I enjoyed reading the manuscript "Influence of groundwater recharge projections on climate-driven subsurface warming: insights from numerical modeling." The manuscript presents a clear objective, employs a sound methodology to achieve it, and reports results that are well presented and supported by the discussion and conclusions. Most of my comments focus on improving clarity in certain sections.

We would like to thank reviewer#2 for their positive feedback on our study, as well as for the in-depth review and constructive suggestions. In what follows, we provide our point-by-point responses to each comment (our answers are marked in red).

Line 167: Could you please explain why these seven GCMs were selected? Since many GCMs are available, I am curious whether these models have specific characteristics that make them particularly suitable for this study. Clarifying this would also address the question of why seven—why not five, ten, or the full ensemble?

Thank you for this helpful comment. The selection of the seven GCMs in our study follows the same strategy used in Nguyen et al. (2024), where models were chosen based on a combination of performance- and independence-based criteria. In that work, GCMs were evaluated using the ClimWIP (Climate model Weighting by Independence and Performance) method (Brunner et al., 2020) as implemented in ESMValTool v2.6.0 (Eyring et al., 2016). The selection relied on quantitative metrics—specifically each model's distance to ERA5 over Europe for 1985–2014 temperature and sea-level pressure climatology, annual variability, and temperature trends—as well as qualitative considerations of model independence and spread following the recommendations of Merrifield et al. (2023). This approach identifies a subset of CMIP6 models that provides a representative range of climate responses while avoiding unnecessary inter-dependencies between closely related models and reducing the computational load for both the weather generator and subsequent impact modeling.

We will clarify this briefly in the updated manuscript.

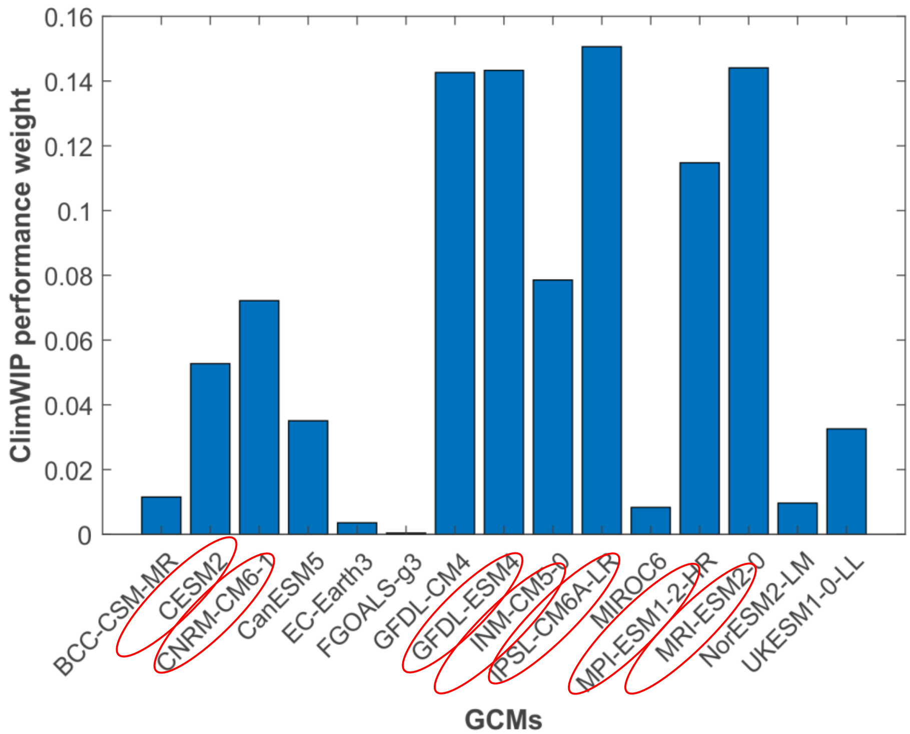

Figure 1: Performance weights for 15 GCMs resulting from the ClimWIP procedure based on the preselected evaluation criteria for the historical period (modified after Nguyen et al., 2024). 7 GCMs selected for the current study are marked in red.

For completeness, could you report the area of the model domain? The mesh spans 170 × 150 km, but (as I understand it) the numerical model only simulates part of that rectangle. Are some cells within the bounding rectangle not active? A brief clarification would help.

The FEM mesh was not rectangular. The provided dimensions are average extents in X and Y direction. The actual shape of the FEM model is irregular, as shown in Figures 1b and 1c. The model boundary follows topographic elements (major rivers and divides), largely corresponding to the boundaries of the Brandenburg federal land. None of the cells were deactivated. We will specify the area of the model domain in the corrected manuscript (27.6 thousand km^2).

It would be helpful to include a short paragraph or a few sentences describing the computational cost of the simulations and the computing resources used (e.g., HPC system, number of cores, total runtime).

This suggestion was also given by reviewer#1. Below is the detailed information about the simulations. We will include a concise description in the updated manuscript.

Computational experiments for the climate and hydrological components—including the non-stationary weather generator, the mHM hydrological model, and the workflow connecting both—were performed on the GLIC high-performance computing (HPC) system at the GFZ German Research Centre for Geosciences. The system uses the SLURM workload manager, a standard HPC scheduling environment that assigns resources and coordinates parallel workloads. This setup allowed us to run the modelling chain efficiently across 50 compute nodes in parallel. Each node provided 40 GB of RAM, and each task used four CPU cores. With this configuration, the complete workflow—from the nsRWG simulations to the mHM impact modelling—required roughly 10 days of wall-clock time.

The groundwater model was simulated separately on a local workstation equipped with an Intel Core Ultra 7 155H processor (16 cores) and 32 GB of RAM. Each of the six scenarios required ~3,500 computational timesteps to cover a 145-year period with monthly boundary conditions, resulting in simulation times between 13 and 26 hours.

Line 219: Why were thermal and hydraulic parameters calibrated for only two layers? A short justification would improve clarity.

We agree that the revised manuscript would benefit from such justification. In general, there are two reasons:

1. The upper two units (Quaternary and post-Rupelian Tertiary) are the most relevant for the study objective, since they directly experience the impact of the varying upper boundary conditions. The influence on the deeper aquifers is local and depends mostly on the thickness distribution of the Rupelian aquitard, which is known fairly well and doesn't require addition calibration. Moreover, above the Rupelian aquitard advection dominates the heat transport, whereas below, heat conduction plays the main role. Therefore, the calibration was performed for the parameters of the advective diffusive domain, which matter more for the time scale of the model.
2. The necessary information about the deeper units is sparser. Selected constant effective rock properties for the deeper units are consistent with published parametrized models of the same area (Noack et al., 2013; Frick et al., 2019). In these papers, sensitivity studies on some key parameters have been already tackled, e.g., effective permeability of stratigraphic units.

Line 227: What proportion (in percentage) of the model's lateral boundaries coincide with rivers?

The proportion of the river boundaries is approximately 50% (basically eastern and northwestern edges corresponding to Oder and Elbe rivers respectively). In the revised manuscript we will replace the phrase "At lateral model edges a constant-head Type I (Dirichlet) boundary condition (BC) was assigned, given that they largely correspond to major rivers" with are more quantitative statement.

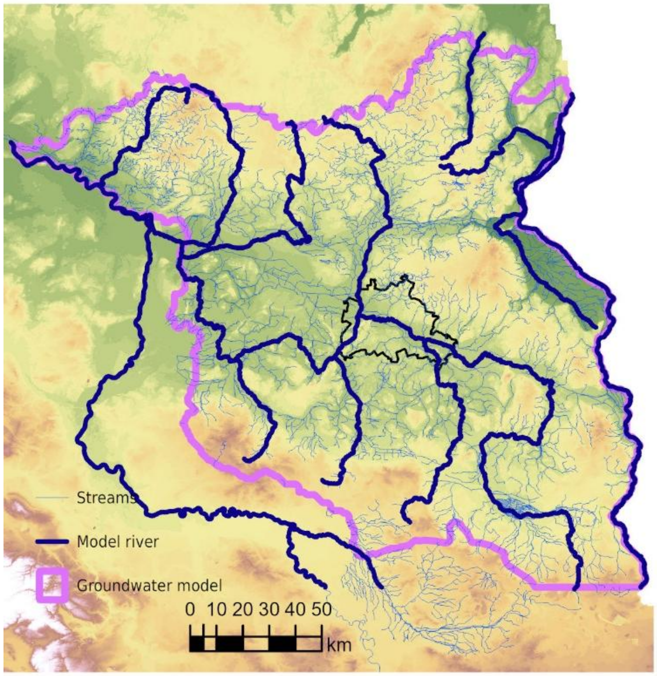

Figure 2: River network of the study area (LfU, 2023) and the model boundaries.

Some darker points appear in Figure 5a. Do these points represent a specific subset of observations, or are they a rendering artifact? Additionally, could you add a measure of bias between modeled and observed heads and temperatures?

Yes, this is an artefact of the overlapping points. The bias for heads was 3.70, and for temperature it was -0.42. We thank reviewer#2 for these suggestions. The new version of the plot reflects these changes (Figure 3).

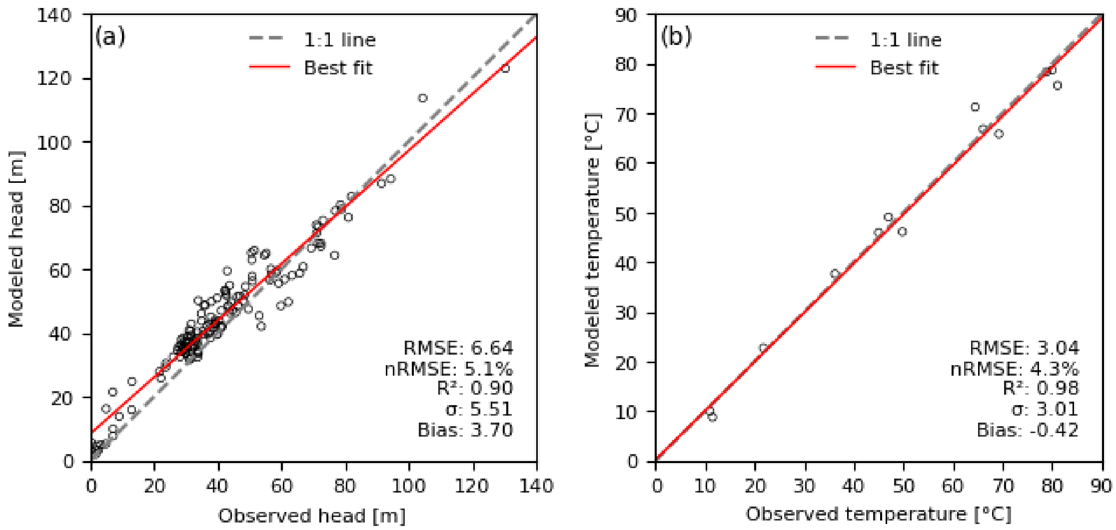

Figure 3: Steady-state groundwater model calibration results: simulated versus observed hydraulic head and temperature at monitoring points (locations shown in Figure 1c). RMSE – Root Mean Square Error, nRMSE – normalized to value range; R2 – coefficient of determination; σ – standard deviation.

In Figure 6 why were these three wells chosen? Would it not be more informative to select wells from three distinct regions with different expected drawdown responses (Fig. 7a)?

These three wells were selected to illustrate typical pathways of GWL evolution for wells with different initial water table depths: < 5 m, 5-20 m, and >20 m. The proposal of reviewer#2 is also valid. However, we have two limitations: (1) the regions in Figure 7 do not overlap between scenarios; (2) there are no suitable wells from the regions with the highest drawdown range with adequately long coverage of the historic record. Therefore, we propose to keep the wells presented in the original manuscript.

Line 386: The text states that Profile A has a shallow seasonal envelope entirely above the water table, implying no advective transfer of the seasonal signal into the saturated zone. Isn't this also the case for Profile D?

Yes, it is also the case for the Profile D, although the main reason for including Profile D was to demonstrate the thermo-hydraulic effect of eroded Rupelian in the text and later in Figure 9. The text will be adjusted accordingly.

Section 6.1 is clearly written, and Figure 13 is well designed. However, I'm a bit confused: GD3 is described as an area where "the Rupelian is locally eroded, allowing direct hydraulic connection between Quaternary aquifers and deeper formations". In Figure 13, the region labeled GD3 appears to still contain Rupelian clay (gray unit), which does not reflect such a direct connection. Could you clarify this?

We thank reviewer#2 for pointing on this inconsistency. GD2 and GD3 labels must be swapped in the schematic (Figure 4) to be consistent with the description in the text.

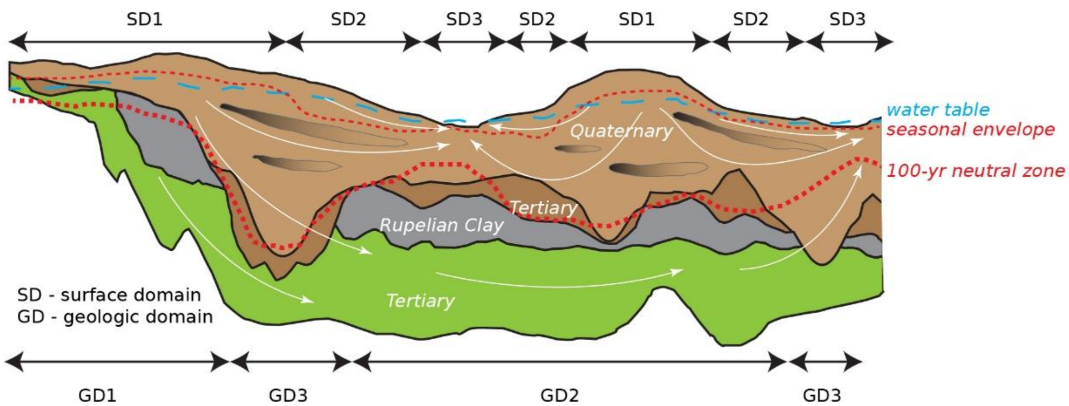

Figure 4: Schematic cross-section, illustrating patterns of groundwater flow and the thermal field for different surface and geological domains.

Line 517: The sentence "The groundwater system is … reversible" may cause confusion. It is clear that your model does not simulate mechanical deformation of the subsurface (and this is okay), yet changes in groundwater storage can lead to changes in hydraulic properties, which may be irreversible depending on geomechanical conditions (Galloway & Burbey, 2011, not my paper, just a reference). I recommend clarifying the assumptions under which the groundwater system is considered reversible in your simulations or rephrasing the sentence to avoid misunderstanding.

We agree with reviewer#2 comment. In our original comment, we were referring to projected head changes without considering compaction effects and associated storativity reductions. In Galloway and Burbey (2011), the driver of regional land subsidence is aquifer overexploitation. Such a process in the study area could potentially be triggered due to drainage of peatlands, dewatering of open-pit mines, and extensive pumping in the urban area (Wolkersdorfer and Thiem, 1999; Landgraf, 2022).

That being said, we are not aware of any studies of climate change directly contributing to the subsidence due to reduction of groundwater recharge. We propose to revise the statement as follows:

Our results also indicate that the tested magnitude of projected recharge changes (-10 to 20 %) only influences groundwater dynamics (e.g., hydraulic gradients and flux) to a limited extent, as compared with stronger changes due to pumping. More fundamental changes in basin-scale flow due to climatic variability *would require sustained forcing beyond the 2100 horizon and evaluation through thermo-hydraulic-mechanical simulations.*

References:

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth Syst. Dynam., 11, 995-1012, 10.5194/esd-11-995-2020, 2020.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, 10.5194/gmd-9-1937-2016, 2016.

Frick, M., Scheck-Wenderoth, M., Schneider, M., and Cacace, M.: Surface to Groundwater Interactions beneath the City of Berlin: Results from 3D Models, Geofluids, 2019, 4129016, https://doi.org/10.1155/2019/4129016, 2019.

Galloway, D. L. and Burbey, T. J.: Review: Regional land subsidence accompanying groundwater extraction, Hydrogeology Journal, 19, 1459-1486, 10.1007/s10040-011-0775-5, 2011.

Landgraf, L.: Das Moorschutzfachkonzept Brandenburgs–wie gelingt der Klimaschutz auf Moorböden in der Praxis?, TELMA-Berichte der Deutschen Gesellschaft für Moor-und Torfkunde, 52, 129-154, 2022.

LfU: Landesamt für Umwelt Brandenburg - Auskunftsplattform Wasser (APW), https://apw.brandenburg.de/, last access: 01/10/2024.

Merrifield, A. L., Brunner, L., Lorenz, R., Humphrey, V., and Knutti, R.: Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications, Geosci. Model Dev., 16, 4715-4747, 10.5194/gmd-16-4715-2023, 2023.

Nguyen, V. D., Vorogushyn, S., Nissen, K., Brunner, L., and Merz, B.: A non-stationary climate-informed weather generator for assessing future flood risks, Adv. Stat. Clim. Meteorol. Oceanogr., 10, 195-216, https://doi.org/10.5194/ascmo-10-195-2024, 2024.

Noack, V., Scheck-Wenderoth, M., Cacace, M., and Schneider, M.: Influence of fluid flow on the regional thermal field: results from 3D numerical modelling for the area of Brandenburg (North German Basin), Environmental earth sciences, 70, 3523-3544, https://doi.org/10.1007/s12665-013-2438-4, 2013.

Wolkersdorfer, C. and Thiem, G.: Ground water withdrawal and land subsidence in northeastern Saxony (Germany), Mine Water and the Environment, 18, 81-92, 1999.