024.

Bart, F., Schmidt, B., Wang, X., Holtmann, A., Meier, F., Otto, M., Scherer, D.: The Central Europe Refined analysis version 2 (CER v2): evaluating three decades of high-resolution precipitation data for the Berlin-Brandenburg metropolitan region, Met. Z., 33(5), 339-363, 2025.

Cao, S., Li, M., Zhu, Z., Wang, Z., Zha, J., Zhao, W., Duanmu, Z., Chen, J., Zheng, Y., Chen, Y., Myneni, R. B., and Piao, S.: Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020, Earth System Science Data, 15, 4877–4899, https://doi.org/10.5194/essd-15-4877-2023, 2023.

Collenteur, R. A., Bakker, M., Klammler, G., and Birk, S.: Estimation of groundwater recharge from groundwater levels using nonlinear transfer function noise models and comparison to lysimeter data, Hydrol. Earth Syst. Sci., 25, 2931–2949, https://doi.org/10.5194/hess-25-2931-2021, 2021.

Kroes, J., van Dam, J., Bartholomeus, R., Groenendijk, P., Heinen, M., Hendriks, R., Mulder, H., Supit, I., and van Walsum, P.: SWAP version 4 - Theory description and user manual, Wageningen Environmental Research Report 2780, https://swap.wur.nl/Documents/Kroes_etal_2017_SWAP_version_4_ESG_Report_2780.pdf, last accessed: 6 January 2025, 2017.

Landesamt für Umwelt Brandenburg: Wasserversorgungsplanung Brandenburg - Sachlicher Teilabschnitt mengenmäßige Grundwasserbewirtschaftung, https://lfu.brandenburg.de/sixcms/media.php/9/Wasserversorgungsplan_barrierefrei.pdf, 2022.

LUA Brandenburg: Flächendeckende Modellierung von Wasserhaushaltsgrößen für das Land Brandenburg - Studien und Tagungsberichte, Band 27 | Startseite | LfU, Tech. Rep. Band 27, Landesumweltamt Brandenburg, https://lfu.brandenburg.de/lfu/de/ueber-uns/veroeffentlichungen/detail/~15-10-2001-flaechendeckende-modellierung-von-wasserhaushaltsgroessen-fuer-das-land-brandenburg-stu#, 2001.

LBGR: Landesamt für Bergbau, Geologie und Rohstoffe Brandenburg, Bodenübersichtskarte des Landes Brandenburg im Maßstab 1:300,000, `https://geo.brandenburg.de/?page=LBGR-Webservices`, last accessed: 30 April 2024, 2024.

LUA Brandenburg: Flächendeckende Modellierung von Wasserhaushaltsgrößen für das Land Brandenburg - Studien und Tagungsberichte, Band 27 | Startseite | LfU, Tech. Rep. Band 27, Landesumweltamt Brandenburg, `https://lfu.brandenburg.de/lfu/de/ueber-uns/veroeffentlichungen/detail/~15-10-2001-flaechendeckende-modellierung-von-wasserhaushaltsgroessen-fuer-das-land-brandenburg-stu#`, 2001.

Pohle, I., Zeilfelder, S., Birner, J., and Creutzfeldt, B.: The 2018–2023 drought in Berlin: impacts and analysis of the perspective of water resources management, Nat. Hazards Earth Syst. Sci., 25, 1293–1313, https://doi.org/10.5194/nhess-25-1293-2025, 2025.

Somogyvári, M., Scherer, D., Bart, F., Fehrenbach, U., Okujeni, A., and Krueger, T.: A hybrid data-driven approach to analyze the drivers of lake level dynamics, Hydrol. Earth Syst. Sci., 28, 4331–4348, https://doi.org/10.5194/hess-28-4331-2024, 2024.

Somogyvári, M., Brill, F., Tsypin, M., Rihm, L., and Krueger, T.: Regional-scale groundwater analysis with dimensionality reduction, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-4031, 2025.

Tsypin, M., et al.: Modeling the influence of climate on groundwater flow and heat regime in Brandenburg (Germany)." Frontiers in Water 6 (2024): 1353394, 2024.