Review of "Effectively Assimilate Satellite Land Surface Temperature into Offline Land Surface Models within Ensemble-based Assimilation Frameworks"

Summary

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.

Major Concerns

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.

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.

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?

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.

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.

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?

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

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.

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

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.

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.

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

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.

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.

Line 290: Please clarify how “error” is defined here.

Line 296: “temperature” should be changed to “soil temperature”.

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

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