Authors' Response to Reviews of

Changes in groundwater-surface water interactions following two centuries of irrigation practices and groundwater use in the Upper Ganges-Yamuna interfluve, North India.

Original comment
AR: Authors' response (blue italic)

Referee #2

General Comments
This paper presents a novel and valuable contribution to understanding long-term groundwater dynamics in the Indo-Gangetic Basin. The two-century hindcast perspective is ambitious and the integration of multiple recharge and abstraction components addresses an important gap in the literature. After moderate revisions, I believe the paper will be suitable for publication.

RC2-GC1
My overarching concern is that the modelling methodology requires clearer description of what the model is actually predicting versus what is imposed through the forcing assumptions. In addition, further discussion on the challenges in constraining hindcast uncertainty (see Forstner et al. 2025) using the Monte Carlo framework would be valuable. Several clarifications would substantially strengthen the paper.
AR: We thank the reviewer for this insightful comment. We agree that the manuscript should more clearly distinguish between externally imposed forcing assumptions and simulated model responses, and we will revise the text accordingly. In our manuscript, recharge and abstraction are prescribed as uncertain external forcings, while groundwater heads and SW-GW exchanges are simulated responses of the groundwater model.

On the model conditioning and uncertainty framework
RC2-GC2
The section titled "Monte Carlo analysis and validation" describes an approach where 1,500 simulations are run and those meeting performance and realism criteria are retained. Could the authors clarify how this approach relates to formal data assimilation or history matching frameworks? In particular, Figure 3 shows that several posterior parameter distributions closely resemble their priors — could the authors discuss what this implies for the information content of the observations, and whether the reported uncertainty bounds reflect a genuinely updated posterior or primarily the prior parameter ranges? Related to this, could the authors reconsider the use of the term "validation" for this process, and "model validity" (line 197)?
AR: The applied methodology is not equivalent to formal data assimilation or Bayesian history matching. The approach is more appropriately described as a Monte Carlo uncertainty analysis combined with behavioural model screening (rejection sampling), in which parameter sets sampled from prior distributions are filtered using observational performance criteria and physically based realism constraints. The retained ensemble therefore represents plausible behavioural model realizations rather than a formally derived posterior distribution.
We agree that the similarity between prior and retained parameter distributions indicates that the available observations only weakly constrain some model parameters. This likely reflects a

combination of limited temporal and spatial groundwater-level observations and that several different parameter combinations can explain the observed groundwater levels equally well. Consequently, uncertainty bounds remain substantially influenced by the prior assumptions and should not be interpreted as fully constrained posterior uncertainty estimates.
In addition, we will revise terminology throughout the manuscript and replace terms such as "validation" and "model validity" with more appropriate terms such as "model performance evaluation" and "history matching".

RC2-GC3
As mentioned by the previous reviewer, is there additional observation data which could have been used to condition the posterior ensemble? For example, a key prediction is groundwater level trends, is there any available data on groundwater levels trends which could have informed the model 'conditioning' rather than only using absolute groundwater heads?
AR: We thank the reviewer for this remark. In the revised manuscript, we will include a comparison between observed and simulated groundwater level trends, alongside the comparison of absolute groundwater heads. Specifically, we calculated the Sen slope for each observation well using the measured heads and compared these observed trend with the simulated groundwater level trend over the corresponding period. We will present this in a new appendix figure showing observed versus simulated Sen slopes for all observation wells.

On goodness-of-fit and trend reproduction
RC2-GC4
The acceptance criteria of MAE ≤ 7.5 m and bias between ±5 m are noted. Could the authors discuss how confidently the reported ~2 m groundwater table decline (line 335) can be attributed to modelled processes given these thresholds? Furthermore, for a hindcasting study, could the authors demonstrate that accepted model runs reproduce observed groundwater level *trends* rather than just absolute head levels? Time series comparisons of observed versus modelled groundwater levels would considerably strengthen confidence in the results.
AR: Many thanks for this important comment. Given the applied thresholds (MAE ≤ 7.5 m and bias within ±5 m), the simulated groundwater level changes should not be interpreted as precise deterministic estimates at the meter scale. Rather, the results provide a robust indication of the overall magnitude and spatial and temporal patterns of long-term groundwater table change. Following this comment, and as also suggested by reviewer #1, we will add additional comparisons between observed and simulated groundwater level time series in the appendix. Furthermore, as already indicated in the previous comment, we now evaluate groundwater level trends explicitly by comparing observed and simulated Sen slopes for each observation well over the corresponding monitoring period.

On model structure and predictive insight
RC2-GC5
The authors state at line 142 that the aim is to "capture overall temporal trends in a spatially distributed manner rather than precise estimates for each cell and time step." Given the relatively simple model structure (single layer, homogeneous aquifer) and that recharge and abstraction forcings dominate the output signal, could the authors more explicitly demonstrate what predictive insight the model contributes beyond what could be inferred directly from the forcing trends alone? This is particularly relevant to the SW-GW exchange conclusions, where model structural uncertainty may be as important as the parameter uncertainty quantified through the Monte Carlo approach.

AR: We agree that the added value of the model beyond the imposed forcing trends should be clarified more explicitly. While recharge and abstraction histories are prescribed externally, the model simulates how these forcings propagate through the groundwater system depending on aquifer properties, boundary conditions, and SW–GW interactions. The model provides additional insights into: (i) the spatial redistribution of groundwater-level changes, (ii) time-lagged system responses, and (iii) changes in SW–GW exchange fluxes that emerge from the interaction between hydraulic gradients and surface water boundaries rather than being directly imposed.
We also agree that structural uncertainty is an important limitation. The simplified single-layer homogeneous model was adopted to enable long-term regional-scale hindcasting under limited historical data availability, but this simplification may affect the representation of local heterogeneity and SW–GW exchange processes. We will clarify in the discussion that the presented uncertainty ranges mainly reflect parameter and forcing uncertainty within the adopted model structure, while structural uncertainty is not fully quantified.

Specific Comments
RC2-SC1
I would encourage the authors to use more precise terminology throughout — for example, referring to the accepted runs as a "posterior ensemble" and the filtering process as "rejection sampling" or "informal history matching" rather than "validation." The term "model validity" (line 197) should also be reconsidered. Similarly, “realizations” rather than “scenarios” (see Anderson et al, 2015) for multiple references to more widely used terminology in uncertainty quantification methods.
AR: Thank you for the good suggestions. We agree that the proposed terminology more accurately reflects the applied methodology. We will use the suggested terminology in the revised manuscript.

RC2-SC2
Line 70 (Aim): Could the authors clarify what the model is actually predicting — specifically, that groundwater table depth and SW-GW exchange fluxes are the primary model outputs, while the recharge and abstraction components are forcing terms whose uncertainty is propagated through the ensemble?
AR: We agree that aim of the manuscript should more clearly distinguished between externally imposed forcing assumptions and simulated model responses. In our manuscript, recharge and abstraction are prescribed as uncertain external forcings, while groundwater heads and SW-GW exchanges are simulated responses of the groundwater model. We, therefore, will revise the aim in the introduction for clarity into this:

“This study aims to quantify the spatio-temporal evolution of groundwater recharge and abstraction components, and their effects on groundwater table and GW–SW interactions in the Upper Ganges–Yamuna interfluve of the IGB. First, the spatio-temporal evolution of groundwater recharge and abstraction components is reconstructed for the period from pre-canal conditions (< 1830) to recent times (2016). These recharge and abstraction estimates are subsequently used as uncertain external forcings in a spatially explicit, physically based groundwater model (MODFLOW 6). The model simulates groundwater heads, groundwater storage evolution, and SW–GW exchange fluxes. This analysis improves understanding of the contribution of different water-use sectors to groundwater dynamics and provides a basis for sustainable groundwater management.”

RC2-SC3

Line 109 ("multiple schematizations"): This phrasing is ambiguous. Please clarify whether this refers to multiple conceptual models or model realizations used in the uncertainty quantification.

AR: To clarify it, we changed it to "multiple model realisations with different parameter values for….".

RC2-SC4

Line 116 (yearly timesteps): Given that seasonal water table fluctuations in monsoon-dominated systems can exceed the magnitude of the long-term trends being studied, could the authors provide further justification for the annual timestep and discuss how this may affect the estimated timing and magnitude of the exfiltration-to-infiltration transition in the Hindon River?

AR: (This response addresses comments RC1-GC4, RC2-SC4 and CC2, which concern the same issue). We agree that the use of yearly time steps limits the representation of seasonal groundwater dynamics in this monsoon-dominated system. With a finer temporal resolution (e.g. monthly timesteps), the model could capture seasonal reversals in SW–GW exchange and distinguish, for example, between dry-season infiltration and monsoon-driven exfiltration. However, the objective of this study was to assess long-term regional trends over a two-century period, for which yearly net SW–GW exchange was considered the most relevant metric. In particular, for evaluating the long-term spreading of polluted surface water into the groundwater system, the annual net exchange is more important than seasonal fluctuations. For example, if polluted river water infiltrates into the aquifer during the dry season but is subsequently discharged back to the river during the monsoon, the net annual exchange remains exfiltrating and the long-term transfer of pollutants to groundwater would be limited.

We acknowledge that a higher temporal resolution could provide additional insight into seasonal dynamics and the timing of exchange reversals, but such simulations were beyond the scope of the present long-term regional analysis and would substantially increase computational demands. Furthermore, additional model complexity associated with a finer temporal resolution could introduce extra sources of uncertainty. We will clarify this limitation in the revised manuscript.

RC2-SC5

Line 123 (CHD boundaries): Were the constant head boundaries for the Ganges and Yamuna held fixed over the full 200-year simulation? If so, could the authors discuss the potential influence of this assumption on results, given known changes to river stages from dam construction and diversions over this period?

AR: Yes, the constant head boundaries for the Ganges and Yamuna were kept fixed throughout the simulation period. We agree that river stages have likely changed over the past two centuries due to construction of dams, diversions, and other large-scale hydrological alterations. However, historical river stage observations were not available for this study, partly due to data restrictions. Additionally, because the Ganges and Yamuna are large, deeply incised rivers, we expect the regional groundwater gradients controlled by these boundary conditions to remain relatively stable, even if river stages have varied over time.

Although river stages from larger-scale hydrological studies could potentially have been incorporated, this was considered outside the scope of the present study. Our objective was specifically to isolate and evaluate the effects of local land- and water-use changes (i.e. recharge and abstraction changes) on the regional groundwater system. By keeping the *boundary conditions fixed, simulated groundwater changes can, therefore, be attributed primarily to changes in local forcing rather than external influences from upstream river regulation or Himalayan climate-driven discharge changes. We acknowledge that this simplification may affect the magnitude of simulated SW–GW exchanges near the model boundaries and will clarify this limitation in the revised manuscript*.

RC2-SC6
Line 155 ("scenarios"): Consider replacing with "realizations" or "ensemble members" throughout, as these terms better reflect the probabilistic nature of the Monte Carlo approach.
AR: Agreed, we will change "scenarios" into "realisations".

RC2-SC7
Lines 181–190 (realism criteria): Could the authors discuss whether these filtering criteria introduce any systematic bias toward particular parameter combinations given the model doesn't account for uncertainty in spatial heterogeneity in parameters? (ie. could the model be compensating for processes not captured numerically?) It would be useful to discuss how many realizations are filtered based on each criteria.
AR: We agree that parameter compensation could play a role in the results. In our framework, different parameter sets can indeed compensate for one another, for example, a higher recharge factor of specific recharge component may be offset by lower recharge factor of another recharge component or lower abstraction component factor or higher hydraulic conductivity. This is reflected in the spread of accepted realisations.
Regarding subsurface simplification, the model assumes spatially uniform aquifer properties, whereas the real system is heterogeneous. Some of this unresolved spatial subsurface variability may therefore be implicitly absorbed into the spatially explicit forcings terms (recharge and abstractions), which can lead to compensatory effects between parameters and forcings. However, we do not find evidence that the realism filters introduce a strong systematic bias toward a specific subset of parameter combinations.
We will clarify this point in the revised manuscript and will also report the number of rejected realizations associated with each individual realism criteria to improve transparency.

RC2-SC8
Figure 1.2: Could the authors clarify how irrigation canals are represented in the model conceptualization, and why rivers and canals were treated differently?
AR: (This response addresses both comments RC1-SC5 and RC2-SC8, which concern the same issue). Irrigation canals were represented as recharge sources rather than as explicit surface-water boundary conditions (e.g. using the RIV package). This conceptualization assumes that canals are generally located at relatively elevated positions in the landscape while groundwater tables are comparatively deeper, resulting in predominantly downward leakage from canals to the aquifer. This is a realistic assumption backed by field observations. Under this assumption, canal leakage is treated as disconnected surface water recharge and is, therefore, not directly influenced by simulated groundwater levels.
This approach was also adopted because detailed data on canal bathymetry, canal-bottom elevations, and surface water levels were unavailable. We will clarify this modelling choice and its implications in the revised manuscript.

RC2-SC9

Figures 8 and 10: Could the authors be more explicit about which periods are being differenced (ie. from end of last period)?

AR: We agree that the description of the differencing approach in figures 8 and 10 was not sufficiently clear. The first and last panels show the mean of respectively the "natural" and "current" periods. The panels in between show the mean of the period minus the mean of the previous period. E.g., the difference map with period label "1800-1830 -> 1830-1900" shows the mean of the period 1830-1900 minus the mean of the period 1800-1830. We will make figures 8 and 10 more clear.

RC2-SC10

Pre-1900 uncertainty bounds (Figure 5): Could the authors explain why there is an apparent narrowing of uncertainty bounds prior to 1900? Given this coincides with the inclusion of meterological data which captures greater interannual variability, could this indicate the hindcast uncertainty is being underpredicted?

AR: The apparent narrowing of uncertainty bounds prior to 1900 is mainly related to the reduced temporal variability in the forcing data during this period. Before 1900, long-term average meteorological conditions were used, which do not capture interannual weather variability and, therefore, result in smoother recharge forcing and reduced ensemble spread.

In addition, fewer recharge and abstraction components were active during the pre-1900 period. The system was primarily influenced by rainfall recharge, canal leakage, and irrigation return flow, whereas in later periods additional recharge and abstraction components contributed their own uncertainties. As a result, total forcing uncertainty was lower in the earlier time periods.

We agree that this simplification may lead to an underestimation of hindcast uncertainty prior to 1900. However, the main conclusions of the study are based primarily on the median behaviour of the posterior ensemble, which would likely remain similar even if additional interannual meteorological variability was included. Incorporating transient historical climate variability over the full hindcast period would substantially increase model complexity and was considered beyond the scope of the present study. We will clarify this limitation in the revised manuscript.

RC2-SC11

Appendix A.6 (irrigation return flow): The return flow factor of 0.162 is applied uniformly across the domain. Could the authors discuss whether this uniform parameterisation may introduce systematic spatial bias, given known variability in irrigation method, soil type, and crop distribution across the region?

AR: We agree that irrigation return flow likely varies spatially due to differences in irrigation methods, soil properties, crop types, and local management practices. However, detailed spatial datasets on these factors were not available for the study area, particularly for the historical reconstruction period. Furthermore, temporal variability in crop rotations and irrigation practices is expected to smooth spatial differences. We, therefore, applied a uniform return flow factor across the model domain as a necessary simplification.

We acknowledge that this assumption may introduce spatial biases in the estimated recharge patterns at the local scale. For this reason, the model results should primarily be interpreted at the broader regional scale and with emphasis on long-term temporal trends rather than cell-specific estimates. We will clarify this limitation in the revised manuscript.