## RC1: 'Comment on egusphere-2025-2124', Anonymous Referee #1

This manuscript develops and presents a novel multivariate data assimilation framework and implements it in an integrated hydrology model. This approach is tested in a catchment in Germany using groundwater and soil moisture observations. It is interesting that this approach resulted in improved state variables and also in some minor improvement to related fluxes such as ET.

*Response: Thanks for the cogent summary of our work.*

The manuscript is clearly written and organized. I recommend this manuscript be published pending some minor revisions and / or responses to my comments below.

*Response: Thanks for your positive feedback.*

1. Are (or how are) the _PAR runs spun up to account for potential impact of the parameters (e.g. ln(Ksat)) on model equilibrium states? The methods section does a nice job describing this process in general and it appears that great care was taken here to ensure a good model spin up, but it is still a bit unclear how this interacted with the adjustment of model inputs for these cases.

*Response: We appreciate the reviewer's attention to the spin-up procedure. The concern about the interaction between parameter updates in the _PAR experiments and the model equilibrium state is indeed important. In our setup, all experiments were initialized from the same carefully spun-up model state. For the _PAR runs, the saturated hydraulic conductivity (log $K_s$) was updated every seven days using a damping factor of 0.1, ensuring that parameter changes were gradual and did not disturb so much the equilibrium state, thereby avoiding the need for additional spin-ups. In addition, we validated the stability of the updated parameters by applying $K_s$ values obtained from one year (e.g., 2016) in open loop simulations of other years (2017-2018), confirming that no spin-up-related issues were introduced. To clarify this point further, we have revised the Section 3.2 (Configuration of Data Assimilation Experiments) as follows (lines 362-368):*

"For all DA experiments, the suffix _PAR indicates that, in addition to state updates, the saturated hydraulic conductivity (log $K_s$) was updated every seven days using a damping factor of 0.1. The _PAR runs were initialized from the same spun-up equilibrium state as their corresponding state-update experiments, and the gradual parameter updates ensured that changes remained small and did not disturb the equilibrium state too much, thereby avoiding the need for additional spin-ups. Furthermore, parameter validation involved applying $K_s$ updated from one year to OL simulations in other years (e.g., using updated $K_s$ from 2016 in 2017-2018)."

2. Table 3 would benefit from some narrative text to describe the different experiments, in words, to help better describe them.

*Response: Your point is taken. Following the advice, we have added more descriptions of the data assimilation experiments listed in Table 3 to facilitate understanding. The new text has been added in Section 3.2 Configuration of Data Assimilation Experiments (lines 351-368 in the revised manuscript) and reads:*

"11 DA experiments (Table 3) were conducted to assess assimilation performance, differing in observation type, state vector composition, and localization strategy. The open loop (OL) experiment, performed without assimilation, served as the reference for DA comparisons. SM_DA assimilated daily SM observations from CRNS (observation error of 0.03 $cm^3/cm^3$) with 100 km localization radius. GWL_DA assimilated weekly GWL observations, with an observational error of 0.05 m, using a 5 km localization radius, updating only hydraulic head ($h$) in the saturated zone. FC_DA assimilated both SM and GWL using the fully coupled DA strategy, with the state vector including $h$ and $\theta$ in all subsurface layers. $\theta$ and $h$ were updated daily and weekly, respectively, both with a 5 km localization radius. WC_DA used the weakly coupled scheme, with $h$ updated only in the saturated zone and $\theta$ only in the unsaturated zone; all other settings were the same as FC_DA. Moreover, WC_DA_r followed the same setup as WC_DA, except that the localization radius differed between the two variables: 5 km for GWL and 100 km for SM. For all DA experiments, the suffix _PAR indicates that, in addition to state updates, the saturated hydraulic conductivity (log $K_s$) was updated every seven days using a damping factor of 0.1. The _PAR runs were initialized from the same spun-up equilibrium state as their corresponding state-update experiments, and the gradual parameter updates ensured that changes remained small and did not disturb the equilibrium state too much, thereby avoiding the need for additional spin-ups. Furthermore, parameter validation involved applying $K_s$ updated from one year to OL simulations in other years (e.g., using updated $K_s$ from 2016 in 2017-2018)."

3. In the experiments that assimilate model derived outputs (such as WTD) and those that also adjust ln(Ksat) there does not appear to be a difference in temporal behavior. That is, often if there is bias in say soil moisture it will drift back to a value different than the observation after the assimilation period over time. Adjusting ln(Ksat) should help correct this model "bias" but I'm not seeing this behavior in the results (but agree with the overall improvement the authors point out). I'm wondering if they can comment on this more?

*Response: We thank the reviewer for this valuable observation. In our experiments, several factors may limit the visible differences in temporal evolution between simulations with and without saturated hydraulic conductivity ($K_s$) updates. First, model biases are influenced not only by $K_s$ but also by other factors, such as forcing uncertainty and structural model errors, which can dominate the temporal evolution of soil moisture and groundwater states. Second, $K_s$ updates were applied with a fixed damping factor of 0.1, which constrained the adjustment range, particularly for slowly responding groundwater states. Nevertheless, adjusting $K_s$ during assimilation can still reduce model bias in variables. As a result,* *while temporal patterns may appear similar, experiments with parameter updates show clear improvements in performance metrics and long-term mean states. Independent validations using the revised $K_s$ confirmed enhanced predictions of both GWL and SM. We have added a brief discussion of these points in Section 5.2 (lines 819-833) to clarify:*

"The primary objective of multivariate DA is to enhance the accuracy of both state variables and associated parameter estimates. This research focused on updating $K_s$, identified as a critical parameter for the subsurface groundwater system. Although the temporal evolution of assimilated states may not show large differences between experiments with and without $K_s$ updates, this does not imply that parameter updating is ineffective. For example, in our experiments, $K_s$ updates led to reductions in ubRMSE of more than 10% for both GWL and SM compared with state-only assimilation. However, the immediate temporal impact of $K_s$ updates may be limited, partly due to the constrained adjustment range applied by the fixed damping factor (0.1) and the slow response of groundwater states. Moreover, model biases are also influenced by other factors, including forcing uncertainty and structural model errors, which may play a dominant role in the temporal evolution of SM and groundwater states. Nevertheless, parameter-updating experiments improved performance metrics and long-term mean states, demonstrating their value in correcting systematic model biases that cannot be fully addressed by state assimilation alone. Independent validations using the revised $K_s$ confirmed enhanced predictions of both GWL and SM. These results highlight the importance of considering both state and parameter updates in multivariate assimilation frameworks to achieve more reliable hydrologic predictions."

## RC2: 'Comment on egusphere-2025-2124', Anonymous Referee #2

This is a well written and well executed technical study on multivariate DA for a catchment in Germany. The authors have a history of such studies, and this is another nice example.

*Response: Many thanks for your positive feedback.*

My main suggestion, to improve the potential impact of the paper, is that the authors expand on some of their comments because – at least to me – they are central to the motivation and implications of the study. However, they are currently not given appropriate attention.

*Response: Your point is taken. Please see our responses and revisions addressing your concerns below.*

Main comments:

[1] Why is there a trade-off between groundwater levels and soil moisture? How strongly are SM and GWL connected? How does this connection depend on the location within the catchment? Is it larger when GWL are shallower, and does it (for CNRS estimates) disconnect for deeper GWL?

*Response: We thank the reviewer for raising such important point. The trade-off between SM and GWL arises from their physical coupling through pressure head relationships and varies spatially: it is strong where groundwater is shallow and weaker where it is deep. Under deeper groundwater conditions, assimilating one variable may have little effect on the other or even slightly degrade its estimates. The limited improvement or occasional slight degradation of the non-assimilated variable can also be partly attributed to small ensemble sizes, which make covariance estimates less reliable. In addition, non-Gaussian soil moisture distributions under dry conditions can reduce the effectiveness of assimilating one variable in improving the estimates of the other. This effect is particularly pronounced for upper soil states when GWL is assimilated, or for deeper subsurface states when SM is assimilated from dry soils. These mechanisms explain why assimilation generally improves the targeted variable while its effect on the non-assimilated one depends on local hydrogeological and statistical conditions. These interpretations are supported by the results presented in Section 4.3 "Multivariate Data Assimilation of Soil Moisture and Groundwater Level" (Figures 7 and 8; Figures S9-S10 for other years). We have also added text in Section 5.1 (lines 658-678) to clarify this mechanism:*

"The observed deterioration may stem from spurious inter-variable covariances generated during the state estimation process. These covariances can modify the natural trade-offs between SM and GWL that arise from their physical coupling through soil water retention and pressure head relationships. Specifically, changes in shallow groundwater directly affect SM in the unsaturated zone, while SM dynamics control recharge and thus influence GWL. Importantly, the strength of this connection is not spatially uniform. In areas with shallow groundwater tables, SM and GWL are tightly coupled, so assimilating one variable has stronger impacts on the other. In contrast, with deeper groundwater, the hydraulic link between SM and GWL weakens, and under such conditions this connection can be functionally disconnected, resulting in assimilating one variable having little or no effect on the other, and in some cases, minor degradations may occur. Such degradations may be partially caused by small ensemble sizes, which make estimated covariances less reliable, especially for weaker correlations. In addition, non-Gaussianity related to drier soil conditions may impair the effectiveness of assimilating one variable on improving the estimates of the other. This effect particularly impacts the upper soil states when GWL is assimilated, or the deeper subsurface states when SM is assimilated from dry soils. This issue can also be partly attributed to the use of point-scale observations, given that neutron sensing stations and groundwater monitoring wells are unevenly distributed across the study area. When assimilation targets only one state component (e.g., GWL), it is difficult to reduce uncertainties in hydrologically connected states (such as SM) at non-adjacent spatial locations. Such spatial heterogeneity and statistical limitation explain why assimilation of a single variable can improve its own estimates while occasionally causing small degradations in the other, depending on local hydrogeological settings."

[2] You state that you propose a "novel multivariate assimilation method*. Can you be more explicit about its novelty, given that you cite references with previous multivariate DA examples. What is novel about the algorithm introduced in this study? I am not doubting its novelty, but you do not elaborate on

this issue when introducing the algorithm, only in the discussion section.

*Response: While multivariate assimilation has been applied in previous hydrological studies, our approach differs in several key methodological aspects. Specifically: We introduce a weakly coupled multivariate DA scheme in which different observation types (CRNS-based SM and GWL) update only their physically related model states (unsaturated and saturated zones, respectively), thereby reducing spurious cross-variable correlations. We apply different localization radii for SM and GWL within the assimilation, accounting for their distinct spatial correlation characteristics, whereas previous DA studies (e.g., TSMP) typically used the same localization radius for both variables in joint assimilation. Additionally, the scheme updates SM and GWL at different intervals according to their typical behavior, with SM updated daily and groundwater weekly, reflecting their different timescales. We agree that these aspects should be stated earlier in the manuscript. Accordingly, we have revised Section 2.4 ("Localized Ensemble Kalman Filter for Data Assimilation", lines 295-308) to highlight these aspects:*

"Earlier research by Zhang et al. (2018) showed that in TSMP, assimilating SM and/or GWL enables updates to all relevant subsurface states via DA. In this fully coupled DA configuration of Zhang et al. (2018), cross-variable covariances ensured that observations of one variable (e.g., SM) could directly adjust others (e.g., GWL). Later, Hung et al. (2022) applied GWL assimilation restricted to the saturated zones and demonstrated that this approach outperformed the fully coupled strategy of Zhang et al. (2018). In this study, we develop a new weakly coupled DA scheme that introduces separate update restrictions for each observation type: GWL observations are used to update only saturated cells, while SM observations are used to update only unsaturated zones. This design minimizes potential spurious cross-variable correlations and enhances the robustness of multivariate assimilation. Additionally, updates are applied asynchronously to account for the different temporal dynamics of the variables: SM, which changes more rapidly, is typically updated daily, whereas groundwater, with slower dynamics, is updated weekly. Furthermore, unlike previous DA studies of TSMP, which generally used the same localization radius for joint GWL and SM assimilation, our approach applies different localization radii for the two variables, accounting for their distinct spatial correlation characteristics."

[3] Lines 268-269. The authors state that: "With a localization radius of ~100 km, exceeding the domain size, assimilation effects covered the entire area." Can you please elaborate on the meaning of this statement?

*Response: In the assimilation process, the localization radius defines the spatial extent to which observations influence model state updates. Our goal is to compare LEnKF multivariate assimilation of SM and GWL with previous single-variable EnKF studies in the Rur catchment (Baatz et al., 2017; Li et al., 2023a), where assimilation of 12 CRNS stations improved SM across the entire catchment. To maintain comparability, we set the localization radius to ~100 km, which is sufficient to cover all 100 × 162 grid cells (500 m resolution) within the catchment. Using a much larger radius, e.g., 1000 km, would not noticeably change catchment-scale effects, but could introduce unphysical correlations in the LEnKF ensemble. Therefore, a radius of ~100 km provides a balance between physical realism and consistency* *with earlier studies. We have added the following clarification in the revised manuscript (lines 285-288):*

"The Rur catchment model consists of 100 × 162 grid cells with a resolution of 500 m × 500 m. Following previous EnKF studies using 12 CRNS stations (Baatz et al., 2017; Li et al., 2023a), we set the localization radius to ~100 km to ensure that assimilation effects cover the entire study area."

[4] The model improves through the assimilation, but what have you learned about the model? And especially its limitations? What does the updating reveal about problems within the model (that make the updating necessary)?

*Response: We thank the reviewer for this important question. Our results show that the assimilation not only improves model states but also reveals several limitations of the model. These limitations can be grouped into several categories, which are discussed in detail in the revised discussion section:*

***① Cross-variable trade-offs:***

*The updating process shows that single-variable assimilation improves the targeted variable but degrades the non-assimilated one. This reflects the model's limited ability to represent the nonlinear coupling between soil and groundwater, and highlights the influence of the uneven distribution of observations. This limitation increases the complexity of DA and may compromise filter performance. These issues are discussed in the revised manuscript as follows (lines 657-674):*

"When assimilation is limited to a single variable, either SM or GWL, it generally enhances the assimilated variable but frequently decreases the reliability of the non-assimilated one. The observed deterioration may stem from spurious inter-variable covariances generated during the state estimation process. These covariances can modify the natural trade-offs between SM and GWL that arise from their physical coupling through soil water retention and pressure head relationships. Specifically, changes in shallow groundwater directly affect SM in the unsaturated zone, while SM dynamics control recharge and thus influence GWL. Importantly, the strength of this connection is not spatially uniform. In areas with shallow groundwater tables, SM and GWL are tightly coupled, so assimilating one variable has stronger impacts on the other. In contrast, with deeper groundwater, the hydraulic link between SM and GWL weakens, and under such conditions this connection can be functionally disconnected, resulting in assimilating one variable having little or no effect on the other, and in some cases, minor degradations may occur. Such degradations may be partially caused by small ensemble sizes, which make estimated covariances less reliable, especially for weaker correlations. In addition, non-Gaussianity related to drier soil conditions may impair the effectiveness of assimilating one variable on improving the estimates of the other. This effect particularly impacts the upper soil states when GWL is assimilated, or the deeper subsurface states when SM is assimilated from dry soils. This issue can also be partly attributed to the use of point-scale observations, given that neutron sensing stations and groundwater monitoring wells are unevenly distributed across the study area."

*And lines 810-816:*

"The EnKF, originally developed to address nonlinearity in dynamic modeling systems, has demonstrated effectiveness in coupled terrestrial simulations. This nonlinearity primarily arises from the complex interdependencies among state variables, such as the coupling between SM and GWL through pressure head dynamics (Camporese et al., 2009b). This inherent nonlinearity complicates the design of multivariate assimilation schemes. As a result, determining the most suitable observational inputs and evaluating the compromises associated with integrating diverse variables continue to pose major obstacles for upcoming investigations."

***②Spatial resolution limitations:***

*Coarser spatial resolution typically smooths terrain features, which reduces gradients in both surface and groundwater flows and likely contributes to persistent discrepancies in simulated GWL. This aspect is discussed in the manuscript (lines 756-765) as detailed below:*

"While advancements have been made, the assimilation results still indicate unresolved uncertainties that should be addressed in future work. Part of this uncertainty arises from the model's use of coarse spatial discretization. Coarser spatial resolution typically smooths terrain features, which reduces gradients in both surface and groundwater flows and likely contributes to persistent discrepancies in simulated GWL. Moreover, DA tends to be less effective in the presence of such systematic biases. For example, Xue et al. (2021) systematically evaluated hydrological simulations over High Mountain Asia using models at different spatial resolutions, and found that coarse model resolution introduced systematic biases in runoff, particularly over complex terrain, thereby limiting the effectiveness of DA. Future research could explore finer spatial resolutions (e.g., 100 m) to more accurately represent groundwater systems linked to narrow valleys, thereby minimizing biases caused by coarse spatial discretization and improving DA performance."

***③Structural simplifications:***

*The updating process also highlights structural simplifications in the model. For example, GWL assimilation assumes hydrostatic equilibrium, while real-world aquifer systems are more complex, with layered aquifers, aquitards, and anthropogenic withdrawals. These simplifications contribute to systematic biases and make updating necessary. This point is discussed in Section 5.2 (Uncertainty Analysis and Enhancement Strategies) of the manuscript (lines 790-800).*

"Beyond spatial resolution and observation distribution, structural deficiencies in the model may contribute to persistent uncertainties and further complicate the effective application of DA with real-world observations. This study performs GWL assimilation under the simplifying assumption of hydrostatic equilibrium, even though real-world conditions are considerably more complex. Multiple aquifers can coexist in a vertically layered system, separated by intervening aquitards. Additionally, fault lines may act as horizontal barriers that disrupt aquifer continuity, potentially altering groundwater flow patterns and their spatial distribution. Anthropogenic groundwater withdrawal also significantly affects aquifers. This is particularly evident in the Rur catchment, where hydrogeological conditions are strongly influenced by water management practices aimed at preventing water accumulation in open-cast lignite mines (Bogena et al., 2018). These processes are insufficiently represented in the current model, which contributes to systematic biases and makes updating necessary."

***④ Parameterization issue:***

*The experiments show that updating only saturated hydraulic conductivity ($K_s$) improves long-term states but has little short-term effect. This reveals that systematic biases cannot be corrected by state updates alone and may require additional parameter and model improvements. This is discussed in Section 5.2 of the manuscript (lines 819-831).*

"The primary objective of multivariate DA is to enhance the accuracy of both state variables and associated parameter estimates. This research focused on updating $K_s$, identified as a critical parameter for the subsurface groundwater system. Although the temporal evolution of assimilated states may not show large differences between experiments with and without $K_s$ updates, this does not imply that parameter updating is ineffective. For example, in our experiments, $K_s$ updates led to reductions in ubRMSE of more than 10% for both GWL and SM compared with state-only assimilation. However, the immediate temporal impact of $K_s$ updates may be limited, partly due to the constrained adjustment range applied by the fixed damping factor (0.1) and the slow response of groundwater states. Moreover, model biases are also influenced by other factors, including forcing uncertainty and structural model errors, which may play a dominant role in the temporal evolution of SM and groundwater states. Nevertheless, parameter-updating experiments improved performance metrics and long-term mean states, demonstrating their value in correcting systematic model biases that cannot be fully addressed by state assimilation alone."

[5] (section 4.1) When assimilating SM only, ET and GWL changed by just a few percent. How relevant is this change? How does this change compare to – say – making slightly different assumptions about the noise and noise structure? Or is this small change the equivalent to essentially no change? You state in your abstract that: "However, assimilating GWL independently had a negative effect on SM representation, and similarly, assimilating SM alone degraded GWL predictions." This effect seems very minimal, and I do not really see a significant decline in SM performance when GWL is assimilated.

*Response: We thank the reviewer for the comment. In the revised manuscript, we explicitly report that while univariate assimilation improves the targeted variable, it may also introduce minor degradations in the non-assimilated variable. The corresponding text in Section 4.1 (Univariate Soil Moisture Assimilation, lines 402-407) has been revised to more explicitly describe the changes in GWL under univariate SM assimilation.*

"However, GWL ubRMSE metric showed variable changes when only SM was assimilated, ranging from -7% to 15% across individual years, where positive values indicate a deterioration in performance. Over the 2016-2018 period, the average change in ubRMSE was small, corresponding to 3.87% for SM_DA and -0.41% for SM_DA_PAR. Overall, SM assimilation had a minor negative effect on GWL, with some annual variability."

*The text in Section 4.2 (Univariate Groundwater Level Assimilation, lines 459-465) has also been* *updated to provide a clearer description of how SM is affected by univariate GWL assimilation.*

"In contrast to the large improvements in GWL, univariate GWL assimilation generally had a limited negative impact on SM, with interannual variability. Over the 2016-2018 period, the average SM ubRMSE was 0.09 $cm^3/cm^3$ in GWL_DA and 0.11 $cm^3/cm^3$ in GWL_DA_PAR, compared with 0.09$cm^3/cm^3$ in OL. In individual years, SM ubRMSE ranged from 0.09 to 0.10 $cm^3/cm^3$ in GWL_DA, corresponding to annual changes of 0-25% compared with OL value in each respective year. In GWL_DA_PAR, SM ubRMSE further increased to 0.10~0.11 $cm^3/cm^3$, reflecting annual rises of over 20% relative to OL values of the corresponding year."

[6] (section 4.2) How would you reduce the problem that performance declines with distance from assimilation wells?

*Response: Thanks for pointing this out. As noted in the manuscript, assimilation performance indeed declines with distance from observation wells. To address this, several local improvements can be applied. First, adaptive localization can directly mitigate performance decline with distance by adjusting the influence radius based on local conditions. Second, increasing the number of wells and using higher spatial resolution reduces grid-cell aggregation and wet biases, allowing more wells to be effectively assimilated and mitigating performance decline in distant areas. At larger scales, integrating spatially distributed datasets, such as remote sensing products for soil moisture (such as SMOS, SMAP, and AMSR-E/AMSR2) and terrestrial water storage (such as GRACE/GRACE-FO), can provide additional constraints across the domain, further mitigating performance decline with distance. We have added extra discussions of these potential approaches and how they address this issue in the revised manuscript (lines 765-773).*

"Furthermore, the performance of assimilation tends to decline with increasing distance from observation wells, as localized updates have weaker influence on more remote areas. Potential strategies to mitigate this issue include applying adaptive localization radii, assimilating spatially distributed datasets (e.g., RS products), or increasing the number of groundwater wells to enhance spatial coverage. Employing higher spatial resolution reduces the likelihood of multiple observation wells being located within a single grid cell, thereby allowing a larger number of wells to be effectively assimilated. It also reduces wet biases in simulated GWL, decreasing the probability of wells falling within river or near-river grid cells and thereby increasing the number of observations that can be reliably assimilated."

*And lines 782-789:*

"To broaden the applicability of this approach, future studies could focus on integrating more widely accessible datasets, such as terrestrial water storage variations derived from GRACE/GRACE-FO (Tapley et al., 2019; Khaki et al., 2017) or RS-based SM products (Bayat et al., 2021). Such spatially distributed observations could also help to reduce the decline in assimilation performance with distance from individual ground-based observations, thereby providing additional constraints across larger areas. However, these data products are unfortunately too coarse to resolve hydrological processes in our study area, highlighting the need for higher-resolution observations for effective local-scale assimilation."

[7] (discussion section) Can the independent updating of different parts of the model lead to water balance issues?

*Response: We thank the reviewer for this important question. Independent updates of SM and groundwater states in our framework can introduce temporary local water balance perturbations, which are a normal consequence of data assimilation and may persist depending on site-specific conditions. In our coupled TSMP system, these perturbations are dynamically adjusted through surface-subsurface interactions, preventing systematic errors at the catchment scale. While not necessarily larger than in uncoupled models, these local imbalances are dynamically regulated in coupled systems through interactions between surface and subsurface processes. This point has been added in the revised discussion (Section 5.1, lines 719-728).*

"Beyond improving state estimates, the impact of independent updates on water balance needs to be considered. During assimilation, SM and groundwater states are modified directly, which can temporarily disturb the local water balance. These imbalances may persist for a period depending on site-specific conditions. Such local imbalances are common in data assimilation, but the tight coupling between CLM and ParFlow ensures that surface and subsurface fluxes redistribute these adjustments through the model's physical processes. Consequently, at the catchment scale, independent updates do not induce systematic water balance errors, as they only alter storage states and local imbalances are mitigated by the coupled land-subsurface dynamics. Compared to uncoupled models, these local imbalances are not necessarily larger, but in coupled systems they are redistributed differently due to interactions between surface and subsurface processes."

[8] In the conclusion you again state that: "However, assimilating GWL data alone negatively affected SM prediction accuracy, and similarly, assimilating SM data alone reduced the accuracy of GWL estimates." However, the improvement during univariate assimilation is more than a factor 10 than the reduction in the other variables. I think it would be good to discuss this aspect a bit more transparently.

*Response: We thank the reviewer for this suggestion. In the revised manuscript, we have clarified the relative magnitudes of improvements and deteriorations during univariate assimilation. The results indicate that the benefits of univariate assimilation outweigh the negative effects on the non-assimilated variable and highlight the need for multivariate approaches to improve both variables simultaneously. Specifically, in the discussion section (lines: 678-683), we now state:*

"However, the observed reductions in the non-assimilated variable are relatively small compared with the improvements in the assimilated variable, suggesting that univariate assimilation still provides substantial benefits for the targeted state. These limitations of univariate assimilation underscore the value of multivariate approaches, which may better account for the coupled dynamics of SM and GWL and improve the accuracy of both states simultaneously."

*In the conclusion (lines 861-867), we have quantified these effects:*

"However, assimilating GWL data alone negatively affected SM prediction accuracy, with the 2016-2018 average ubRMSE increasing by approximately 20%. Similarly, assimilating SM data alone reduced the accuracy of GWL estimates, leading to a less than 4% rise in the 2016-2018 average ubRMSE. Overall, the improvements in the targeted state clearly exceeded the limited deteriorations in the non-assimilated state, demonstrating the benefit of univariate assimilation. This also highlights the importance of multivariate approaches for achieving simultaneous improvements in both variables."

## CC1: 'Comment on egusphere-2025-2124', Nima Zafarmomen

This is a carefully done study and the findings are of considerable interest. And the submission is worth of publication. Following are some minor comments:

1. The paper presents a novel multivariate data assimilation (DA) framework for integrating cosmic-ray neutron sensor (CRNS) soil moisture (SM) and groundwater level (GWL) data into the Terrestrial System Modeling Platform (TSMP). The approach is innovative in its use of a weakly coupled DA scheme to independently update saturated and unsaturated zones, addressing limitations of fully coupled DA systems. However, the manuscript could strengthen its claim of novelty by explicitly comparing the proposed method to existing multivariate DA frameworks in other coupled models (e.g., CATHY, Flux-PIHM, MIKE-SHE, and SWAT-MODLFOW) beyond the brief mentions in the introduction and discussion. For example, "Assimilation of Sentinel-Based Leaf Area Index for Modeling Surface-Ground Water Interactions in Irrigation Districts" used the integrated model fully coupled DA systems and you can cite it to make a stronge fundation for you research. Moreover, the introduction effectively highlights the importance of SM and GWL in terrestrial hydrology and the role of DA in reducing model uncertainties. However, it could better contextualize the study by discussing recent advancements in CRNS technology and its adoption in DA frameworks. For instance, referencing studies like Bogena et al. (2022) or Schrön et al. (2017) earlier in the introduction would clarify why CRNS is a superior choice compared to traditional in-situ or RS-based SM data.

*Response: We thank the reviewer for these constructive comments and for recognizing the value of our study. We have expanded the introduction to provide a more explicit overview of existing multivariate DA approaches implemented in coupled hydrological models, including CATHY, Flux-PIHM, and MIKE-SHE. Reference to related multivariate DA framework in SWAT-MODFLOW has also been added to better introduce our approach within existing studies. In the revised manuscript, the introduction now includes the following text reviewing existing multivariate DA studies (lines 88-103):*

"Previous studies have applied multivariate EnKF within coupled models like CATHY and Flux-PIHM to jointly assimilate multiple observations, including SM, groundwater, discharge, and land surface fluxes, demonstrating improved estimates of hydrologic states and parameters (Camporese et al., 2009a; Shi et al., 2014; Botto et al., 2018; Shi et al., 2015). Despite being tested primarily on small experimental catchments, these multivariate DA frameworks remain computationally intensive and may involve trade-offs among variables. Some parameters can only be identified under specific hydrological conditions, particularly in strongly nonlinear problems involving the unsaturated zones. To overcome these challenges, some studies have explored alternative multivariate DA strategies within coupled models. Using MIKE-SHE, Zhang et al. (2016) highlighted the importance of spatial and variable-based localization in jointly assimilating SM and groundwater head. Yet, its unsaturated flow is still modeled in one dimension, limiting full system representation. More recently, Zafarmomen et al. (2024) demonstrated that a multivariate particle filter framework assimilating Sentinel-based leaf area index (LAI) and streamflow in a coupled SWAT-MODFLOW model improved estimates of vegetation and hydrologic states. However, the loosely coupled model, in which surface and groundwater components interact via data exchange, may not fully capture integrated dynamics of saturated and unsaturated zones."

*Additionally, the discussion section has been extended to compare our proposed weakly coupled multivariate DA framework with these previous studies, highlighting the novel aspects and strengths of our method. The revised text reads as follows (lines 694-703):*

"In multivariate DA, previous studies have shown that challenges persist despite methodological advances. Shi et al. (2015) combined model states and global calibration coefficients into a high-dimensional joint vector, requiring covariance relaxation, conditional covariance inflation, and quality control to prevent filter divergence and ensure physical plausibility. Zhang et al. (2016) employed distance and variable localization to control spurious correlations in joint SM and groundwater head assimilation, but this approach relies on manually defined rules and may lose physically meaningful cross-variable information. Botto et al. (2018) applied normalization to measurement error covariance matrices and addressed simulated data anomalies and innovation vectors to prevent ill-conditioning of the Kalman gain. While these measures ensure numerical stability, they require careful manual scaling of each variable."

*We have also revised the introduction to better highlight recent advancements in CRNS technology and its integration into data assimilation frameworks, citing relevant studies (e.g., Schrön et al., 2017; Bogena et al., 2022). These revisions clarify why CRNS provides superior soil moisture observations compared to conventional in-situ and RS-based measurements and explain why CRNS was chosen as the data source in our assimilation framework. This revised text has been added in the manuscript (lines 64-76) and is detailed below:*

"As an alternative, Cosmic-Ray Neutron Sensors (CRNS) (Zreda et al., 2008) provide reliable, non-invasive SM estimates at the field scale (~18 ha), with deeper penetration (~80 cm) and reduced bias compared to RS products (Zreda et al., 2012; Köhli et al., 2015; Bogena et al., 2022). Recent advances in CRNS techniques, including improved footprint characterization and revised calibration strategies, have substantially enhanced its robustness (Franz et al., 2013; Köhli et al., 2015; Schrön et al., 2017). As a result, CRNS data have been adopted in diverse applications such as hydrology, snow and vegetation monitoring, and land surface modeling (Fersch et al., 2020; Dimitrova-Petrova et al., 2021; Bogena et al., 2022). With the establishment of long-term monitoring networks, CRNS data have also been increasingly integrated into DA frameworks (Baatz et al., 2017; Cooper et al., 2021; Patil et al., 2021). By bridging the scale gap between point measurements and model grids, CRNS serves an effective data source in DA frameworks, thereby reducing model uncertainties and enhancing the reliability of terrestrial hydrology simulations (Shuttleworth et al., 2013; Han et al., 2015; Baatz et al., 2017; Mwangi et al., 2020)."

2. The description of the Rur catchment and data sources (Section 2.1 and 2.3) is thorough, but the rationale for selecting specific CRNS sites and GWL wells is not fully explained. For example, why were only 78 out of 616 wells used for DA, and how was the median GWL selection criterion determined? A brief justification of these choices would enhance transparency.

*Response: We thank the reviewer for highlighting the need to clarify our data selection criteria. To ensure the reliability and representativeness of groundwater level (GWL) observations for assimilation, we applied multiple selection criteria to wells across the study area. Selected wells were required to have GWL observations at depths of 0-20 m within the unconfined upper aquifer, provide at least monthly records during the assimilation period, and should not be located in persistently saturated or river-adjacent grid cells. In addition, when multiple wells were located within a single model grid cell, the well with the median GWL was selected as representative. From a groundwater modeling and parameterization perspective, this ensures that the assimilated observation reflects typical conditions within the grid cell and provides a stable and representative input for data assimilation, while wells with consistently higher or lower GWL may represent extreme or unrepresentative conditions and could introduce biases if used. The median GWL is therefore a robust choice for assimilating state variables and model parameters. Following these criteria, 78 wells were selected for data assimilation, and the remaining wells were reserved for independent validation. A brief explanation of these selection criteria has been added to Section 2.3.3 (Field Measurements of Soil Moisture, Groundwater, and Evapotranspiration) in the revised manuscript (lines 213-225) to clarify our data selection criteria.*

"Groundwater table depth data for assimilation and independent validation were obtained from the Geoportal NRW platform (www.geoportal.nrw, accessed May 2, 2025). Given the weak hydraulic connectivity between the RZSM and the deep confined aquifer, this study focused on assimilating data from the unconfined upper aquifer. Wells selected exhibited observation depths between 0 to 20 meters and supplied records with at least monthly observations. In total, 616 wells met these criteria during the 2016-2018 period (Fig. 1). Due to the 500 m model resolution and the spatial clustering of observation wells near rivers, multiple wells were often located within a single grid cell or within river cells. To ensure representative observations for assimilation, the median GWL was chosen among multiple wells within a grid cell to minimize potential biases from unusually high or low groundwater levels. Additionally, grid cells adjacent to stream networks were excluded from the assimilation process, as persistent saturation in these areas caused large discrepancies with observed values. Accordingly, wells situated in river grid cells were excluded from the assimilation. Following these screening procedures, 78 wells were selected for DA, while the remaining 465 wells were reserved for independent validation."

3. The results section provides a comprehensive analysis of ubRMSE, RMSE, and correlation coefficients across multiple experiments. However, the focus on ubRMSE as the primary metric is not fully justified. While ubRMSE accounts for bias, a discussion of why it is prioritized over RMSE or other metrics (e.g., mean absolute error) would clarify its relevance.

*Response: In this study, three performance metrics, ubRMSE, RMSE, and R were calculated. ubRMSE is highlighted in the manuscript because it is commonly used in the remote sensing DA literature, which facilitates comparison with existing studies. Focusing on a single metric in the main text also helps avoid redundancy. RMSE and R are also reported in Supplementary Tables to provide a comprehensive assessment of model performance. Section 3.3 (Model Performance Assessment, lines 380-382) has been revised to clarify this motivation.*

"Among these, ubRMSE was emphasized in our analysis because it is widely applied in DA research and facilitates comparison with previous studies. To avoid redundancy, detailed results for RMSE and R are presented in the Supplementary Tables to ensure a comprehensive evaluation of model performance."

4. The discussion compares the proposed method to Hung et al. (2022) and Zhang et al. (2016), but it could be expanded to include other multivariate DA studies (e.g., Botto et al., 2018; Shi et al., 2015) to highlight how the weakly coupled approach addresses their limitations. For instance, how does the proposed method mitigate the issue of spurious correlations noted in Zhang et al. (2016)?

*Response: We thank the reviewer for this valuable suggestion. We have expanded discussion to include Botto et al. (2018) and Shi et al. (2015) to highlight limitations of previous multivariate DA, which can propagate observation errors and generate spurious correlations, as noted by Zhang et al. (2016). In contrast, our weakly coupled DA scheme updates variables sequentially, applying variable-specific spatial localization to restrict cross-variable influence to physically meaningful scales. This approach reduces spurious correlations and maintains physically consistent interactions. The revised manuscript now includes this extended discussion (5.1 Benefits and Challenges of the New Multivariate Data Assimilation Framework, lines 694-718).*

"In multivariate DA, previous studies have shown that challenges persist despite methodological advances. Shi et al. (2015) combined model states and global calibration coefficients into a high-dimensional joint vector, requiring covariance relaxation, conditional covariance inflation, and quality control to prevent filter divergence and ensure physical plausibility. Zhang et al. (2016) employed distance and variable localization to control spurious correlations in joint SM and groundwater head assimilation, but this approach relies on manually defined rules and may lose physically meaningful cross-variable information. Botto et al. (2018) applied normalization to measurement error covariance matrices and addressed simulated data anomalies and innovation vectors to prevent ill-conditioning of the Kalman gain. While these measures ensure numerical stability, they require careful manual scaling of each variable.

In contrast, the weakly coupled DA scheme adopted in this study updates states and parameters sequentially, with each variable employing its own spatial localization and independent updates. This allows saturated zone pressure to be updated using GWL observations, while SM estimates in the unsaturated zone are adjusted based on CRNS-derived measurements. The use of variable-specific localization parameters further improves the representation of their distinct spatial characteristics, reduces the influence of spatially distant uncertainties, and limits unphysical information propagation. Importantly, this framework achieves these benefits without requiring extensive manual tuning or high-dimensional corrective procedures, which are often needed in traditional multivariate DA approaches. Additionally, asynchronous assimilation enables different update intervals for each variable: SM, which varies rapidly, is typically updated daily, whereas groundwater, with slower dynamics, is updated weekly. This approach allows coupled models to better accommodate the differing timescales of fast-evolving and slowly changing processes and to assimilate multiple variables from diverse data inputs. These characteristics enhance the robustness and reliability of the assimilation framework in real-world catchments, where observations are spatially heterogeneous and hydrological processes operate across multiple timescales."

5.Terms like "weakly coupled DA" and "fully coupled DA" are used consistently but may confuse readers unfamiliar with DA jargon.

*Response: We thank the reviewer for this helpful comment. To improve clarity for readers unfamiliar with data assimilation terminology, we have added explicit definitions and explanations of the terms "weakly coupled DA" and "fully coupled DA" in the section 2.4 (Localized Ensemble Kalman Filter for Data Assimilation). The revised manuscript now includes the following explanations to clarify these terms (lines 295-308):*

"Earlier research by Zhang et al. (2018) showed that in TSMP, assimilating SM and/or GWL enables updates to all relevant subsurface states via DA. In this fully coupled DA configuration of Zhang et al. (2018), cross-variable covariances ensured that observations of one variable (e.g., SM) could directly adjust others (e.g., GWL). Later, Hung et al. (2022) applied GWL assimilation restricted to the saturated zones and demonstrated that this approach outperformed the fully coupled strategy of Zhang et al. (2018). In this study, we develop a new weakly coupled DA scheme that introduces separate update restrictions for each observation type: GWL observations are used to update only saturated cells, while SM observations are used to update only unsaturated zones. This design minimizes potential spurious cross-variable correlations and enhances the robustness of multivariate assimilation. Additionally, updates are applied asynchronously to account for the different temporal dynamics of the variables: SM, which changes more rapidly, is typically updated daily, whereas groundwater, with slower dynamics, is updated weekly. Furthermore, unlike previous DA studies of TSMP, which generally used the same localization radius for joint GWL and SM assimilation, our approach applies different localization radii for the two variables, accounting for their distinct spatial correlation characteristics.”

## New references added in the revised manuscript:


Bayat, B., Camacho, F., Nickeson, J., Cosh, M., Bolten, J., Vereecken, H., and Montzka, C.: Toward operational validation systems for global satellite-based terrestrial essential climate variables, International Journal of Applied Earth Observation and Geoinformation, 95, 102240, 10.1016/j.jag.2020.102240, 2021.

Cooper, E., Blyth, E., Cooper, H., Ellis, R., Pinnington, E., and Dadson, S. J.: Using data assimilation to optimize pedotransfer functions using field-scale in situ soil moisture observations, Hydrology and Earth System Sciences, 25, 2445-2458, 10.5194/hess-25-2445-2021, 2021.

Dimitrova-Petrova, K., Rosolem, R., Soulsby, C., Wilkinson, M. E., Lilly, A., and Geris, J.: Combining static and portable Cosmic ray neutron sensor data to assess catchment scale heterogeneity in soil water storage and their integrated role in catchment runoff response, Journal of Hydrology, 601, 126659, 10.1016/j.jhydrol.2021.126659, 2021.

Fersch, B., Francke, T., Heistermann, M., Schrön, M., Döpper, V., Jakobi, J., Baroni, G., Blume, T., Bogena, H., Budach, C., Gränzig, T., Förster, M., Güntner, A., Hendricks Franssen, H. J., Kasner, M., Köhli, M., Kleinschmit, B., Kunstmann, H., Patil, A., Rasche, D., Scheiffele, L., Schmidt, U., Szulc-Seyfried, S., Weimar, J., Zacharias, S., Zreda, M., Heber, B., Kiese, R., Mares, V., Mollenhauer, H., Völksch, I., and Oswald, S.: A dense network of cosmic-ray neutron sensors for soil moisture observation in a highly instrumented pre-Alpine headwater catchment in Germany, Earth Syst. Sci. Data, 12, 2289-2309, 10.5194/essd-12-2289-2020, 2020.

Franz, T. E., Zreda, M., Rosolem, R., and Ferre, T. P. A.: A universal calibration function for determination of soil moisture with cosmic-ray neutrons, Hydrology and Earth System Sciences, 17, 453-460, 10.5194/hess-17-453-2013, 2013.

Patil, A., Fersch, B., Hendricks Franssen, H.-J., and Kunstmann, H.: Assimilation of Cosmogenic Neutron Counts for Improved Soil Moisture Prediction in a Distributed Land Surface Model, Frontiers in Water, 3, 10.3389/frwa.2021.729592, 2021.

Xue, Y., Houser, P. R., Maggioni, V., Mei, Y., Kumar, S. V., and Yoon, Y.: Evaluation of High Mountain Asia‐Land Data Assimilation System (Version 1) From 2003 to 2016, Part I: A Hyper‐Resolution Terrestrial Modeling System, Journal of Geophysical Research: Atmospheres, 126, 10.1029/2020jd034131, 2021.

Zafarmomen, N., Alizadeh, H., Bayat, M., Ehtiat, M., and Moradkhani, H.: Assimilation of Sentinel‐Based Leaf Area Index for Modeling Surface‐Ground Water Interactions in Irrigation Districts, Water Resources Research, 60, 10.1029/2023wr036080, 2024.