Responses to Reviewers' Comments for Manuscript egusphere-2025-6159

# Effectively Assimilate Satellite Land Surface Temperature into Offline Land Surface Models within Ensemble-based Assimilation Frameworks

Yunhao Fu, et al.

May 27, 2026

# Authors' Response to the Editor and Reviewers:

We sincerely thank the editor, the three referees, and the community reviewer for their constructive and detailed comments. We have revised the manuscript by integrating all comments into a single coherent version. The major revisions include clarifying the novelty of the joint-update LST assimilation scheme, expanding the methodological description of the updating and observation processing, correcting the localization formulation to the Gaspari–Cohn function, adding an experimental workflow figure, adding independent satellite and flux-tower validation, and strengthening the Discussion on parameter perturbations and comparison with previous studies. All responses below are organized in the original order of the Referee and Community comments. Bold Equations/Tables/Figures are from the revised manuscript; non-bold are part of this response.

# Referee Comment 1

**General Comments.**

This manuscript addresses a long-standing challenge in the field of Land Data Assimilation: how to effectively assimilate Land Surface Temperature (LST) into offline Land Surface Models (LSMs), given that LST varies drastically over time and is predominantly driven by downward atmospheric radiation forcing. Based on the Common Land Model (CoLM) and the Local Ensemble Transform Kalman Filter (LETKF) algorithm, the authors propose a novel scheme to implicitly and jointly update surface soil temperature and soil moisture by utilizing the cross-correlations among variables.

Overall, this research topic holds strong scientific value and practical operational significance. The manuscript is logically structured with a rigorous experimental design, effectively overcoming the limitations of single-variable updates in traditional observation operators. It provides valuable insights for researchers in meteorology, hydrology, and Earth system modeling. However, to further enhance the scientific rigor of this study, there is still room for optimization and expansion. Specifically, the authors need to further justify the physical rationality of key parameter settings within the assimilation framework (such as the large-scale localization strategy, parameter perturbation magnitudes, and observation representativeness errors). Additionally, it is essential to demonstrate the comparative advantages of the LST observation operator, supplement the evaluation with independent in-situ validation, and provide an in-depth mechanistic interpretation of the degraded assimilation performance in specific regions. In light of these considerations, I recommend that the manuscript be accepted after Minor Revisions.

**Response:**

We thank the reviewer for the positive assessment and constructive suggestions. Detailed responses are provided below.

**Comment 1**

1. Localization Strategy in Land DA

The study adopts a distance-dependent decay localization strategy, with the localization radius D set to decrease linearly from 730 km at the equator to 146 km at the poles. While such a large-scale spatial smoothing strategy is common in atmospheric assimilation systems, the land surface exhibits extreme spatial heterogeneity (strongly influenced by topography, vegetation cover, soil texture, etc.). Forcing covariance propagation across different underlying surfaces (e.g., boundaries between lakes and deserts, or transition zones of different vegetation types) is highly prone to introducing spurious correlations. The authors are advised to thoroughly discuss the physical rationality of employing such a large spatial localization radius (up to 730 km) in land assimilation, and clarify whether matching constraints for underlying surface characteristics were considered during the covariance propagation process.

**Response:**

The 730 km mentioned is the cutoff distance where the localization function drops to exactly zero [5]; the corresponding half-width is 365 km.

This localization scale is physically justified for two reasons:

1. It is considerably smaller than typical atmospheric system scales.

2. While larger than the minimum scale of land surface heterogeneity, it properly accounts for the spatial representation of aggregated satellite observations.

We have modified the text to add "cutoff" and explicitly clarified this rationale. Additionally, we noticed that the localization formula in the original manuscript was mistakenly written in the form of an exponential function instead of a Gaspari–Cohn function, which has now been corrected in the revised manuscript.

**Changes:** Section 2.3, L202-212 and **Equation 15** and **Equation 16**.

Comment 2

2. Model Parameter Perturbations

To maintain a reasonable ensemble spread, the authors introduced random perturbations to 14 tunable parameters in the CoLM model (see Table 1). Please provide further explanation regarding the physical justification for selecting these specific parameters and the rationality of setting their maximum perturbed magnitudes. Additionally, it is recommended to briefly discuss which of these 14 parameters contribute most significantly to the simulated variance of LST and the soil hydro-thermal cross-covariance.

**Response:**

**Table 1** in our manuscript summarizes the physical meanings for these 14 tunable parameters defined in the CoLM. For a more detailed explanation, please refer to the CoLM technical guide (`http://globalchange.bnu.edu.cn/download/doc/CoLM/CoLM_Technical_Guide.pdf`).

Regarding the perturbation magnitude, preliminary tests indicated that a 10% perturbation yielded an ensemble spread that was small relative to the observation error. We therefore increased the magnitude to 15% to generate a reasonable ensemble spread.

Owing to the highly nonlinear nature of land surface processes, any single one of these parameters can influence nearly all quantities in an LSM. Thus, it is not possible to isolate the sole contribution of each parameter. From a direct sensitivity perspective, LST is more sensitive to parameters related to surface energy balance and outgoing radiation (i.e., tuning factor, roughness lengths for soil and snow, drag coefficient, maximum transpiration, and critical temperature); and the hydro-thermal cross-covariance might be more directly governed by the parameters related to energy storage (i.e., maximum transpiration, critical temperature, wimp, irreducible water saturation of snow, restriction for minimum soil potential, pond depth, and fraction of model area with high water table).

**Changes:** Discussion, L574-579.

**Comment 3**

3. Lack of Independent In-situ Validation

The authors utilized ERA5-Land, GLDAS, and MERRA2 as evaluation benchmarks. Although these datasets offer the advantage of global coverage, they are inherently products driven by specific model parameterization schemes and atmospheric forcings, inevitably containing their own systematic biases. Please explain why independent in-situ observation data were not utilized for validation. If conditions permit, it is strongly recommended to supplement the study with site-level validation in representative regions to further enhance the objectivity and persuasiveness of the assimilation performance evaluation.

**Response:**

We have introduced independent satellite and in-situ observations into the revised manuscript to robustly evaluate the assimilation performance. These consist of datasets retrieved from multiple satellites, including ATSR-2 [11], AVHRR [7], and MODIS [8], as well as in-situ observations from the FLUXNET2015 [10] and AmeriFlux [9] networks.

**Changes:** Section 3.1.3, L282-286; Section 3.3, L330-332; Section 4.2.2, L433-440 and **Figure 8**; Section 4.2.3, L530-537 and **Figure 14**.

**Comment 4**

4. Observation Error Representation

Given the extremely high spatial resolution of MODIS LST, the authors used a simple area-weighted processing method and excluded cloud-contaminated pixels when aggregating it to the 0.5-degree model grids. It is recommended to further clarify how the observation error covariance matrix was explicitly specified in the LETKF update. When aggregating high-resolution observational data, did the observation error account not only for the retrieval error of MODIS itself but also reasonably incorporate the "representativeness error" caused by scale mismatch? This is crucial for the appropriate allocation of weights in the final analysis increments.

**Response:**

As in other data assimilation studies, the observation errors are assumed to be independent, so the observation covariance is diagonal. The error variance ($\sigma^2$) for each grid cell is a direct aggregation of the MODIS `lst_uncertainty` field, following the same quality control and aggregation procedures applied to the LST data.

We did not explicitly introduce a separate representativeness error term for scale mismatch. Preliminary tests indicated that the standard error of 0.01° LST with respect to the aggregated LST in a box of 0.5° × 0.5° is less than 5% of the aggregated retrieval uncertainty. This demonstrates that the representativeness error is negligible compared to the retrieval error.

**Changes:** Section 3.1.2, L245-248.

**Comment 5**

5. LST Operator

The accuracy of the observation operator itself plays a decisive role in the effectiveness of the assimilation. Since the method for calculating LST varies across different models, please explain the method used to calculate LST in CoLM and elaborate on its advantages or differences compared to observation operators in other models. This will help demonstrate the rationality of the adopted methodology.

**Response:**

The simplest way to calculate radiometric LST is based on the net longwave radiation using the Stefan-Boltzmann relationship, $LST_{sim} = (\frac{LW}{\sigma})^{\frac{1}{4}}$. In CoLM, $LW$ is constructed by combining thermal emissions from the sunlit canopy, shaded canopy, and the ground, while explicitly accounting for vegetation cover, canopy gap fraction, and surface emissivity. **Equation 2** is the formulation that the CoLM utilizes to calculate the LST, representing the realistic LST calculation within CoLM's two-big-leaf framework [4].

**Changes:** Section 2.2, L127-130.

**Comment 6**

6. Degradation in Specific Regions

Figures 8 and 9 show that the simulation performance deteriorated in certain regions after LST assimilation (indicated by the red areas). Please provide an in-depth analysis of the physical mechanisms or algorithmic reasons leading to the degradation in these specific areas. Is this negative impact constrained by special local underlying surface conditions, limitations in the observation error settings, or a failure of the cross-correlations among physical variables? Please discuss whether this phenomenon is reasonable within the current assimilation framework.

**Response:**

Degradation in soil water in western North America and northern Asia largely occurs in late spring or late autumn when freezing and thawing happen frequently. Assimilating the LST may alter the timing of thaw or freeze, which in turn causes the soil water to be discrepant from reference datasets.

A lake-land breeze-induced secondary circulation acts as a negative feedback mechanism that reduces the lake-land thermal contrast. However, in an offline LSM, the lack of this negative feedback can result in the degradation around the Great Lakes (largest in summer) because the assimilation updates only land grids while excluding water bodies.

The mechanism for northern Australia is already detailed in the manuscript.

**Changes:** Section 4.2.2, L481-485.

## Referee Comment 2

**General Comments.**

Reviewing of the manuscript "Effectively Assimilate Satellite Land Surface Temperature into Offline Land Surface Models within Ensemble-based Assimilation Frameworks" by Y. Fu et al. submitted to Geoscientific Model Development (Manuscript Number: eguusphere-2025-6159). This work aims to improve the offline Land Surface Model by data assimilation techniques to assimilate land surface temperature, updates the soil temperature and soil moister simultaneously. This process influences significantly the land-atmosphere interactions and the soil temperature in different layers. The following is my comments in details:

**Response:**

We thank the reviewer for the careful reading and constructive comments. Detailed responses to each comment are provided below.

**Comment 1**

1. The manuscript uses a lot of abbreviations, which is not necessary and influence the readability of this work. For example, the Data Assimilation (DA), Common Land Model (CoLM), Extended Kalman Filter (EKF). There are a lot of more examples than these, and I will not list them all. The authors please reduce some unnecessary abbreviations to improve the fluency and readability of this manuscript.

**Response:**

We have carefully reviewed the manuscript and removed several unnecessary abbreviations that appeared infrequently, including GCOS, ECV, EKF, EnSRF, EAKF, LEKF, WFD, CHTESSEL, and EH. Meanwhile, abbreviations such as CoLM (Common Land Model), DA (data assimilation), LST, and LSM have been retained to maintain conciseness throughout the manuscript.

**Changes:** Introduction, L21 and L79; Section 2.3, L139, L141 and L197; Section 3.1.1, L218; Section 3.1.3, L263; Caption for **Figure 3**; Section 4.1, L348-350.

**Comment 2**

2. The subtitles are overly simple. For example, Section 2.1 CoLM, 2.2 Observation operator, 2.3 LETKF. Again, there are more examples than these, and it is the responsibility of the authors to read through the entire manuscript to improve them.

**Response:**

We have carefully reviewed the entire manuscript and revised all overly simple subtitles to provide clearer context for each section. Specifically, we have made the following updates:

Section 2.1 “CoLM” is now “CoLM Configuration and Ensemble Generation”.

Section 2.2 “Observation operator” is now “Land Surface Temperature Observation Operator”.

Section 2.3 “LETKF” is now “The Local Ensemble Transform Kalman Filter Scheme”.

Section 4.3.1 “Temperature” is now “Impacts on Soil and Snow Temperature”.

Section 4.3.2 “Water content” is now “Improvements in Snow Properties and Soil Moisture Content”.

Section 4.3.3 “Evaporation and heat flux” is now “Modulations of Evaporation and Surface Heat Fluxes”.

**Changes:** Section 2.1, L93; Section 2.2, L121; Section 2.3, L135; Section 4.2.1, L356; Section 4.2.2, L409; Section 4.2.3, L499.

## Comment 3

3. Error or mistakes in writing. For example, Page 1 Line 22, delete the comma in “...... (, GCOS).” Page 5 Lines 9-12, add “is” in each expression following the variables. BTW, it is a little bit awkward to use down arrow following LW, please consider revising it.

**Response:**

Thank you for your careful reading. We have addressed all of your comments in the revised manuscript, as detailed below:

1. We have corrected the citation formatting by removing the misplaced comma in “..., (, GCOS)”.
2. We have added the verb “is” after each variable in the definitions.
3. We have replaced the downward arrow following “LW” with “DLW”.

**Changes:** Introduction, L22-23; Section 2.2, L124-127 and **Equation 2**.

## Comment 4

4. Add a full flowchart of this model coupling work including the data assimilation processes. It is not necessary to list the detailed equations but show the concepts of this entire process.

**Response:**

We have added a comprehensive flowchart illustrating the CoLM-LETKF land data assimilation workflow, which includes a 52-year spin-up, ensemble initialization, and 3-hour forecast cycles with parameterized perturbations. Alongside the DA experiment, an open-loop experiment (without data assimilation) was conducted to serve as a baseline. For the sake of completeness, the equations are retained in the text.

**Changes:** Section 3.2, L289-290 and **Figure 2**.

**Comment 5**

5. 1.3 Evaluation datasets. The writing style of this Section reads a little bit like AI-generated. For example, the point "-ERA5-Land" or "-GLDS" or "-MERRA2" followed by a paragraph". Please consider revising and re-writing for fluency.

**Response:**

We have rewritten Section 3.1.3 to improve its narrative fluency.

**Changes:** Section 3.1.3, L259-276.

**Comment 6**

6. The format of this work should be consistent throughout the whole manuscript. For example, leave space when starting a new paragraph in the front.

**Response:**

We appreciate the reviewer's attention to the formatting details of the manuscript. We have thoroughly checked the text to ensure consistency. The absence of indentation at the beginning of sections is due to the default settings of the official GMD LaTeX template (`https://www.geoscientific-model-development.net/Copernicus_LaTeX_Package.zip`), which we used for this submission. In this template, the first paragraph of each section is not indented, whereas all subsequent paragraphs are properly indented. We have strictly adhered to this official formatting standard throughout the revision.

**Comment 7**

7. Section 4.1 for Equation (19), this is the skill metrics and should be removed to the Method Section. Please double check the definitions of RMSD and unbiased RMSD, if the mathematical formulae are correct or not.

**Response:**

Regarding the placement of the "Evaluation Metrics" subsection (including **Equation 20**), we have moved it to a position between the "Experiments" (Section 3.2) and "Results" (Section 4) sections. We chose this specific location because the application of these metrics—and the subsequently presented Table—logically builds upon the experimental setup and the validation datasets introduced immediately prior.

We have also double-checked the formulations for both RMSD and unbiased RMSD. The equations presented are fully consistent with the calculation methods defined in Appendix B of [3].

**Changes:** Section 3.3.

**Comment 8**

8. The Table 2 is not necessary, please delete it.

**Response:**

This table has been removed from the revised manuscript, as the experimental configurations are already fully described in the text.

**Changes:** Section 3.2.

**Comment 9**

9. When comparison for AEXP or OEXP driven by ERA5 or GLDAS or MERRA2, what the observational data are used or is there any in-situ data to evaluate the model performance? This is an important issue, since the comparison and validation are throughout the whole manuscript.

**Response:**

We have introduced independent satellite and in-situ observations into the revised manuscript to robustly evaluate the assimilation performance. These consist of datasets retrieved from multiple satellites, including ATSR-2 [11], AVHRR [7], and MODIS [8], as well as in-situ observations from the FLUXNET2015 [10] and AmeriFlux [9] networks.

**Changes:** Section 3.1.3, L282-286; Section 3.3, L330-332; Section 4.2.2, L433-440 and **Figure 8**; Section 4.2.3, L530-537 and **Figure 14**.

**Comment 10**

10. Figure 4: The use of “MAM” and “SON”. What do these abbreviations mean? Please be clear.

**Response:**

We have clarified these abbreviations in the revised manuscript. 'MAM' and 'SON' stand for the months March-April-May and September-October-November, respectively.

**Changes:** Section 4.2.1, **Figure 6**; Section 4.2.2, **Figure 10**.

### Comment 11

11. Figure 10: why the evaporation from the bare soil is missing for AEXP-OEXP vs. GLDAS?

**Response:**

The evaporation from bare soil compared based on GLDAS is presented in **Figure 13**(c).

### Comment 12

12. This work needs substantial improvements for the Discussion, e.g., 1. Compare the work/accuracy of land surface temperature, soil moister, water content, soil humidity simulated by model compared with previous or similar studies? The physical mechanisms behind the parameter tuning strategy needs to substantially improve.

**Response:**

Regarding the first point, we have expanded the Discussion section to compare our assimilation performance with relevant previous studies. In general, the degree of improvement observed in these previous studies is very similar to what we achieved in our current work. For instance, Huang, Li, and Lu [6] assimilated LST solely to update soil temperature; thus, their analysis concentrated primarily on soil temperature improvements, reporting an error reduction of approximately 1 K for the soil profile at specific sites. Their findings are highly consistent with our results, where the bias in soil temperature over Northeast Asia is reduced by approximately 1.0 K, 1.5 K, and 2.0 K with increasing depth. Furthermore, Chen et al. [2] assimilated LST to update soil temperature alongside brightness temperature to update soil moisture, focusing mainly on the Tibetan Plateau. They reported a reduction in the RMSD for shallow-layer soil moisture ranging from 19% to 71% across various sites. In our study, the proposed LST assimilation achieves a comparable impact on soil moisture, reducing the unbiased RMSD for the upper soil layer by approximately 40% across Northern Asia.

For the second issue, the tuning parameters encompass six key physical processes, which are distinguished by different colors in Table 1: surface turbulent exchange (blue); canopy hydrology (cyan); runoff generation (orange); soil thermal dynamics (red); snow hydrology (purple); and soil water dynamics (magenta). The Crank-Nicholson factor is explicitly excluded ($\mathbf{\Delta} = 0$) as it is a numerical scheme parameter rather than a physical one. Please note that aside from the added color coding, the content of Table 1 is identical to the original manuscript, provided here to facilitate your review.

Table 1: Tunable parameters, their default values, and maximum perturbed magnitudes in CoLM. The parameters are color-coded based on the six key physical processes they regulate: surface turbulent exchange, canopy hydrology, runoff generation, soil thermal dynamics, snow hydrology, and soil water dynamics. The numerical parameter (black) is excluded from perturbation.

| **Tunable parameter** | **Default ($\lambda$)** | **Max Perturb. ($\Delta$)** |
|---|---|---|
| Roughness length for soil [m] | 0.01 | 0.0015 |
| Roughness length for snow [m] | 0.0024 | 0.00036 |
| Drag coefficient for soil under a canopy [-] | 0.004 | 0.0006 |
| Maximum dew [mm] | 0.1 | 0.015 |
| Fraction of model area with high water table | 0.3 | 0.045 |
| Tuning factor (turn first layer T into surface T) | 0.34 | 0.051 |
| Crank-Nicolson factor (between 0 and 1) | 0.5 | 0.0 |
| Irreducible water saturation of snow | 0.033 | 0.00495 |
| Water impermeable if porosity less than wimp | 0.05 | 0.0075 |
| Pond depth [mm] | 10.0 | 1.5 |
| Wilting point potential [mm] | -1.5E5 | 2.25E4 |
| Restriction for minimum of soil potential [mm] | -1.0E8 | 1.5E7 |
| Maximum transpiration for moist soil+100% veg [mm/s] | 0.0002 | 0.00003 |
| Critical temperature [K] (to determine rain or snow) | 0.0 | 0.15 |

**Changes:** Discussion, L554-560 and L574-579.

# Referee Comment 3

**General Comments.**

The authors proposed an LST assimilation scheme that jointly update soil moisture and soil temperature in the upper soil layers of CoLM. Global experiments show notable improvements in several land surface variables, including soil temperature, snow temperature, snow depth, soil moisture, and surface fluxes. The main contribution is to demonstrate that LST assimilation can be more effective when its information is transferred into coupled water-energy state variables, particularly in freeze-thaw regions and humid areas where vertical water and heat exchanges help preserve and propagate the assimilation signal. The paper is generally well structured and clearly written. However, I have several concerns and suggestions that should be addressed before publication. I therefore recommend MAJOR revision before the manuscript can be considered for publication in GMD.

**Response:**

We thank the reviewer for the detailed and constructive review. Detailed responses are provided below.

Comment 1

**1. Novelty and positioning of the proposed scheme**

The novelty of the manuscript should be clarified more explicitly. LETKF and LST assimilation have been widely explored in previous studies, so the main contribution appears to be the joint update of soil temperature and soil moisture. However, the mechanism by which this joint update improves snow-related variables in freeze-thaw regions is not sufficiently demonstrated. For example, an overestimated LST could be corrected either by reducing soil temperature or by increasing soil moisture, and these two pathways may have different or even competing effects on the surface water-energy balance. The authors should explain how the LETKF balances these increments and why this leads to better snow temperature, snow cover, and snow depth. Moreover, if the proposed mechanism is mainly related to freeze-thaw processes, it is unclear why the strongest soil moisture improvements occur mainly in humid and relatively warm regions rather than in the same freeze-thaw regions. I therefore suggest adding sensitivity experiments that update only soil temperature and only soil moisture, respectively, to demonstrate the added value of the proposed joint-update scheme.

**Response:**

Assimilating LST to update only the surface soil temperature yields marginal improvements, as the surface state rapidly relaxes back to the trajectory dictated by atmospheric forcing. Motivated by this finding, we propose to jointly update soil temperature (ST) and moisture (SM) in the top two layers.

The proposed scheme leverages the persistent memory of soil moisture and the thermal inertia of soil temperature to prolong the assimilation effect.

Data assimilation inherently distributes the observation innovation to state variables based on the background error covariance rather than physical constraints. This covariance reflects the physical sensitivities and the strength of energy-water coupling within the land surface model. The joint update precisely exploits the covariance between soil temperature and moisture.

This joint-update scheme is proposed to sustain the impact of assimilation, yet it also yields substantial improvements in snow-related variables. As shown in Fig. 1, the joint-update scheme (ST+SM) significantly outperforms the temperature-only update (ST) in capturing snow cover using ERA5-Land. Updating soil moisture modulates the heat capacity and thermal conductivity of the underlying soil. This physically consistent adjustment to the soil's thermal properties regulates the snow energy budget far more effectively than updating temperature alone.

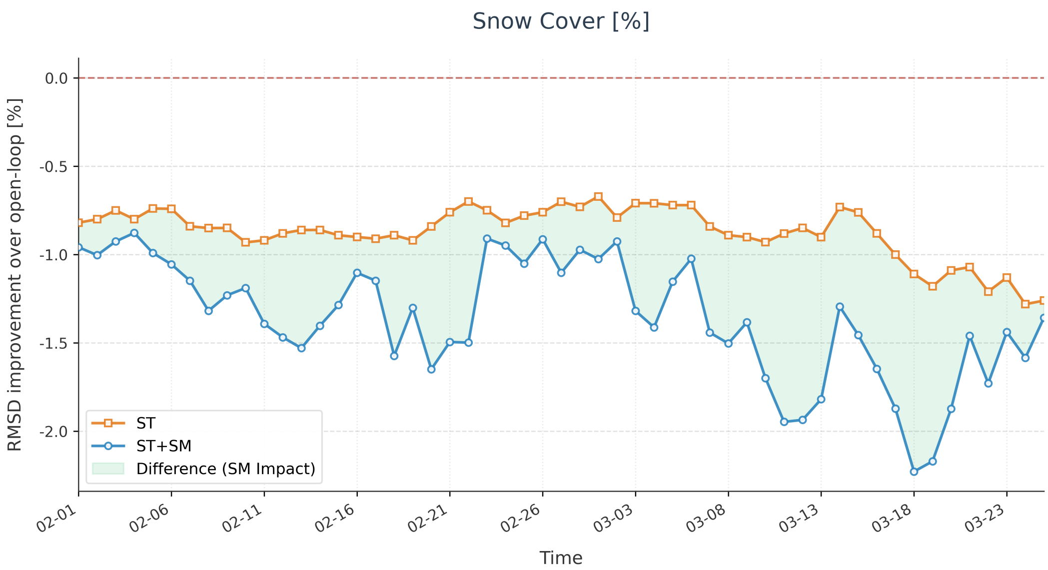


Figure 1: Snow cover $\mathrm{RMSD_d}$ [%] for the joint-update (ST+SM) and temperature-only (ST) assimilation schemes relative to the open-loop experiment.

**Figure 1** in the revised manuscript compares the joint-update scheme and the temperature-only scheme against in-situ observations from the AmeriFlux networks [9], demonstrating that the proposed scheme consistently improves both water and energy states. We have clarified this novelty and the underlying mechanisms in the Introduction section of the revised manuscript.

**Changes:** Introduction, L62-76 and **Figure 1**.

Comment 2

**2. Some questions on your methods**

Several methodological choices are not sufficiently clear, which makes the framework difficult to reproduce and the claimed innovation difficult to assess.

1) The authors generate ensemble spread mainly by perturbing model parameters, rather than atmospheric forcing. This choice needs justification, especially because parameter perturbations may affect water and energy conservation. The authors should clarify whether conservation is maintained or diagnosed. If not, the assimilation experiment should be compared with an ensemble-mean open-loop experiment, rather than a single deterministic open-loop run, to separate the impact of ensemble generation from the impact of LST assimilation.

**Response:**

Fig. 2 presents two-month (from January 1 to March 1, 2001) single-member experiments with and without perturbing tunable parameters. The temporal evolution of the global mean energy and water balance error is nearly identical between the two experiments. This demonstrates that perturbing tunable parameters does not compromise the model's intrinsic physical balance or conservation properties, thereby justifying the experimental design adopted in this study.

**Changes:** Section 2.1, L118-120.

Comment 3

2) The updated state vector should be defined more explicitly. What specific soil layers are included in the "upper soil layers"? Are soil temperature, liquid water, and ice content updated simultaneously? Are snow-layer temperature and water/ice content also updated, or only soil variables?

**Response:**

The updated state variables are defined in Section 2.1. The "upper soil layers" specifically refer to the top two valid layers ($l_{\text{top}} \leq l \leq l_{\text{top}} + 1$). Temperature, liquid water, and solid water/ice content within these valid layers are updated simultaneously. Furthermore, when snow is present, the snow temperature and water/ice content are also updated in the same manner.

**Changes:** Section 2.1, L109-110.

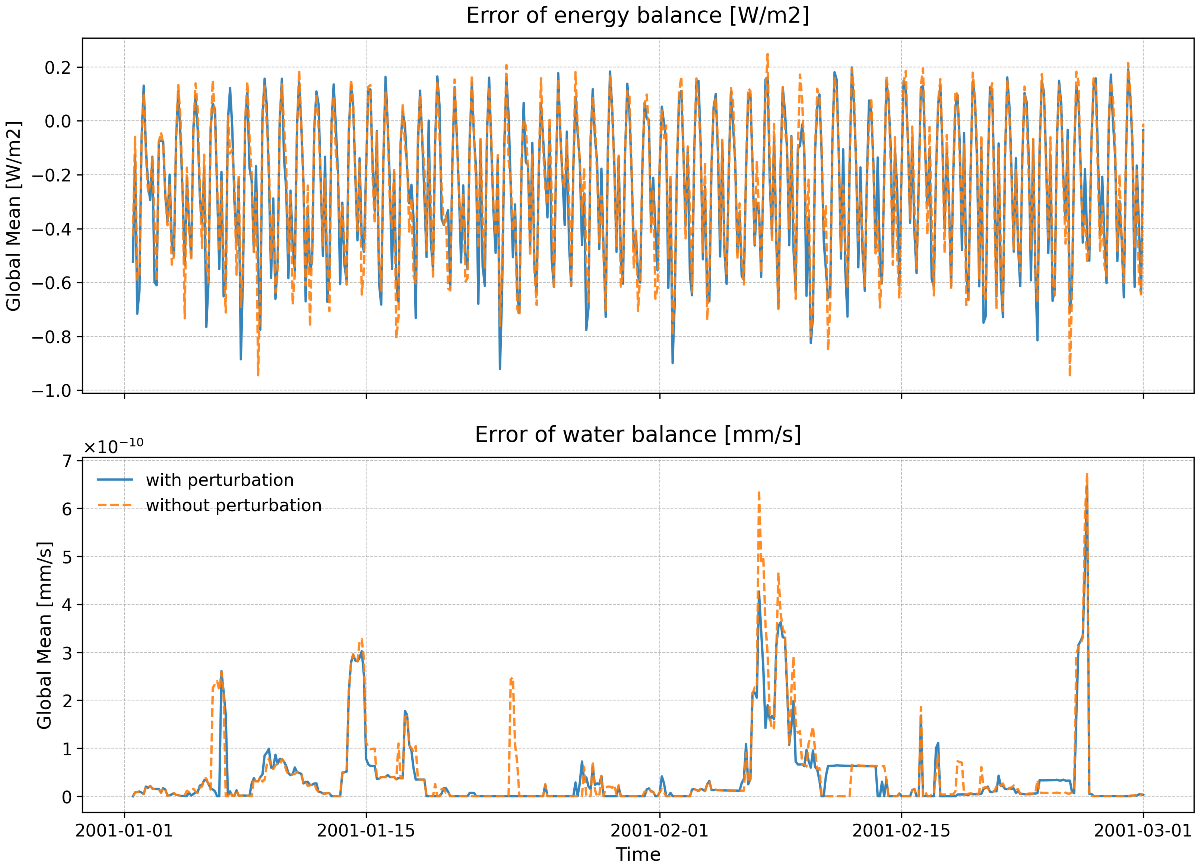


Figure 2: Temporal evolution of the global mean error of energy balance (top panel) and water balance (bottom panel) for simulations with (solid blue line) and without (dashed orange line) perturbations of tunable parameters from January 1 to March 1, 2001.

### Comment 4

3) Line 210: Where do the observation error values come from? Do they include only MODIS retrieval error, or also representativeness error from aggregating observations to the 0.5° grid? The latitude-dependent weighting/threshold should also be justified. the specification of the observation error needs to be clarified.

**Response:**

As in other typical data assimilation studies, the observation errors are assumed to be independent, so the observation covariance matrix is diagonal. The observation errors are directly derived from the MODIS lst_uncertainty field and aggregated to the 0.5° model grid. We did not explicitly introduce the representativeness error, as preliminary tests indicated that the standard deviation of the raw 0.01° LST within a 0.5° grid cell is less than 5% of the aggregated retrieval uncertainty, making the representativeness error negligible.

We use latitude-dependent weighting to account for the varying physical area of the raw MODIS pixels at different latitudes during aggregation, which ensures accurate spatial representativeness. Finally, the threshold of 1500 valid pixels (60% coverage) is maintained to guarantee statistical reliability.

**Changes:** Section 3.1.2, L241-248.

**Comment 5**

4) Several key thresholds and parameters require justification or sensitivity tests, including the localization radius and its latitude dependence, the 8 K innovation rejection threshold, and the 3-standard-deviation limit on increments. These choices can strongly affect the analysis increments and the number of assimilated observations.

**Response:**

The localization radius uses a cutoff distance of 730 km at the equator. This radius is smaller than typical atmospheric system scales, and while larger than the minimum scale of land surface heterogeneity, it properly accounts for the spatial representation of aggregated satellite observations. The latitude dependence is simply implemented to avoid unreasonable analysis increments at high latitudes. Regarding the 8 K innovation rejection threshold, our sensitivity tests indicated that tighter thresholds (5 K) overly reject valid observations driven by the LST diurnal cycle, while looser thresholds (9 K) fail to filter gross errors. The 8 K threshold empirically provides a good balance. Finally, the 3-standard-deviation limit is a standard measure to clip the outlier analysis.

**Changes:** Section 2.3, L208-212; Section 3.1.2, L254-255; Section 3.2, L302.

**Comment 6**

5) The temporal treatment of MODIS LST observations is unclear. When assimilating every 3 hours, are all observations within the 3-hour window used at the analysis time, or only observations valid exactly at that time?

**Response:**

For a specific analysis time $t$, all observations whose actual scan times fall within the $[t-1.5, t+1.5]$ hours window are aggregated and assimilated.

**Changes:** Section 3.1.2, L232-235.

**Comment 7**

6) The 52-year spin-up and the "time-lag method" should be described more clearly, especially how the initial ensemble is generated.

**Response:**

To generate the initial ensemble, we first conducted a 52-year spin-up simulation to allow the model to reach a stable equilibrium state. Following this spin-up, the model was run for an additional full year with three-hourly outputs, from which 50 initial members were randomly sampled. This time-lag approach helps the selected initial members maintain the intrinsic energy and water balance of the model while providing a reasonable climatological spread of land surface states.

**Changes:** Section 3.2, L295-298.

**Comment 8**

7) The authors should explain why different variables are evaluated with different metrics and reference datasets, and whether this affects the consistency of the conclusions.

**Response:**

We have clarified the rationale for selecting specific evaluation metrics and reference datasets in the revised manuscript. The metrics were chosen based on the typical error structures of the variables. Soil moisture evaluation uses unbiased RMSD to remove systematic biases caused by differing soil layer definitions across products, thereby robustly assessing dynamic variations. Conversely, the other variables (temperatures, snow depth, and fluxes) rely on standard RMSD and BIAS to quantify absolute accuracy. The selection of reference datasets depends on data availability and regional reliability. Specifically, GLDAS lacks snow surface temperature and snow cover fraction data, while MERRA2 lacks bare soil evaporation data. Furthermore, because ERA5-Land soil temperature in permafrost regions of Asia is known to exhibit large biases [1], we excluded it to ensure evaluation accuracy.

**Changes:** Section 3.3, L320-323 and L330-332.

**Comment 9**

8) How to process imbalance of water and energy during DA?

**Response:**

Because data assimilation alone lacks explicit physical constraints, we implemented specific processing steps to mitigate potential water and energy imbalances. Specifically, snow mass is updated only when all ensemble members share the same number of snow layers. In addition, a non-negative limit is systematically applied to soil moisture. Furthermore, when the top two soil layers are updated, if the top layer contains solid ice while the second layer has none, a small amount of ice is redistributed from the first layer to the second to ensure stable model integration.

**Changes:** Section 3.2, L306-310.

Comment 10

**3. Need for more independent validation**

The evaluation relies mainly on ERA5-Land, GLDAS, and MERRA2. Although using multiple products is helpful, these datasets are still model-based and are strongly affected by their own forcing, parameterizations, and bias corrections. They do not provide fully independent observations of land surface states. Therefore, I suggest adding more independent validation, such as soil moisture/surface flux from FLUXNET or other in-situ networks, and/or satellite-based products (e.g., MODIS) for relevant land surface variables. This would make the reported improvements more convincing, especially for soil moisture, snow variables, and surface fluxes.

**Response:**

We have introduced independent satellite and in-situ observations into the revised manuscript to robustly evaluate the assimilation performance. These consist of datasets retrieved from multiple satellites, including ATSR-2 [11], AVHRR [7], and MODIS [8], as well as in-situ observations from the FLUXNET2015 [10] and AmeriFlux [9] networks.

**Changes:** Section 3.1.3, L282-286; Section 3.3, L330-332; Section 4.2.2, L433-440 and **Figure 8**; Section 4.2.3, L530-537 and **Figure 14**.

Comment 11

**Specific comments**

Line 75: I do not agree with this statement. LETKF is also not the core contribution of this study, so I suggest not highlighting it here.

**Response:**

We've removed this sentence.

**Changes:** Introduction, L83-84.

### Comment 12

Line 85: Please specify the version of CoLM used in this study, e.g., CoLM 2014 or CoLM 2024.

**Response:**

The version of CoLM used in this study is based on CoLM 2014 and is operationally employed as the land surface model within the GRAPES (Global/Regional Assimilation and Prediction Enhanced System) global model at the China Meteorological Administration. This information has been specified in the revised manuscript.

**Changes:** Section 2.1, L98-100.

### Comment 13

Section 2.2: This section mainly describes a physical process already implemented in CoLM and does not appear to be a methodological contribution of this study. If retained, it should be reorganized to clearly explain how soil moisture and soil temperature are linked to LST in the proposed assimilation framework; otherwise, it may not be necessary.

**Response:**

We clarify that the observation operator is not claimed as a methodological contribution in this study. However, providing a clear description of the observation operator is a standard and necessary component in data assimilation literature.

Furthermore, this section demonstrates that LST is directly related to soil temperature. The connection between LST and soil moisture is bridged by the temperature-moisture covariance. This cross-variable update mechanism is detailed in Section 2.3. To make this clearer, we have added a sentence at the end of this section pointing to Section 2.3, where this cross-variable update mechanism is explicitly detailed.

**Changes:** Section 2.2, L134.

### Comment 14

Section 4.2: The improvement in LST is very small. This may be related to the choice of perturbing model parameters rather than atmospheric forcing. Please discuss this possible limitation.

**Response:**

We agree that the limited improvement in LST is closely related to the use of deterministic atmospheric forcing. The central goal of our study is to explore how to effectively assimilate LST within an offline framework, even under such forcing conditions. We generated background uncertainty through initial state sampling and parameter perturbations rather than forcing perturbations. Because LST is a fast-varying variable highly sensitive to external atmospheric drivers, the unperturbed forcing strongly constrains the surface energy balance. Consequently, even after LST is corrected at the assimilation step, it quickly relaxes back toward the trajectory dictated by the fixed forcing. Nevertheless, as demonstrated by other results in the manuscript, the assimilation of LST effectively propagates its impact to deeper soil layers through adjustments in soil temperature and moisture. To address the forcing-related limitation, we plan to use a land-atmosphere coupled model in future work to account for atmospheric feedbacks, as noted in the original manuscript.

**Changes:** Section 4.1, L344-347.

### Comment 15

Line 290: Please clarify how "error" is defined here.

**Response:**

"Error" refers to the observation error, which specifically represents the inherent uncertainty associated with the observational measurements. We have clarified this in the revised text.

**Changes:** Section 4.1, L351.

### Comment 16

Line 296: "temperature" should be changed to "soil temperature".

**Response:**

We revised the subsection title to "Impacts on Soil and Snow Temperature" to make the evaluated variables explicit.

**Changes:** Section 4.2.1, L356.

### Comment 17

Figure 3: The open-loop experiment appears to perform better in winter. Please explain why this occurs.

**Response:**

In winter, the open-loop experiment exhibits a systematic negative bias in soil temperature. However, assimilating LST alters the snow state, effectively converting "old snow" into "new snow". The enhanced insulation of the new snow then traps excessive heat within the soil layers, shifting the soil temperature toward a positive bias. This reversal from negative to positive bias ultimately leads to the higher RMSD observed in winter.

### Comment 18

Figure 4: In February, performance degrades in many regions. Please explain the possible reasons for these degradations.

**Response:**

The degradation observed in February is physically linked to the soil temperature biases discussed earlier. Since the atmospheric forcing is prescribed, the thermal exchange at the snow-atmosphere interface is effectively constrained. At the same time, the underlying soil—now anomalously warm due to the enhanced insulation of the updated snowpack—drives a stronger upward ground heat flux into the snow layer. This persistent upward energy transfer artificially elevates the snow temperature, leading to the degradation in RMSD during peak winter conditions.

## Community Comment 1

**General Comments.** This paper proposes a new way to assimilate satellite land surface temperature, specifically MODIS LST, into an offlineland surface model. The core idea is that LST should not update only temperature-related states, but should jointly update upper-layer soil temperature and soil moisture/water phase, using the cross-covariances inside an ensemble Kalman filtering framework. The authors implement this in CoLM with LETKF, apparently for the first time in this model, and run a global 0.5° experiment for 2001 with a 2-month free forecast afterward. The scientific motivation is reasonable: in offline LSMs, LST is hard to assimilate effectively because it changes rapidly and is strongly controlled by atmospheric forcing, so direct gains in LST itself tend to be short-lived. The manuscript argues that by storing the analysis signal in both energy and water states, the effect can persist longer and propagate vertically into deeper soil and snow states. Potentially worth publishing in HESS.

**Response:**

We thank the community comment for the constructive suggestions. Detailed responses are provided below.

**Comment 1**

The manuscript would benefit from a clearer statement of the exact novelty relative to Huang et al. (2008) and Chen et al. (2021). At present, the novelty is there, but it is somewhat buried in the introduction.

**Response:**

The core distinction lies in the assimilation strategy. Our study proposes a joint assimilation framework that uses a single observation—Land Surface Temperature (LST)—to simultaneously update both surface soil temperature and soil moisture. By contrast, Huang, Li, and Lu [6] assimilated LST to update only the soil temperature profile. Meanwhile, Chen et al. [2] assimilated LST for soil temperature and brightness temperature (BT) for soil moisture, but these act as independent updates rather than a joint assimilation.

**Changes:** Introduction, L65-68.

**Comment 2**

Please clarify whether the 3-hour assimilation frequency is dictated by model cycling convenience or by the temporal handling of MODIS Terra observations, which are not naturally available every 3 hours globally. The observation-time treatment needs to be easier to follow.

**Response:**

The 3-hour assimilation frequency represents a necessary balance between the physical characteristics of LST and model dynamics. Because LST is fast-varying, a narrow assimilation window is required to accurately capture its temporal changes. Conversely, the window cannot be too short, as the model requires sufficient spin-up time to dynamically adjust after each assimilation step. Therefore, a 3-hour cycle optimally satisfies both constraints.

For each 3-hour data assimilation cycle centered at 00:00, 03:00, ..., 21:00 UTC, only the satellite pixels whose actual scan times fall within this window are extracted and assimilated at that specific step.

**Changes:** Section 3.1.2, L232-235.

**Comment 3**

The threshold requiring at least 1500 raw MODIS observations per 0.5° grid cell seems high and may create regionally uneven sampling. Please justify this choice and show how much spatial coverage is lost because of it.

**Response:**

The threshold of 1500 valid pixels (representing approximately 60% coverage) was chosen to ensure both spatial representativeness and high statistical confidence. A 0.5° grid cell in the CoLM model typically encompasses highly heterogeneous land cover and topography. Aggregating heavily clouded grids using an insufficient number of clear-sky pixels not only lacks statistical reliability but also inevitably introduces severe cloud-edge biases. Assimilating such unrepresentative observations would actively force the land surface model into an unrealistic physical state. Therefore, preserving physical consistency and statistical robustness must be prioritized over absolute spatial completeness.

To explicitly address the concern regarding coverage loss, we quantified the spatial gaps over a continuous one-month period (February 2001). We found that relaxing the threshold from 1500 down to 1000 valid pixels only marginally improves the spatial coverage of the observed grids by approximately 8.8%. This demonstrates that the data gaps are primarily driven by the physical reality of persistent cloud cover rather than the stringency of our algorithmic threshold. Consequently, compromising the representativeness and statistical standards would yield very limited gain in coverage while risking the introduction of significant assimilation errors.

**Changes:** Section 3.1.2, L244.

**Comment 4**

The innovation rejection threshold of 8 K should be justified with either prior literature or a sensitivity test.

**Response:**

We evaluated multiple threshold values (5 K to 9 K) to examine their impact on filtering the Observation-Minus-Background (OMB) innovations. It is worth noting that our assimilation window spans $\pm 1.5$ hours. Given the pronounced diurnal cycle of land surface temperature (LST), this temporal span inherently introduces relatively large, yet physically valid, OMB deviations. Consequently, our tests indicated that a tighter threshold (e.g., 5 K) aggressively rejected a large portion of these valid observations, while a looser threshold (e.g., 9 K or higher) failed to adequately filter out gross errors and unrepresentative extremes. The 8 K threshold was empirically found to provide a good balance for the observation quality control in this study, effectively removing gross errors while safely retaining sufficient high-quality observations.

**Changes:** Section 3.1.2, L254-255.

**Comment 5**

The paper should report ensemble-spread diagnostics more explicitly, since ensemble collapse/spread deficiency is discussed as a central issue.

**Response:**

To address this concern, we have added ensemble-spread diagnostics to the Discussion section. Taking the Northern Hemisphere as an example, the temperature spread in the first soil layer ranges from 0.3 K to 1.4 K. This relatively low spread is expected due to strong atmospheric forcing, which further justifies our approach of using LST to update both soil temperature and moisture. For soil moisture, the spread increases from approximately 0.5 kg m$^{-2}$ in the top layer to 10.0 kg m$^{-2}$ in the deepest layer, exhibiting depth-dependent and seasonal characteristics. Finally, as highlighted in the Conclusion, maintaining a reasonable ensemble spread in offline land surface models is fundamentally limited by atmospheric forcing. Overcoming this limitation requires a two-way coupled land-atmosphere assimilation framework.

**Changes:** Discussion, L566-571.

Comment 6

Please clarify whether perturbing the 14 tunable parameters is time-invariant per member or re-sampled through time. This matters for interpretation of model-error representation.

**Response:**

The 14 tunable parameters are re-sampled at the beginning of each 3-hour assimilation cycle for each ensemble member rather than being time-invariant. We have added this clarification to ensure the model-error representation is accurately interpreted.

**Changes:** Section 2.1, L117-118.

Comment 7

I do strongly recommend that the authors cite Zafarmomen et al. (2024), "Assimilation of Sentinel-Based Leaf Area Index for Modeling Surface-Ground Water Interactions in Irrigation Districts", as it is highly relevant to the present study.

**Response:**

Zafarmomen et al. [12] successfully assimilated high-resolution satellite-based Leaf Area Index into a coupled hydrological model to improve surface-groundwater simulations. This work is highly relevant to our study as it highlights the growing application of emerging remote sensing observations in hydrologic data assimilation.

**Changes:** Introduction, L28-29.

# References


[1] Bin Cao et al. "The ERA5-Land Soil Temperature Bias in Permafrost Regions". In: *The Cryosphere* 14.8 (Aug. 2020), pp. 2581–2595. DOI: 10.5194/tc-14-2581-2020.

[2] Weijing Chen et al. "Retrieving Accurate Soil Moisture over the Tibetan Plateau Using Multisource Remote Sensing Data Assimilation with Simultaneous State and Parameter Estimations". In: *Journal of Hydrometeorology* 22.10 (Oct. 2021), pp. 2751–2766. DOI: 10.1175/JHM-D-20-0298.1.

[3] Andreas Colliander et al. “Validation of Soil Moisture Data Products From the NASA SMAP Mission”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 364–392. DOI: 10.1109/JSTARS.2021.3124743.

[4] Yongjiu Dai, Robert E. Dickinson, and Ying-Ping Wang. “A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance”. In: *Journal of Climate* 17.12 (June 2004), pp. 2281–2299. DOI: 10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2.

[5] Gregory Gaspari and Stephen E. Cohn. “Construction of Correlation Functions in Two and Three Dimensions”. In: *Quarterly Journal of the Royal Meteorological Society* 125.554 (Jan. 1999), pp. 723–757. DOI: 10.1002/qj.49712555417.

[6] C Huang, X Li, and L Lu. “Retrieving Soil Temperature Profile by Assimilating MODIS LST Products with Ensemble Kalman Filter”. In: *Remote Sensing of Environment* 112.4 (Apr. 2008), pp. 1320–1336. DOI: 10.1016/j.rse.2007.03.028.

[7] Kathrin Naegeli et al. *ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction - Snow on Ground (SCFG) from AVHRR (1982 - 2018), Version 2.0.* application/xml. NERC EDS Centre for Environmental Data Analysis, 2022. DOI: 10.5285/3F034F4A08854EB59D58E1FA92D207B6.

[8] Thomas Nagler et al. *ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction - Snow on Ground (SCFG) from MODIS (2000-2020), Version 2.0.* application/xml. NERC EDS Centre for Environmental Data Analysis, 2022. DOI: 10.5285/8847A05EEDA646A29DA58B42BDF2A87C.

[9] K.A. Novick et al. “The AmeriFlux Network: A Coalition of the Willing”. In: *Agricultural and Forest Meteorology* 249 (Feb. 2018), pp. 444–456. DOI: 10.1016/j.agrformet.2017.10.009.

[10] Gilberto Pastorello et al. “The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data”. In: *Scientific Data* 7.1 (July 9, 2020), p. 225. DOI: 10.1038/s41597-020-0534-3.

[11] Rune Solberg et al. *ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction – Snow on Ground (SCFG) from ATSR-2 (1995 – 2003), Version 1.0.* application/xml. NERC EDS Centre for Environmental Data Analysis, 2023. DOI: 10.5285/0AEBA0C203C2447B9553A78F99D3A276.

[12] Nima Zafarmomen et al. “Assimilation of Sentinel-Based Leaf Area Index for Modeling Surface-Ground Water Interactions in Irrigation Districts”. In: *Water Resources Research* 60.10 (Oct. 2024), e2023WR036080. DOI: 10.1029/2023WR036080.