Geosci. Model Dev., 19, 9177–9201, 2026
https://doi.org/10.5194/gmd-19-9177-2026

Geoscientific Model Development
Open Access

Development and technical paper

Effectively assimilate satellite land surface temperature into offline land surface models within ensemble-based assimilation frameworks

Yunhao Fu[1,2], Yongjun Zheng[1,3], and Jingjia Luo[1,3]

[1]State Key Laboratory of Climate System Prediction and Risk Management (CPRM)/Institute of Climate Application Research (ICAR), Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China

Correspondence: Yongjun Zheng (yongjun.zheng@nuist.edu.cn)

Received: 10 December 2025 – Discussion started: 2 March 2026
Revised: 28 August 2026 – Accepted: 11 September 2026 – Published: 28 September 2026

Abstract. Land surface temperature (LST) plays a vital role in controlling the water and energy fluxes at the interface between the land and atmosphere, and the main aim of assimilating LST observations into Land Surface Models (LSMs) is to not only provide better initial conditions for the LSM itself, but also yield more accurate land–atmosphere interactions. While observation systems provide a vast amount of satellite-derived LST observations in recent years, they are not as widely used as soil-moisture observations in land data assimilation (DA), in research or in operations, owing to the fast temporally varying nature of LST. This study proposes a new scheme to effectively improve the impact of LST assimilation by simultaneously updating soil temperature and soil moisture within the upper soil layers. The proposed scheme outperforms the conventional strategy of updating temperature only, and it also yields significant benefits for hydrological variables such as soil moisture and snow. The DA method and LSM used in this study are the Local Ensemble Transform Kalman Filter (LETKF) and the Common Land Model (CoLM), respectively. Moderate Resolution Imaging Spectroradiometer (MODIS) derived LST is assimilated into CoLM every 3 h using the proposed scheme. The assimilation and open-loop experiments are conducted for one year with a global resolution of $0.5° \times 0.5°$. The LST shows marginal enhancement after assimilation, owing to its fast-varying nature dominated by atmospheric forcings. However, the BIAS in soil temperature over Northeast Asia is reduced significantly, with a magnitude of 1.0, 1.5, and 2.0 K for the layers within 0–10, 40–100, and 100–200 cm, respectively. Prominent improvements in snow temperature and snow depth are observed over Northeast Asia, with a reduction in root mean square difference (RMSD) of approximately 4 K and 150 mm, respectively. The improvements in soil water content are also notable, particularly over humid tropical regions. The largest reductions in unbiased RMSD of soil water content over the Amazon Rainforest are approximately 0.06, 0.12, 0.15, and 6.00 $kg\,m^{-2}$ for the layers within 0–10, 10–40, 40–100, and 100–200 cm, respectively. These consistent improvements in both the energy and water components of CoLM demonstrate the effectiveness of the proposed scheme and the importance of LST assimilation for land-surface-process modeling.

1 Introduction

Land surface temperature (LST) is defined as a Global Climate Observing System Essential Climate Variable for its irreplaceable value in assessing both the land-atmosphere interaction and the land surface process itself (Global Climate Observing System, 2022). Through soil moisture and LST, understanding of the land surface process from a climatic perspective has made significant progress in the last few decades (Jin and Dickinson, 2002; Kalnay and Cai, 2003; Seneviratne et al., 2010). To a large extent, this progress can be seen as a result of the rapid development of satellite-based observations at the land surface, which provide much broader spatial coverage than ground-based measurements and yield supplementary information for regions with few in situ observations (De Lannoy et al., 2022). These newly emerging remote sensing observations have gained popularity in the field of land and hydrologic data assimilation (DA) (McLaughlin, 1995; Swinbank et al., 2003; Reichle, 2008; Khaki et al., 2020; Zafarmomen et al., 2024). Together with the conventional observations, remote sensing observations have been assimilated into Land Surface Models (LSMs) in many operational centers worldwide (Rodell et al., 2004; Balsamo et al., 2007; Drusch, 2007).

For the long memory (Vinnikov et al., 1996; Koster and Suarez, 2001) and continuous influence of the land surface on the atmosphere, which is somewhat comparable to the sea-atmosphere interaction (Koster et al., 2000, 2004), satellite-based soil moisture has been assimilated into several LSMs and led to significant improvements in both operational and research contexts (Reichle et al., 2002; Crow and Wood, 2003; De Rosnay et al., 2013, 2014; Muñoz-Sabater et al., 2019; Bonan et al., 2020). In the land DA community, the variables directly associated with the atmosphere such as evaporation (Meng et al., 2009) or LST itself (Bosilovich et al., 2007), rather than the variables within the upper soil layers such as soil temperature and moisture, are updated when assimilating LST. This preference is a consequence of the extremely volatile energy stored in the upper soil layers, which is strongly controlled by the incoming solar radiation and outgoing thermal radiation. Therefore, it seems to be more effective to assimilate LST into the atmosphere through land-atmosphere interaction. While assimilating LST into atmosphere could be easily achieved in a coupled model, where the interaction between land and atmosphere is bidirectional (Browne et al., 2019; Zhang et al., 2020), it is not applicable in an offline LSM, where the atmosphere only acts as boundary conditions for the land surface.

Based on the offline Common Land Model version 3.0 (CoLM, Dai et al., 2003) and the Ensemble Kalman Filter (EnKF, Evensen, 1994; Burgers et al., 1998; Evensen, 2003), a one-dimensional land DA scheme was designed to update the soil temperature profile on sites by assimilating LST (Huang et al., 2008). Through a simple observation operator based on the linear relationship between LST and in situ ground temperature, the assimilation results exhibited significant enhancements over the open loop (i.e., without DA). This improvement stemmed primarily from utilizing LST to update soil temperature encompassing both the upper and the deeper soil layers, as the deeper soil layers prevent the adjusted energy from being quickly dissipated and therefore keep the impact longer than the upper ones. Although the main motivation of this early study was to demonstrate the potential of assimilating remote sensing observations into the CoLM using the EnKF, it is noteworthy that directly using LST to update the soil temperature profile may not be physically plausible. This is because the influence of LST is typically limited to the surface and does not extend to a significant depth (Jin and Dickinson, 2010). Despite other attempts to assimilate LST into offline LSMs, such as storing its impact within the soil water content (Lakshmi, 2000) or employing a bias correction method before performing DA (Bosilovich et al., 2007; Reichle et al., 2010), the question of how to effectively assimilate LST into LSMs remains open in an offline context.

Following the work by Huang et al. (2008), Chen et al. (2021) assimilated brightness temperature to update soil moisture and LST to update soil temperature, respectively, into the CoLM using the EnKF with an observation operator based on physics. The updated states are also influenced by distant observations, which helps to mitigate spatial discontinuities in the analysis increments. Although the improvements are prominent in both soil moisture and temperature when assimilating multi-source remote sensing measurements, it is challenging to distinguish the contribution of LST among the different observations.

The aforementioned studies demonstrate the promising benefits of assimilating LST into LSMs; however, their assimilation strategies are limited to updating individual states separately. Specifically, Huang et al. (2008) assimilated LST to update only the soil temperature profile, whereas Chen et al. (2021) assimilated LST and brightness temperature separately to update soil temperature and soil moisture, respectively. Initially, assimilating LST to update only soil temperature yielded marginal improvements because the surface thermal memory is rapidly lost as the state relaxes back to the trajectory dictated by atmospheric forcing. Therefore, a joint update of soil temperature and soil moisture in the top two layers was introduced, revealing a promising improvement (Fig. 1). In addition, the joint update also improves snow-related variables notably compared with updating only soil temperature (results not shown here). By jointly updating these variables, the persistent memory of soil moisture and the thermal inertia of soil temperature are leveraged to sustain the post-assimilation impact. Although soil moisture is not explicitly used in the LST observation operator (see Sect. 2.2), this joint update can be achieved implicitly within an ensemble-based assimilation framework via the background error covariance between soil water content and soil temperature (see Sect. 2.3). Hence, this paper proposes jointly updating soil temperature and soil moisture when assimilating LST in an offline land surface model within an ensemble framework, and systematically evaluates the assimilation performance.

The land DA has evolved from incorporating near-surface meteorological observations to initialize soil moisture in LSMs (Mahfouf, 1991), to assimilating brightness temperature to update soil profile properties by solving the inverse problem with the Extended Kalman Filter (Entekhabi et al., 1994), and to assimilating brightness temperature with the EnKF (Crow and Wood, 2003). This study adopts the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) for its ease of implementation and high degree of parallelism. The LETKF has been widely utilized in atmospheric research (Bonavita et al., 2010) and operational settings (Miyoshi et al., 2010; Hamrud et al., 2015), and it

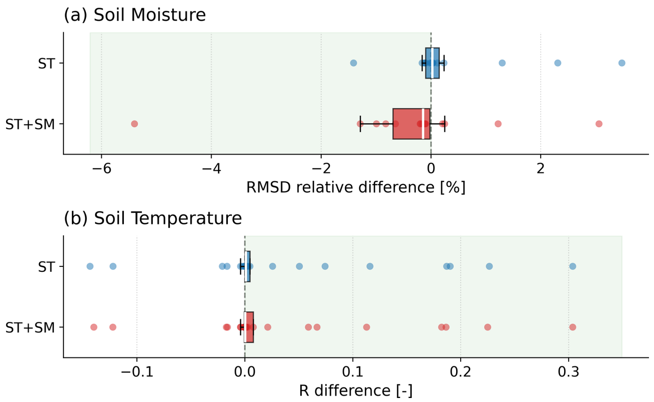

Figure 1. Comparison of assimilation performance between the joint-update (ST+SM) and temperature-only (ST) schemes across AmeriFlux sites. Panel **(a)** shows the RMSD relative difference (%) for soil moisture. Panel **(b)** shows the correlation coefficient (R) difference for soil temperature.

has been applied to several LSMs for land studies (Han et al., 2014; Seo et al., 2021; Kurosawa et al., 2023). The LST retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) is used in this study for its robust quality (Pérez-Planells and Martin, 2020) and extensive applications to LSMs (Yu et al., 2014; Fu et al., 2020; Huang et al., 2008).

This paper demonstrates jointly updating soil temperature and moisture to effectively assimilate LST into offline land surface models within ensemble-based assimilation frameworks. Section 2 briefly introduces the CoLM, the LST observation operator and the LETKF method. Section 3 describes the atmospheric forcing, the LST observations, the validation datasets, and the experimental design. Comprehensive evaluations of the performance of LST assimilation are presented in Sect. 4. Section 5 discusses the issues encountered in this study and gives prospects. Finally, Sect. 6 gives the conclusions of this study.

2 Methods

2.1 CoLM Configuration and Ensemble Generation

CoLM was developed by a collaborative effort of scientists with the conception of a modular LSM for weather prediction and climate research (Dai et al., 2003). It incorporates the most advantageous features of three well-documented and exceptional LSMs (Bonan, 1996; Dickinson et al., 1993; Dai and Zeng, 1997), and its technical details are well documented (Ji and Dai, 2010). It has undergone critical improvements over the original version in various aspects, among which the most notable feature is the two-big-leaf scheme, making more accurate representation of the leaf temperature possible (Dai et al., 2004). The specific version used in this study is from the offline component of the GRAPES (Global/Regional Assimilation and Prediction Enhanced System) model at the China Meteorological Administration, and it is developed based on CoLM 2014.

It comprises one vegetation layer, 10 soil layers with thicknesses of 0.018, 0.028, 0.045, 0.075, 0.124, 0.204, 0.336, 0.554, 0.913, and 1.137 m, and dynamic snow layers with levels up to 5. The vertical layer index l is from -4 to 10 with $-4 \leq l \leq 0$ for snow layers and $1 \leq l \leq 10$ for soil layers. CoLM adopts the concept of mosaic (Avissar, 1992; Koster and Suarez, 1992), allowing each model grid cell to be subdivided into different patches, each for a unique land cover type.

CoLM's integration accounts for both energy and water processes, with theoretical details provided in Dai et al. (2003) and implementation specifics outlined in Ji and Dai (2010). In this study, the state variables updated by DA include sunlit leaf temperature (T_{lsun}) and shaded leaf temperature (T_{lsha}) in the canopy; ground temperature (T_{g}, temperature at the top valid layer); soil temperature (T_l), soil liquid water (W_l^{liq}), and soil solid water (W_l^{ice}) within the top two valid layers ($l_{\mathrm{top}} \leq l \leq l_{\mathrm{top}}+1$) – these soil variables are updated simultaneously. When snow is present, snow temperature and water/ice content are updated analogously. The index of the top valid layer $l_{\mathrm{top}} = 1$ if no snowpack was accumulated above the ground, otherwise $l_{\mathrm{top}} = l$, where l is the first layer (starting from -4 to 0) for which $W_l^{\mathrm{liq}} + W_l^{\mathrm{ice}} > 0$.

For the assimilation experiment, 14 tunable parameters (Table 1) in the CoLM are perturbed as follows to account for model uncertainty, which is vital for maintaining a realistic ensemble spread.

$$\widetilde{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda} + r_k \boldsymbol{\Delta}, \quad k = 1, 2, \cdots, K, \tag{1}$$

where K and r_k are the number of ensemble members and a random number drawn from a uniform distribution $U(-1, 1)$ for the kth ensemble member; $\boldsymbol{\lambda}$ and $\boldsymbol{\Delta}$ are the vectors of the default values (the 2nd column in Table 1) and the maximum perturbed magnitudes (the 3rd column in Table 1) of the tunable parameters, respectively. The random factor r_k is re-sampled at the beginning of each assimilation cycle for each ensemble member to represent time-varying model uncertainty. This perturbation scheme maintains a realistic ensemble spread while preserving the model's intrinsic water and energy balance, as confirmed by preliminary diagnostic tests.

Table 1. Tunable parameters, their default values and maximum perturbed magnitudes in CoLM.

Tunable parameter	Default value ($\boldsymbol{\lambda}$)	Maximum perturbed magnitude ($\boldsymbol{\Delta}$)
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]	-1.5×10^5	2.25×10^4
Restriction for minimum of soil potential [mm]	-1.0×10^8	1.5×10^7
Maximum transpiration for moist soil + 100 % veg [$\mathrm{mm\,s^{-1}}$]	0.0002	0.00003
Critical temperature [°C] (to determine rain or snow)	0.0	0.15

2.2 Land Surface Temperature Observation Operator

In this study, the LST is simulated by the CoLM as follows:

$$\mathrm{LST}_{\mathrm{simulated}} = \mathcal{H}(T_{\mathrm{lsun}}, T_{\mathrm{lsha}}, T_{\mathrm{g}}) = \left(\frac{\begin{array}{c} F_{\mathrm{cover}}\sigma(1-c_{\mathrm{g}})\left(F_{\mathrm{sun}}T_{\mathrm{lsun}}^4 + F_{\mathrm{sha}}T_{\mathrm{lsha}}^4\right) \\ +F_{\mathrm{cover}}\sigma\epsilon_{\mathrm{g}}c_{\mathrm{g}}T_{\mathrm{g}}^4 + (1-F_{\mathrm{cover}})(1-\epsilon_{\mathrm{g}})\mathrm{DLW} \\ +(1-F_{\mathrm{cover}})\sigma\epsilon_{\mathrm{g}}T_{\mathrm{g}}^4 \end{array}}{\sigma} \right)^{\frac{1}{4}}, \tag{2}$$

where $\mathcal{H}(\cdot)$ is the nonlinear observation operator that projects the model space onto the observation space, F_{cover} is the fraction of vegetation cover, σ is the Stefan-Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$), c_{g} is the canopy gap fraction for thermal infrared radiation, F_{sun} is the sunlit fraction of the canopy, F_{sha} is the shaded fraction of the canopy, ϵ_{g} is the emissivity of the ground, and DLW is the downward longwave radiation. This operator renders the realistic LST within CoLM's two-big-leaf framework (Dai et al., 2004) by explicitly combining thermal emissions from the sunlit canopy, shaded canopy, and ground. This approach is more physically consistent with the radiometric nature of satellite observations, particularly over partially vegetated surfaces.

It is shown from Eq. (2) that the derivation of LST only accounts for the energetic aspects of state variables, namely T_{lsun}, T_{lsha}, and T_{g}. The intrinsic connection between the energy and water processes implies a cross-covariance between the LST and the water content (W^{liq} or W^{ice}). To enhance the influence of LST assimilation, this study utilizes the implicit cross-covariance to jointly update the temperature and the water content, as described in Sect. 2.3.

2.3 The Local Ensemble Transform Kalman Filter Scheme

The stochastic EnKF perturbs observations for the posterior sample covariance to be consistent with the theoretical one (Burgers et al., 1998). By contrast, the deterministic EnKF updates the ensemble mean and covariance directly, thereby eliminating the sampling error induced by perturbing observations (Whitaker and Hamill, 2002). Deterministic EnKF variants include the Ensemble Square Root Filter (Whitaker and Hamill, 2002), the Ensemble Adjustment Kalman Filter (Anderson, 2001), and the Ensemble Transform Kalman Filter (ETKF, Bishop et al., 2001). The LETKF which combines the ETKF framework with localization strategy of Local Ensemble Kalman Filter (Ott et al., 2004) is classified as one of the Ensemble Square Root Filters (Tippett et al., 2003).

Let M be the dimension of the model state, K be the ensemble size, and P be the number of observations. Let the kth member at time t be $\boldsymbol{x}_k^{\mathrm{g}}(t) \in \mathbb{R}^M$, where the superscript g denotes the global state (hereafter). The ensemble mean is

$$\overline{\boldsymbol{x}}^{\mathrm{g}}(t) = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k^{\mathrm{g}}(t) \in \mathbb{R}^M, \tag{3}$$

and deviation from the ensemble mean of state variables is

$$\mathbf{X}^{\mathrm{g}}(t) = \left[\boldsymbol{x}_1^{\mathrm{g}}(t) - \overline{\boldsymbol{x}}^{\mathrm{g}}(t), \cdots, \boldsymbol{x}_K^{\mathrm{g}}(t) - \overline{\boldsymbol{x}}^{\mathrm{g}}(t)\right] \in \mathbb{R}^{M\times K}. \tag{4}$$

Thus, the sample error covariance is estimated as

$$\mathbf{P}^{\mathrm{g}}(t) = \frac{1}{K-1}\mathbf{X}^{\mathrm{g}}(t)\left(\mathbf{X}^{\mathrm{g}}(t)\right)^{\mathrm{T}} \in \mathbb{R}^{M\times M}, \tag{5}$$

where the superscript T denotes the transpose. In the forecast step, each ensemble member is advanced independently by the nonlinear model $\mathcal{M}$:

$$\boldsymbol{x}_k^{\mathrm{gb}}(t) = \mathcal{M}\left(\boldsymbol{x}_k^{\mathrm{ga}}(t-1), \widetilde{\boldsymbol{\lambda}}_k\right), \quad k = 1, 2, \ldots, K, \tag{6}$$

where the superscripts a and b denote the analysis and background, respectively; $\widetilde{\boldsymbol{\lambda}}_k$ is the tunable parameter calculated in Eq. (1).

For the analysis step, the time index t is hereafter omitted for clarity since only a time step is involved. Let $\boldsymbol{y}^{\text{go}} \in \mathbb{R}^P$ be the actual observations with an error covariance $\mathbf{R}^{\text{g}} \in \mathbb{R}^{P\times P}$. The background simulated observations of the kth member are $\boldsymbol{y}_k^{\text{gb}} = \mathcal{H}(\boldsymbol{x}_k^{\text{gb}}) \in \mathbb{R}^P$, where $\mathcal{H}$ is the nonlinear observation operator. The ensemble mean of the background simulated observations is

$$\overline{\boldsymbol{y}}^{\text{gb}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{y}_k^{\text{gb}} \in \mathbb{R}^P, \tag{7}$$

and the deviation from the ensemble mean of background simulated observations is

$$\mathbf{Y}^{\text{gb}} = \left[\boldsymbol{y}_1^{\text{gb}} - \overline{\boldsymbol{y}}^{\text{gb}}, \cdots, \boldsymbol{y}_K^{\text{gb}} - \overline{\boldsymbol{y}}^{\text{gb}}\right] \in \mathbb{R}^{P\times K}. \tag{8}$$

The LETKF performs the analysis in a pointwise way. For each grid point i, the observations whose distance to the grid point i is less than the prescribed localization radius are the local observations which are assimilated by the LETKF. Let $\boldsymbol{j}$ be the global indices of the local observations of the grid point i and P_i be the number of these local observations. The variables in the left hand side of Eqs. (9) and (10) are therefore local to the grid point i and are used in Eqs. (11)–(14) for the local analysis. Formally,

$$\begin{cases} \boldsymbol{y}^{\text{o}} = \boldsymbol{y}^{\text{go}}(\boldsymbol{j}), \\ \mathbf{R} = \mathbf{R}^{\text{g}}(\boldsymbol{j}, \boldsymbol{j}), \\ \overline{\boldsymbol{y}}^{\text{b}} = \overline{\boldsymbol{y}}^{\text{gb}}(\boldsymbol{j}), \\ \mathbf{Y}^{\text{b}} = \mathbf{Y}^{\text{gb}}(\boldsymbol{j}), \end{cases} \tag{9}$$

are the actual observations, observation error covariance, mean of background simulated observations, ensemble deviation of the background simulated observations within the given radius of the analysis grid point i, respectively. And

$$\begin{cases} \boldsymbol{x}_k^{\text{b}} = \boldsymbol{x}_k^{\text{gb}}(i), \quad k = 1, 2, \cdots, K \\ \overline{\boldsymbol{x}}^{\text{b}} = \overline{\boldsymbol{x}}^{\text{gb}}(i), \\ \mathbf{X}^{\text{b}} = \mathbf{X}^{\text{gb}}(i), \end{cases} \tag{10}$$

are the ensemble member, ensemble mean, and ensemble deviation of the background state at the analysis grid point i, respectively.

The LETKF updates the covariance in the ensemble space as follows:

$$\tilde{\mathbf{P}}^{\text{a}} = \left[(K-1)\mathbf{I} + (\mathbf{Y}^{\text{b}})^{\text{T}}\tilde{\mathbf{R}}^{-1}\mathbf{Y}^{\text{b}}\right]^{-1} \in \mathbb{R}^{K\times K}, \tag{11}$$

where $\mathbf{I} \in \mathbb{R}^{K\times K}$ is an identity matrix, $\tilde{\mathbf{R}} \in \mathbb{R}^{P_i\times P_i}$ is the localized observation error covariance. Inherited from the ETKF, the LETKF covariance update is computationally efficient, especially when $K \ll M$. With the updated covariance $\tilde{\mathbf{P}}^{\text{a}}$, the analysis mean is computed as

$$\overline{\boldsymbol{w}}^{\text{a}} = \tilde{\mathbf{P}}^{\text{a}}(\mathbf{Y}^{\text{b}})^{\text{T}}\tilde{\mathbf{R}}^{-1}(\boldsymbol{y}^{\text{o}} - \overline{\boldsymbol{y}}^{\text{b}}) \in \mathbb{R}^K \Rightarrow \overline{\boldsymbol{x}}^{\text{a}} = \overline{\boldsymbol{x}}^{\text{b}} + \mathbf{X}^{\text{b}}\overline{\boldsymbol{w}}^{\text{a}}, \tag{12}$$

and the analysis ensemble members are constructed as

$$\boldsymbol{x}_k^{\text{a}} = \overline{\boldsymbol{x}}^{\text{a}} + \mathbf{X}^{\text{b}}\mathbf{W}_k^{\text{a}}, \quad \text{where } k = 1, \cdots, K. \tag{13}$$

where the analysis increment $\mathbf{X}^{\text{a}} = \mathbf{X}^{\text{b}}\mathbf{W}^{\text{a}}$ could be computed differently in different deterministic EnKFs. The LETKF adopts a symmetric square-root approach for

$$\mathbf{W}^{\text{a}} = \left[(K-1)\tilde{\mathbf{P}}^{\text{a}}\right]^{\frac{1}{2}}. \tag{14}$$

This manner guarantees that the mean of columns of $\mathbf{X}^{\text{a}}$ is zero; therefore the analysis mean is preserved. Moreover, because $\mathbf{W}^{\text{a}}$ depends continuously on $\tilde{\mathbf{P}}^{\text{a}}$ (Hunt et al., 2007), the analysis ensemble does not change abruptly.

Even some state variables are not directly connected to the observation operator, that is, they are not an input of the observation operator, it is feasible to update them through the cross-covariances. The high correlation between soil temperature and soil moisture therefore justifies the joint update of both variables by assimilating only land surface temperature in this study.

Similar to the sampling error that arises from perturbing observations, the background error covariance suffers from the sampling error due to a small ensemble size – a problem known as rank deficiency. In real applications, the ensemble dimension K is typically orders of magnitude smaller than the model dimension M (e.g. $K \sim 10^2$ and $M \sim 10^8$ for operational numerical weather prediction scenarios), so the estimated background error covariance can be severely degraded and consequently undermine the analysis (Hamill et al., 2001).

A variety of techniques have been developed to mitigate this error, including covariance inflation (Anderson and Anderson, 1999), cross-validation (Houtekamer and Mitchell, 1998), and localization (Houtekamer and Mitchell, 1998). Cross-validation can also be incorporated into the LETKF (Buehner, 2020). Distance-based domain localization – an integral component of the LETKF inherited from the Local Ensemble Kalman Filter – suppresses spurious long-range correlations by tapering the impact of distant observations (Hunt et al., 2007). Covariance localization is often advocated for non-local observations such as satellite radiances whose vertical location is undefined (Houtekamer and Zhang, 2016), but it is challenging to implement within the LETKF framework (Farchi and Bocquet, 2019; Bishop et al., 2017).

To taper the influence of a distant observation j on the grid point i, its observation error variance is inflated according to

$$\tilde{\mathbf{R}}_{j,j} = \frac{\mathbf{R}_{j,j}}{\rho(d(i,j), L)}, \tag{15}$$

where $d(i,j)$ is the distance between the ith grid point and the jth observation, L is half of the cutoff radius or the length-scale parameter, and $\rho(\cdot,\cdot)$ is the fifth-order piecewise rational function defined by Gaspari and Cohn (1999):

$$\rho(d,L)= \begin{cases} 1-\frac{5}{3}r^2+\frac{5}{8}r^3+\frac{1}{2}r^4-\frac{1}{4}r^5, & 0\le r\le 1,\\ 4-5r+\frac{5}{3}r^2+\frac{5}{8}r^3-\frac{1}{2}r^4+\frac{1}{12}r^5-\frac{2}{3r}, & 1<r<2,\\ 0, & r\ge 2, \end{cases} \tag{16}$$

where $r=d/L$. Following Miyoshi et al. (2007), the cutoff distance ($2L$) decreases linearly from 730 km at the equator to 146 km at 80° N/S, corresponding to length-scale parameters L of 365 and 73 km, respectively. The latitude dependence avoids unrealistically broad analysis increments at high latitudes, where the physical grid spacing becomes smaller. This localization scale is physically justified: while it is larger than the minimum scale of land surface heterogeneity, it remains considerably smaller than typical atmospheric system scales and properly accounts for the spatial representation of aggregated satellite observations.

3 Datasets and Experiments

3.1 Forcing, Observation, and Evaluation Datasets

3.1.1 Atmospheric forcing dataset

To run an offline LSM, the atmospheric forcing is needed (e.g., wind, temperature, humidity, precipitation, pressure, short- and long-wave radiation; for details, see Ji and Dai, 2010). In this study, we use the WFDE5 forcing dataset (Cucchi et al., 2020). It is generated by applying the WATCH Forcing Data methodology to the surface meteorological variables from the ERA5 reanalysis (Hersbach et al., 2020). The temporal resolution is hourly. Even though its spatial resolution ($0.5°\times0.5°$) is coarser than that (approximately $31\,\text{km}\times31\,\text{km}$) of the ERA5 reanalysis, the WFDE5 shows comparable performance to the latter and better performance in the estimation of precipitation (Cucchi et al., 2020). Because the soil moisture depends strongly on the scale and intensity of precipitation (Seneviratne et al., 2010), a better estimation of the soil moisture and other relevant state variables (e.g., soil temperature) may benefit the LST observation operator and thus potentially improve the analysis.

3.1.2 MODIS-LST

Observations play a crucial role in the assimilation. Their quality, temporal frequency, spatial resolution, and coverage directly influence both the analysis and the forecast. The land surface temperature observed by the MODIS onboard the Terra polar-orbiting satellite (Ghent et al., 2022) is used in this study, hereafter referred to as MODIS-LST. The MODIS-LST offers global coverage observations twice a day, with a fine spatial resolution of $0.01°\times0.01°$ and a temporal period from 2000 to 2018. Detailed information on this observation dataset can be found in its user guide (Ghent et al., 2021) and algorithm document (Ghent et al., 2023). Validation of MODIS-LST quality against in situ observations demonstrates its robustness during both daytime and nighttime, with biases typically falling within ±2.0 K (Pérez-Planells and Martin, 2020).

This study adopts a 3 h assimilation frequency (centered at 00:00, 03:00, ..., 21:00 UTC) to balance the physical characteristics of LST against model dynamics. Given that LST varies rapidly, a narrow assimilation window is essential to accurately capture its temporal evolution; conversely, the window cannot be excessively short, as the model requires adequate spin-up time to dynamically adjust following each assimilation step.

The size of MODIS-LST pixels is significantly smaller than that of a CoLM grid cell. For better representativeness, the MODIS-LST observations are aggregated into the closest CoLM grid cell as follows:

$$\begin{aligned} \mathrm{LST}^i_{\mathrm{agg}} &= \frac{\sum_j \cos(\phi_j)\mathrm{LST}^j_{\mathrm{raw}}}{\sum_j \cos(\phi_j)},\\ \mathrm{Err}^i_{\mathrm{agg}} &= \frac{\sum_j \cos(\phi_j)\mathrm{Err}^j_{\mathrm{raw}}}{\sum_j \cos(\phi_j)}, \end{aligned} \tag{17}$$

where $\mathrm{LST}^i_{\mathrm{agg}}$ and $\mathrm{Err}^i_{\mathrm{agg}}$ are the aggregated LST and observation error of the ith CoLM grid cell, respectively; $\mathrm{LST}^j_{\mathrm{raw}}$ and $\mathrm{Err}^j_{\mathrm{raw}}$ are the jth raw MODIS-LST and observation error within the ith CoLM grid cell and the time window, respectively; and ϕ_j is the latitude of the corresponding pixel. The latitude-dependent weighting accounts for the varying physical area of raw MODIS pixels during aggregation. The raw MODIS-LST observations with a quality control flag of 0 (indicating cloud or aerosol contamination) or with a value of −32 768 (representing invalid data) do not participate in aggregations. In addition, to ensure sufficient spatial representativeness and statistical confidence, if the total number of raw MODIS-LST observations aggregated into the ith CoLM grid cell is less than 1500, the $\mathrm{LST}^i_{\mathrm{agg}}$ is discarded. The observation errors are assumed to be independent, so the observation error covariance matrix is diagonal. No separate representativeness-error term is introduced because preliminary tests showed 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.

The observation operator described in Sect. 2.2 is applied to each patch of the ith grid cell, and then the resulting patch-level estimates $\mathrm{LST}^{p,i}_{\mathrm{sim}}$ are area-weighted to yield the grid-cell mean simulated LST

$$\begin{aligned} \mathrm{LST}^{p,i}_{\mathrm{sim}} &= \mathcal{H}(T_{\mathrm{lsun}}(p,i), T_{\mathrm{lsha}}(p,i), T_{\mathrm{g}}(p,i)),\\ \mathrm{LST}^i_{\mathrm{sim}} &= \frac{\sum_p w_p \mathrm{LST}^{p,i}_{\mathrm{sim}}}{\sum_p w_p} \end{aligned} \tag{18}$$

where w_p is the fractional area of patch p (patches classified as wetlands, lakes, or ocean are discarded). The observation innovation for grid cell i is therefore $\mathrm{OMB}_i = \mathrm{LST}^i_{\mathrm{agg}} - \mathrm{LST}^i_{\mathrm{sim}}$. The observation is rejected from assimilation whenever the ensemble-mean innovation exceeds 8 K, i.e., $|\frac{1}{K}\sum_{k=1}^{K}\mathrm{OMB}_{i,k}| > 8$. Through sensitivity tests, this threshold was determined to strike a balance between filtering out gross errors and retaining valid observations.

3.1.3 Evaluation datasets

Owing to the scarce in-situ land surface observations and the high uncertainty of land surface reanalysis, three gridded datasets are initially used to robustly evaluate the performance of the land DA in this study.

The first dataset employed is the land component of the fifth generation ECMWF reanalysis (hereafter ERA5-Land; Copernicus Climate Change Service, 2019), which describes the evolution of water and energy processes of the global land surface from 1950 to the present in a consistent manner and is produced by the Carbon Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land forced with the ERA5 reanalysis. Although the temporal resolution is the same as the ERA5 reanalysis (hourly), it provides a finer horizontal resolution (9 km) to capture a more realistic and spatially detailed land process. Although it is an open-loop product, observations are indirectly assimilated into the land surface through the atmospheric forcing. Validation against independent global models and both in situ and satellite observations has shown the added value of ERA5-Land, especially for the estimation of soil moisture and other related variables (Muñoz-Sabater et al., 2021).

In addition to ERA5-Land, the Global Land Data Assimilation System (hereafter GLDAS; Rodell et al., 2004; Beaudoing et al., 2020) is forced with a combination of model and observation data from 2000 to the present. The spatial and temporal resolution of GLDAS-2.1 is 0.25° and 3 h, respectively. It provides a near-global (180° W–180° E, 60° S–90° N) land surface state generated by the Noah LSM (Chen et al., 1996; Koren et al., 1999). Like ERA5-Land, GLDAS is an open-loop product in which observations are incorporated indirectly through the atmospheric forcing.

Finally, the Modern-Era Retrospective analysis for Research and Applications version 2 (hereafter MERRA2; Global Modeling and Assimilation Office and Pawson, 2015) provides global scale land surface diagnostics from 1980 to the present. While precipitation plays the dominant role in the hydrological process of the land, a model-generated precipitation dataset with significant errors in amounts and timing may bias the prediction of an LSM. The MERRA2 is forced with observation-based precipitation that contains smaller errors, yielding a more accurate depiction of the land surface conditions. The spatial resolution is $0.5° \times 0.625°$ (lat × lon), and the temporal resolution is 1 h.

Furthermore, to provide independent validation beyond the aforementioned gridded products, satellite retrievals and in-situ station data are incorporated. For the evaluation of snow cover, independent satellite observations retrieved from the ATSR-2 (Solberg et al., 2023), AVHRR (Naegeli et al., 2022), and MODIS (Nagler et al., 2022) are utilized. Additionally, to evaluate the surface turbulent fluxes, this study incorporates sensible and latent heat flux observations from the globally distributed FLUXNET2015 (Pastorello et al., 2020) and AmeriFlux (Novick et al., 2018) networks.

3.2 Design and Configuration of the Data Assimilation and Open-Loop Experiments

The overall workflow of the experimental design, including the initialization, DA and forecasting processes, is comprehensively illustrated in Fig. 2. Two experiments are performed with the near-balanced and consistent initial state of the CoLM obtained by a 52-year spin-up. The assimilation experiment (AEXP) is a one-year assimilation followed by a two-month ensemble forecast, that is, LST observations are assimilated every three hours from 1 January to 31 December 2001, and an ensemble forecast of two months from 1 January to 28 February 2002. The assimilation window is three hours and the CoLM tunable parameters of each ensemble member are perturbed (see Sect. 2.1). The open-loop experiment (OEXP) is a deterministic forecast of 14 months without perturbation of the tunable parameters. For AEXP, the initial ensembles were generated via the time-lag method by randomly sampling 50 states from an additional one-year simulation following the spin-up. This approach ensures that the initial members maintain physical consistency while providing a representative climatological spread. The forcing, time step, and spatial resolution of both experiments are WFDE5, 1800 s, and $0.5° \times 0.5°$, respectively.

After assimilation, to prevent model shock by restricting updates within the statistically reliable ensemble spread, if the magnitude of the analysis increment of a quantity x is three times larger than the standard deviation of its background, the analysis increment is clipped as follows:

$$\begin{cases} \sigma = \sqrt{\dfrac{\sum_{k=1}^{K}\left(x_k^{\mathrm{b}} - \overline{x}^{\mathrm{b}}\right)^2}{K-1}}, \\ x_k^{\mathrm{a}} = \min\left(x_k^{\mathrm{b}} + 3\sigma, \max\left(x_k^{\mathrm{b}} - 3\sigma, x_k^{\mathrm{a}}\right)\right). \end{cases} \tag{19}$$

To ensure thermodynamic consistency, snow variables are only updated when all ensemble members reach a consensus on the presence of snow and possess an identical number of snow layers. In addition, if the analyzed liquid water (W^{liq}) and solid water (W^{ice}) amounts are negative, they are reset to zero. To maintain stable model integration, a small amount of ice is redistributed from the first soil layer to the second if solid ice is present in the top layer but absent in the second. Finally, the snow water equivalent is recalculated through the

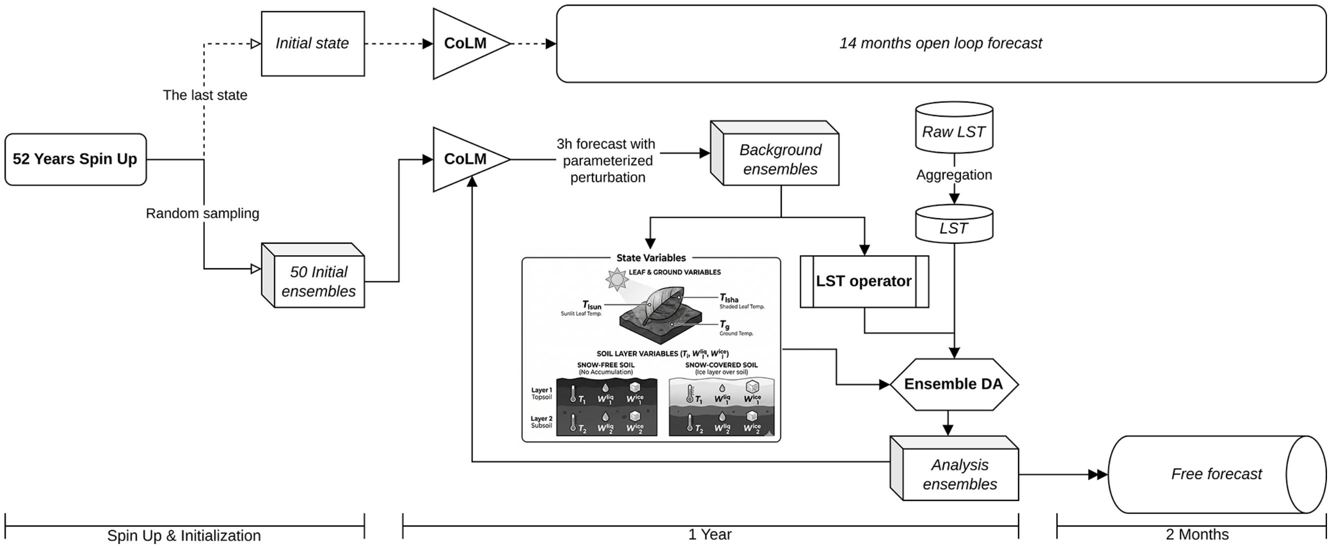

Figure 2. Schematic workflow of the coupling architecture and DA experiments. The process comprises a 52-year spin-up for initialization, a 1-year period comparing the assimilation group (solid lines, continuous 3 h cycles updating leaf, ground, and soil variables) against the open-loop baseline (dashed lines), and a subsequent 2-month free forecast.

accumulation of the updated water mass within the snow layers for consistency.

3.3 Evaluation Metrics

The ensemble mean of the assimilation experiment is compared with the open-loop experiment and the evaluation datasets. The following metrics are used:

$$\begin{cases} \mathrm{RMSD} = \sqrt{\dfrac{\sum_{n=1}^{N}(\hat{y}_n - y_n)^2}{N}}, \\ R = \dfrac{\sum_{n=1}^{N}\hat{y}'_n y'_n}{\sqrt{\sum_{n=1}^{N}\hat{y}'^2_n}\sqrt{\sum_{n=1}^{N}y'^2_n}}, \\ \mathrm{BIAS} = \dfrac{\sum_{n=1}^{N}(\hat{y}_n - y_n)}{N}, \\ \text{unbiased RMSD} = \sqrt{\dfrac{\sum_{n=1}^{N}(\hat{y}'_n - y'_n)^2}{N}}, \end{cases} \tag{20}$$

where $\hat{y}'_n = \hat{y}_n - \overline{\hat{y}}$, $y'_n = y_n - \overline{y}$, with $\overline{\hat{y}} = \frac{\sum_{n=1}^{N}\hat{y}_n}{N}$, $\overline{y} = \frac{\sum_{n=1}^{N}y_n}{N}$, and $\hat{y}_n$ and y_n denoting the model outputs and validation datasets, respectively. We also compute the metric differences (AEXP -— OEXP) to highlight improvements.

The details of the variables along with the metrics used for the evaluation against the gridded products are presented in Table 2. The metrics are selected based on the error characteristics of each variable. Specifically, soil moisture evaluation employs the unbiased RMSD to mitigate systematic biases arising from differing soil layer definitions across products, thereby focusing on the capture of dynamic variations. For other variables, absolute metrics such as RMSD and BIAS are used to evaluate their overall accuracy. The associated metrics include BIAS, root mean square difference (RMSD), unbiased RMSD (RMSD after removing the mean bias), and correlation coefficients (R). To directly demonstrate the effect of the assimilation compared to the open loop, the metric differences between the two experiments are defined as $\mathrm{BIAS_d} = |\mathrm{BIAS_{AEXP}}| - |\mathrm{BIAS_{OEXP}}|$, $\mathrm{RMSD_d} = \mathrm{RMSD_{AEXP}} - \mathrm{RMSD_{OEXP}}$, $(\text{unbiased RMSD})_\mathrm{d} = (\text{unbiased RMSD})_\mathrm{AEXP} - (\text{unbiased RMSD})_\mathrm{OEXP}$, and $R_\mathrm{d} = R_\mathrm{AEXP} - R_\mathrm{OEXP}$, where negative values of $\mathrm{BIAS_d}$, $\mathrm{RMSD_d}$, and $(\text{unbiased RMSD})_\mathrm{d}$ indicate that the assimilation reduces the error relative to the open loop, while positive values of R_d indicate an improved correlation. With respect to the evaluation of soil moisture with the ERA5-Land and MERRA2, instead of the $(\text{unbiased RMSD})_\mathrm{d}$, the relative $(\text{unbiased RMSD})_\mathrm{d} = (\text{unbiased RMSD})_\mathrm{d} / (|(\text{unbiased RMSD})_\mathrm{OEXP}|)$ is used because the denominator can be very small. The selection of reference datasets in Table 2 is primarily dictated by data availability; for instance, GLDAS lacks snow surface temperature and snow cover fraction data, while MERRA2 does not provide bare soil evaporation. For the independent satellite and in-situ observations, the $\mathrm{RMSD_d}$ is specifically adopted to quantify the improvements in snow cover fraction and surface heat fluxes.

4 Results

4.1 Validation based on MODIS-LST

Figure 3a shows the $\mathrm{RMSD_d}$ of LST based on MODIS-LST for 31 January 2001–28 February 2002. It can be seen that,

Table 2. Variables evaluated against the three gridded reanalysis/model products.

Variables in CoLM	ERA5-Land	GLDAS	MERRA2	Metrics associated
Soil temperature in different layers		✓	✓	BIAS, $BIAS_d$, RMSD
Snow temperature of surface	✓		✓	RMSD, $RMSD_d$
Fraction of snow cover	✓		✓	RMSD, *R*
Snow depth	✓	✓	✓	BIAS, $RMSD_d$
Soil moisture in different layers	✓	✓	✓	unbiased RMSD, relative (unbiased RMSD)$_d$
Evaporation from bare soil	✓	✓		RMSD, $RMSD_d$
Latent heat flux	✓	✓	✓	RMSD, $RMSD_d$
Sensible heat flux	✓	✓	✓	RMSD, $RMSD_d$

based on the previous 3 h analysis in AEXP, the predicted background LST is closer to the MODIS-LST compared with OEXP. This is evident from Fig. 3a since there are more negative (cyan) than positive (red) dots for $RMSD_d$. Specifically, the number (1995) of negative values for $RMSD_d$ roughly doubles that (999) of the positive values. This is clearer from the daily mean (solid black line in Fig. 3a). As exhibited in the two-month free forecast (pink background in Fig. 3a), the accumulated effect within the one-year assimilation of MODIS-LST persists for at least 2 months.

The small differences between the two experiments, as indicated by the absolute values of $RMSD_d$ in Fig. 3a (with the maximum absolute difference of about 0.1 K), are consistent with the previous discussions on the nature of LST (Swinbank et al., 2003; Jin and Dickinson, 2010). LST is strongly influenced by land-atmosphere interactions, whereas in an offline model, this bidirectional interaction reduces to an unidirectional forcing from the atmosphere. Furthermore, the deterministic atmospheric forcing strictly constrains the background LST. Therefore, for LST, which largely depends on the interaction between the land and atmosphere, it is not surprising that the differences introduced by the assimilation are quickly damped by the same atmospheric forcing during the 3 h forecast lead time.

The example depicted in Fig. 3b illustrates the number of observations centered on 06:00 UTC in the Eastern Hemisphere counting from 1 January 2001 to 31 December 2001. The maximum number is centered at about 80° E. Figure 3c exhibits the error of observations, RMSD for both the AEXP and the OEXP in the Eastern Hemisphere. It is shown that the error of observations (i.e., observation uncertainty) has a considerable fluctuation in different months, with the largest error emerging in March (5.1 K) and the smallest one in August (3.2 K). Compared with the global scale (figures omitted), over Asia the AEXP shows pronounced improvement over the OEXP, with all months exhibiting reductions in RMSD (the largest reduction, 1.43 %, occurs in July), indicating region-dependent LETKF performance and the added value of observations.

4.2 Evaluation based on third-party datasets

4.2.1 Impacts on Soil and Snow Temperature

Comparisons between the two experiments on soil temperature in different layers based on GLDAS and MERRA2 datasets are presented in Figs. 4 and 5. The yearly $BIAS_{OEXP}$ based on GLDAS is illustrated in the first column of Fig. 4. It can be inferred that the CoLM exhibits a large positive BIAS in depicting soil temperature in several regions, including western North America, western South America, South Africa, the Tibetan Plateau and Northeast Asia. These regions typically are arid zones, plateau areas, or permafrost soils. In contrast, relatively small negative BIAS is observed in Alaska, the southern part of the Middle East, and some areas of Northeast Asia.

The second and third rows in Fig. 4 mostly demonstrate the considerable improvements of the assimilation over the open loop based on the third-party datasets. In terms of the yearly $BIAS_d$ based on the GLDAS dataset (Fig. 4b, e and h), the maximum improvements in BIAS are centered on Northeast Asia (BIAS is reduced by approximately 1.0, 1.5, and 2.0 K for depths 0–10, 40–100, and 100–200 cm, respectively). This pattern of improvement is also distributed across most parts of eastern Asia, Europe, and eastern North America, which are mostly humid areas. Although improvements are evident in the Northern Hemisphere, small degradations emerge in the Southern Hemisphere, with the maximum centered on southern South America, South Africa, and Australia, which are typical arid zones in the LSM. These results suggest that the assimilation scheme effectively updates soil temperature in relatively humid areas and reduces bias appropriately. However, the effect of bias reduction is not pronounced in arid zones. A similar pattern can be seen in Fig. 4c, f, and i, which are based on the MERRA2 datasets.

Improvements are also evident from Fig. 5, where the $RMSD_{AEXP}$ is smaller than $RMSD_{OEXP}$ throughout the assimilation period with both GLDAS and MERRA2 benchmark datasets. The reduction of RMSD approximates 0.1 and 0.2 K for depths 0–10 and 100–200 cm, respectively, in particular, on 31 January 2001 with the GLDAS benchmark dataset. It can also be inferred from the temporal variation of

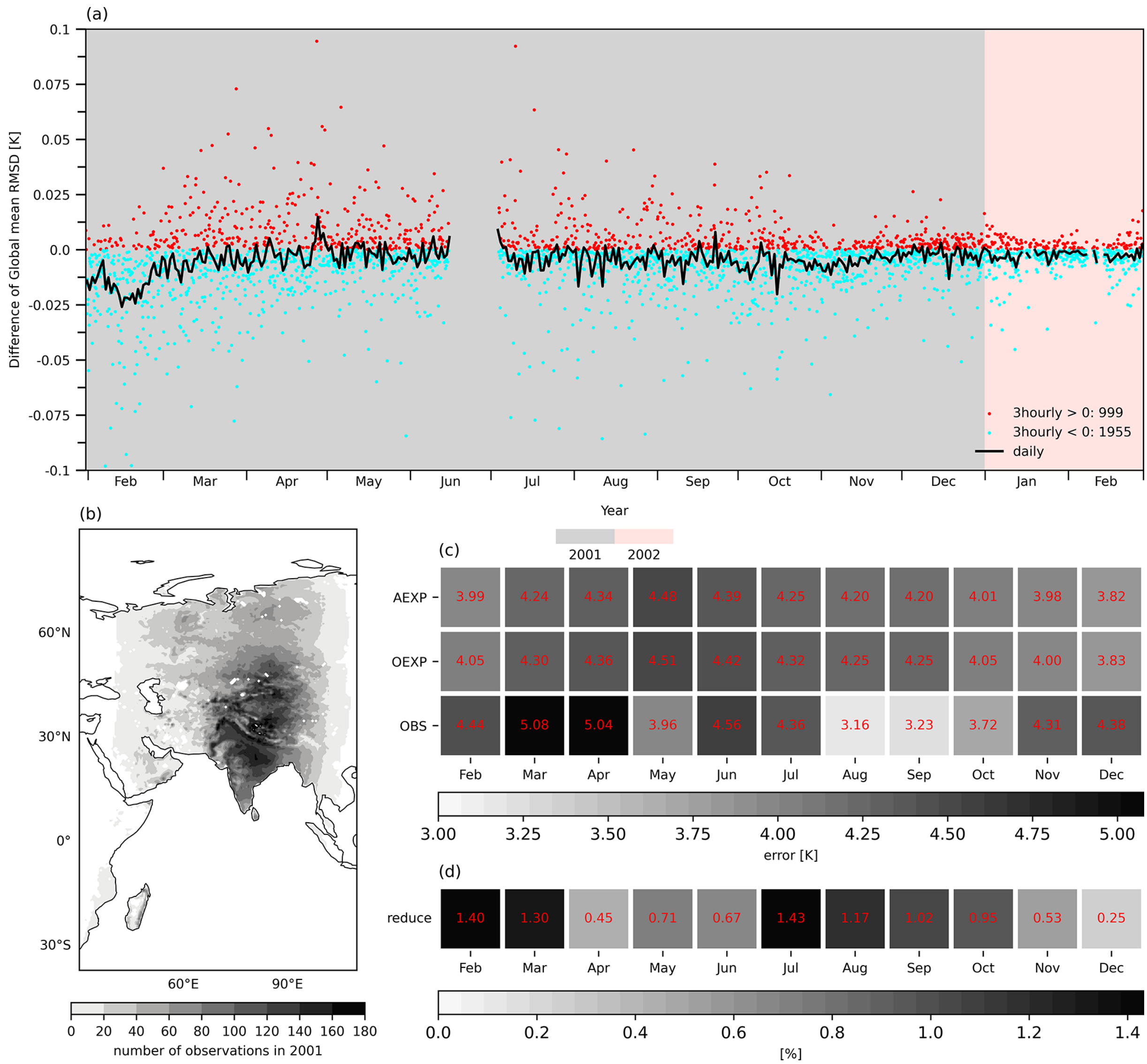

Figure 3. (a) Global LST $RMSD_d$ based on the MODIS-LST from 31 January 2001 to 28 February 2002 in 3 h intervals (red and cyan dots for positive and negative values, respectively) and daily mean (solid black line), as the observations are not available in some days between June and July (blank areas). The number of positive and negative values is shown in the legend. Gray and pink backgrounds represent the period of cycle DA and free forecast, respectively. **(b)** Number of observations centered at 06:00 UTC within the time window from 1 January–31 December 2001. **(c)** Eastern Hemisphere monthly mean spatially averaged error for observations (third row) and the LST RMSD for AEXP (first row) and OEXP (second row) based on the MODIS-LST, and **(d)** relative reduction in percent of LST RMSD from OEXP to AEXP.

RMSD of both experiments that the influence on soil temperature is depth dependent, as can be seen from Fig. 5, as the depth increases, taking MERRA2 as an example, the minimum RMSD occurs in August, September, October and November, indicating different memory effects in different soil layers.

Apart from both temporal and spatial improvements, the most promising finding is the dependence of improvements on the depth of soil layers. While relatively small improvements are observed in the upper soil layers (Figs. 4b and c, 5a), the improvements tend to become more significant in the deeper soil layers (Figs. 4e, f, h, and i, 5b–d), which is consistent with the nature of the land process. Although the soil temperature within the upper layers can be easily modulated by changes in the land-atmosphere interaction, energy stored in the deeper soil layers, especially within 100–200 cm, cannot be directly influenced by this interaction. Consequently, the signals modified by assimilating LST into CoLM can be transferred to and persist longer in these deeper soil layers, maintaining their influence within them for a longer time.

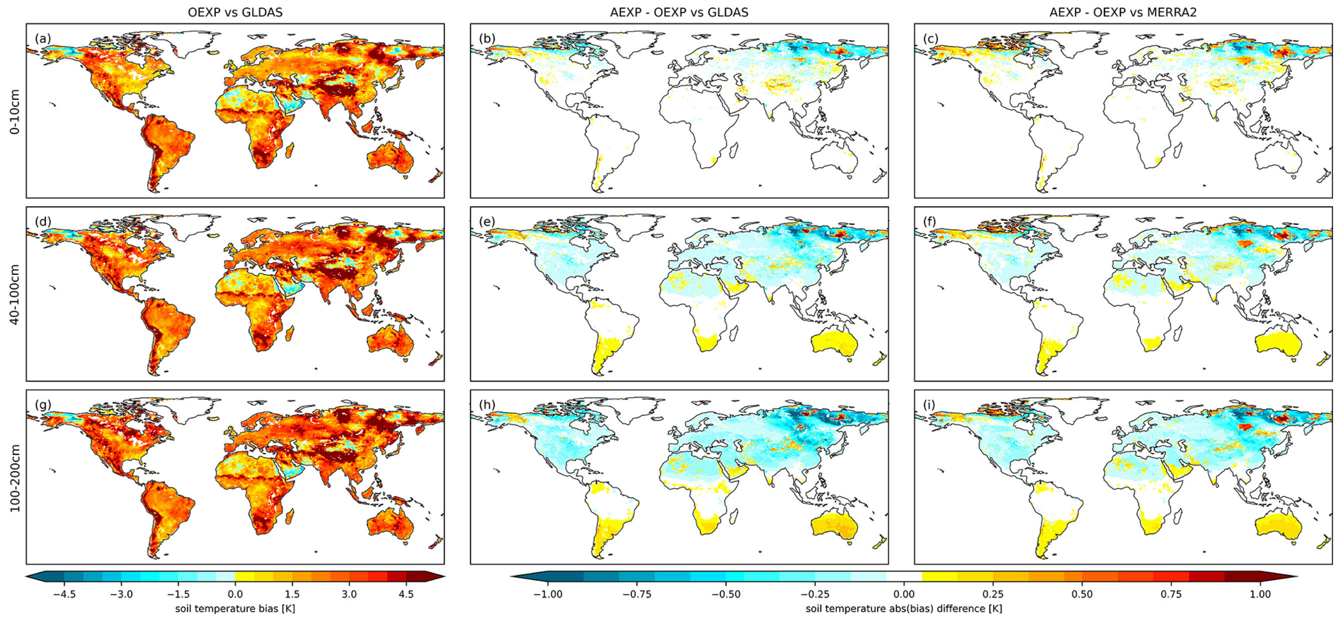

Figure 4. Soil temperature (in K) yearly (31 January 2001 to 31 December 2001) BIAS based on GLDAS of OEXP (first column) and $\mathrm{BIAS_d}$ based on GLDAS (second column) and MERRA2 (third column): **(a–c)** for 0–10 cm, **(d–f)** for 40–100 cm, and **(g–i)** for 100–200 cm.

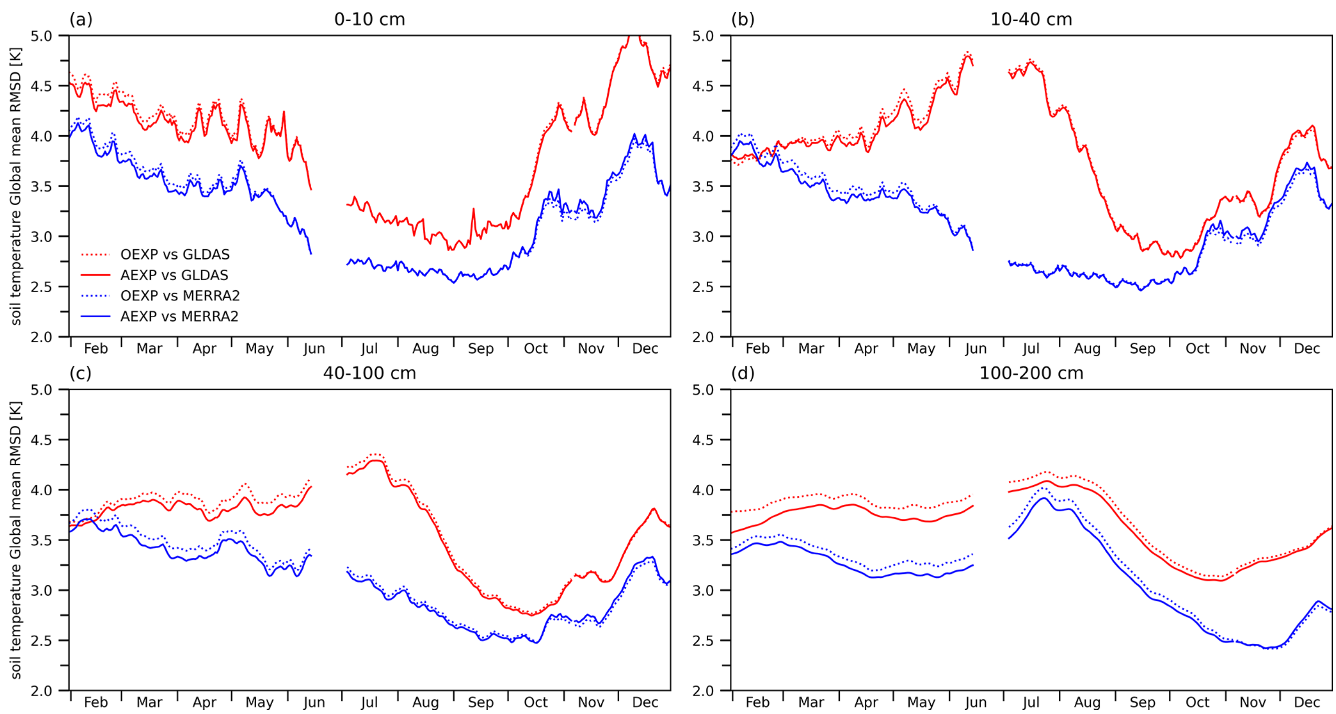

Figure 5. Soil temperature (in K) global RMSD of OEXP (dotted line) and AEXP (solid line) based on GLDAS (red) and MERRA2 (blue): **(a)** for 0–10 cm, **(b)** for 10–40 cm, **(c)** for 40–100 cm, and **(d)** for 100–200 cm.

Furthermore, the modified signals within the deep layers are not directly updated by assimilating MODIS-LST, but through the vertical exchange of water and energy with the upper layers. In other words, although the energy stored in the upper soil layers cannot persist in the upper layers, its influence can be transferred into the deeper layers through vertical diffusion between different soil layers. After MODIS-LST is assimilated into CoLM to update the upper soil temperature and water content, the changed signals within the upper layer are transferred to deeper layers, as the latter can store this signal for a relatively longer time. This result suggests promising potential for the assimilation scheme presented in this study, which utilizes MODIS-LST to jointly update soil temperature and moisture in the upper soil layer, after which the modified signals are transferred into deeper soil layers through the vertical diffusion by CoLM during the forecast phase.

The temperature of snow is vital for estimating both the energy and the water balance of the land surface layer. Figure 6 reveals the effectiveness of using MODIS-LST to update both snow temperature and water content consistently. It can be observed from Fig. 6a that CoLM exhibits relatively large RMSD of snow temperature based on ERA5-Land, centered in Northeast Asia and the northern regions of North America. Similar patterns are evident in Fig. 6d, g, and j for different seasons, with maximum uncertainties centered in the same areas as in Fig. 6a (results based on MERRA2 datasets are similar and not shown here). The largest RMSD values occur in February, indicating the challenges faced by CoLM in accurately representing snow temperature, particularly during winter for high-latitude regions (December also exhibits large RMSD values, not shown here). While snow temperature strongly depends on the water content, we attribute this unsatisfactory depiction of snow temperature to the deficient snow processes in LSM.

Compared with the open loop, improvements in snow temperature by assimilation are evident in Fig. 6b and c. Most regions exhibit improvements, with the maximum enhancements centered in Northeast Asia, similar to the patterns observed for soil temperature shown in Fig. 4b and c. This similarity suggests that snow water content effectively modulates the updated snow temperature, particularly during winter (Fig. 6e and f), with maximum $\mathrm{RMSD_d}$ values larger than 4 K in Northeast Asia for both ERA5-Land and MERRA2 datasets. This indicates that when the snow mass largely covers the soil layers, it insulates the soil beneath.

4.2.2 Improvements in Snow Properties and Soil Moisture Content

Unlike previous studies that used LST to update either soil temperature or soil water content, the assimilation scheme proposed in this study employs MODIS-LST to jointly update both soil temperature and water content in the upper layers through the framework of EnKF. Under this assimilation scheme, it is theoretically capable of producing consistent energy and water fields. Furthermore, one potential drawback of soil moisture assimilation is that the analysis increments only apply to soil moisture mass but cannot change its phase which depends on temperature. However, the joint update for both temperature and water content enables the water stored in snow or soil to change its phase between liquid and solid, especially in middle to high latitudes where the surface temperature is around the freezing/melting point. By simultaneously adjusting both state variables, this analysis allows water to redistribute between phases, yielding more realistic frozen/soil water partitions during the forecast phase.

The fraction of snow cover is a key parameter for indicating the temporal and spatial variation of snow water content. In high latitudes the fraction of snow cover may remain unchanged throughout the year due to the relatively low temperature below the melting point; the temporal variations of snow cover are considerably large in middle to high latitudes where the surface temperature range easily accommodates the condition for snow freezing or melting. In this study, we compare the snow cover from both experiments based on the ERA5-Land and MERRA2 datasets in terms of two global-scale metrics, and present the results in Fig. 7.

It can be inferred from Fig. 7a that the simulation of snow cover in CoLM has relatively large uncertainties in early summer and early winter (RMSD reaches 60 % and 40 % in May, based on ERA5-Land and MERRA2, respectively), revealing a poor simulation of snow melting and freezing. Relatively small uncertainties emerge in winter and summer months when the temporal variations of snow cover are relatively small. In Fig. 7b, both experiments reach their minimum R in summer when the snow merely remains in high-altitude and high-latitude regions. Comparison between two experiments demonstrates that the most significant enhancement is in spring and autumn regardless of the benchmark datasets (ERA5-Land and MERRA2). This improvement underscores the promising potential of assimilating MODIS-LST to adjust the snow cover in CoLM, particularly during periods of snow melting and freezing.

To further substantiate these findings with independent observations, the horizontal distribution of the yearly snow cover (in %) $\mathrm{RMSD_d}$ evaluated against three satellite datasets is presented in Fig. 8. As depicted, the assimilation of LST yields widespread improvements across the Northern Hemisphere, particularly over the mid-to-high latitudes of North America and Eurasia when evaluated against AVHRR (Fig. 8b) and MODIS (Fig. 8c), where most of these areas exhibit an RMSD reduction of approximately 10 %. While the evaluation against ATSR-2 (Fig. 8a) shows some regional degradations, the overall spatial patterns based on these three datasets robustly demonstrate the capability of the joint assimilation scheme to effectively constrain the surface thermal dynamics, thereby yielding a more physically consistent representation of snow accumulation and melting in the mid-to-high latitudes.

While snow cover is a key parameter indicating the fraction of snow within a grid cell, with strong temporal variations in middle-high latitudes and minimal disturbance in high latitudes, snow depth describes the cumulative effect of snow content within a grid cell. Changes in temperature, mass, and phase of water content all directly influence the simulation of snow depth from middle to high latitudes, making it a comprehensive parameter for estimating energy and water budgets related to snow content. Even excluding the permanent snow regions of Greenland and Antarctica, snow

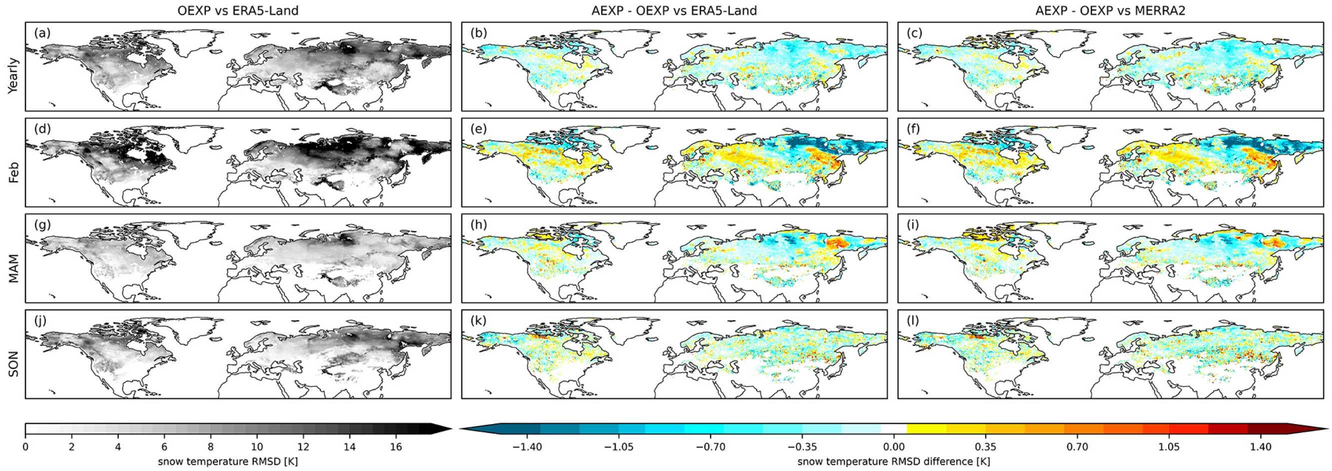

Figure 6. Snow temperature (in K) RMSD based on ERA5-Land (first column) of OEXP and $\mathrm{RMSD_d}$ based on ERA5-Land (second column) and MERRA2 (third column) for **(a–c)** yearly, **(d–f)** February, **(g–i)** MAM (March–April–May), and **(j–l)** SON (September–October–November).

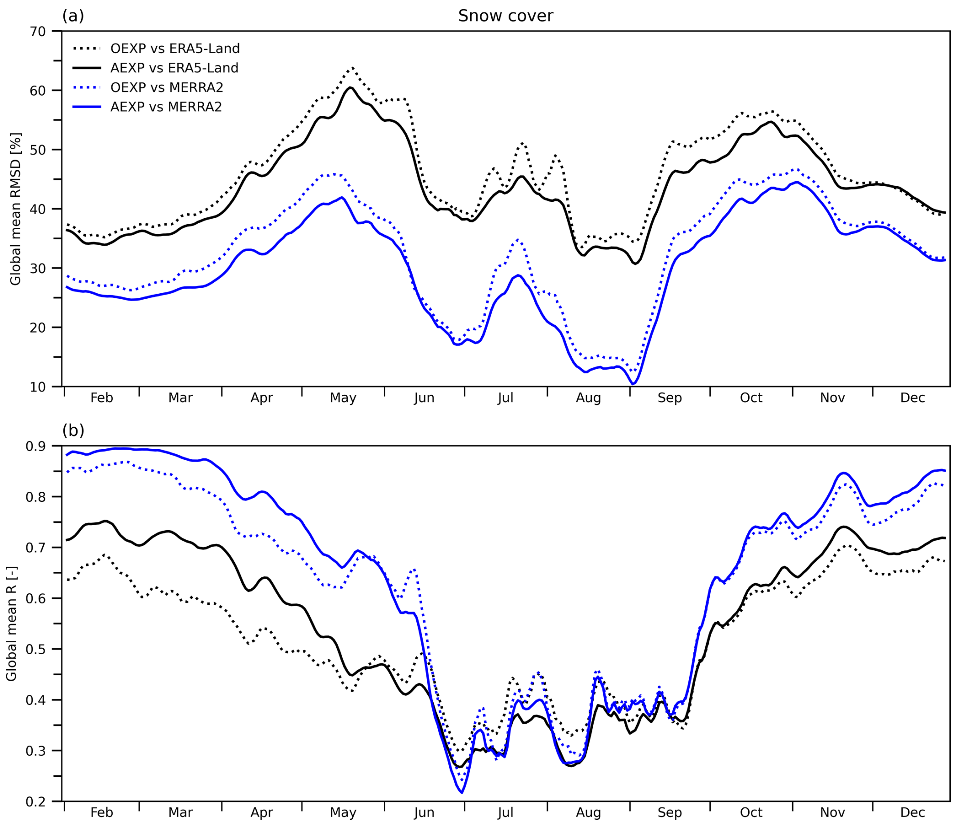

Figure 7. Snow cover (in %) global RMSD **(a)** and R **(b)** of OEXP (dotted) and AEXP (solid) based on ERA5-Land (black) and MERRA2 (blue).

depth still exhibits significant spatial variation across different continents. In this study, the spatial BIAS of snow depth based on the three datasets is provided in Fig. 9.

It can be seen from Fig. 9 that the BIAS for snow depth reaches a minimum in summer, as snow mass is relatively low when the Northern Hemisphere receives much solar radiation. The BIAS based on GLDAS and MERRA2 is consistent over either the globe (Fig. 9a) or different continents (Fig. 9b–d). In contrast, BIAS based on ERA5-Land exhibits a large discrepancy from those based on GLDAS

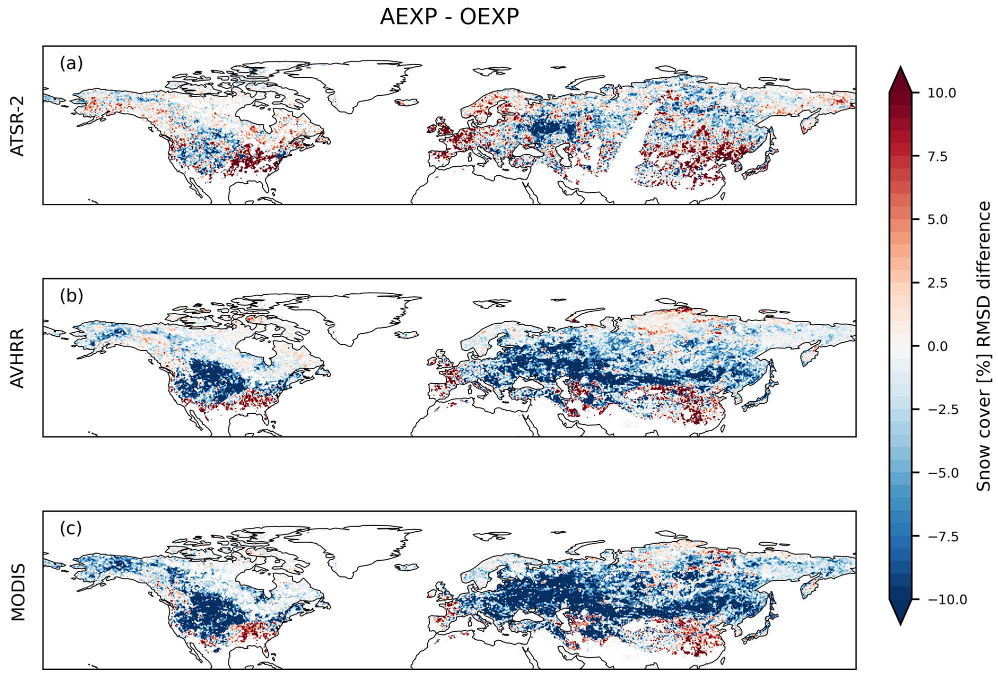

Figure 8. Snow cover (in %) yearly (31 January to 31 December 2001) $\mathrm{RMSD_d}$ evaluated based on different satellite products: **(a)** ATSR-2, **(b)** AVHRR, and **(c)** MODIS.

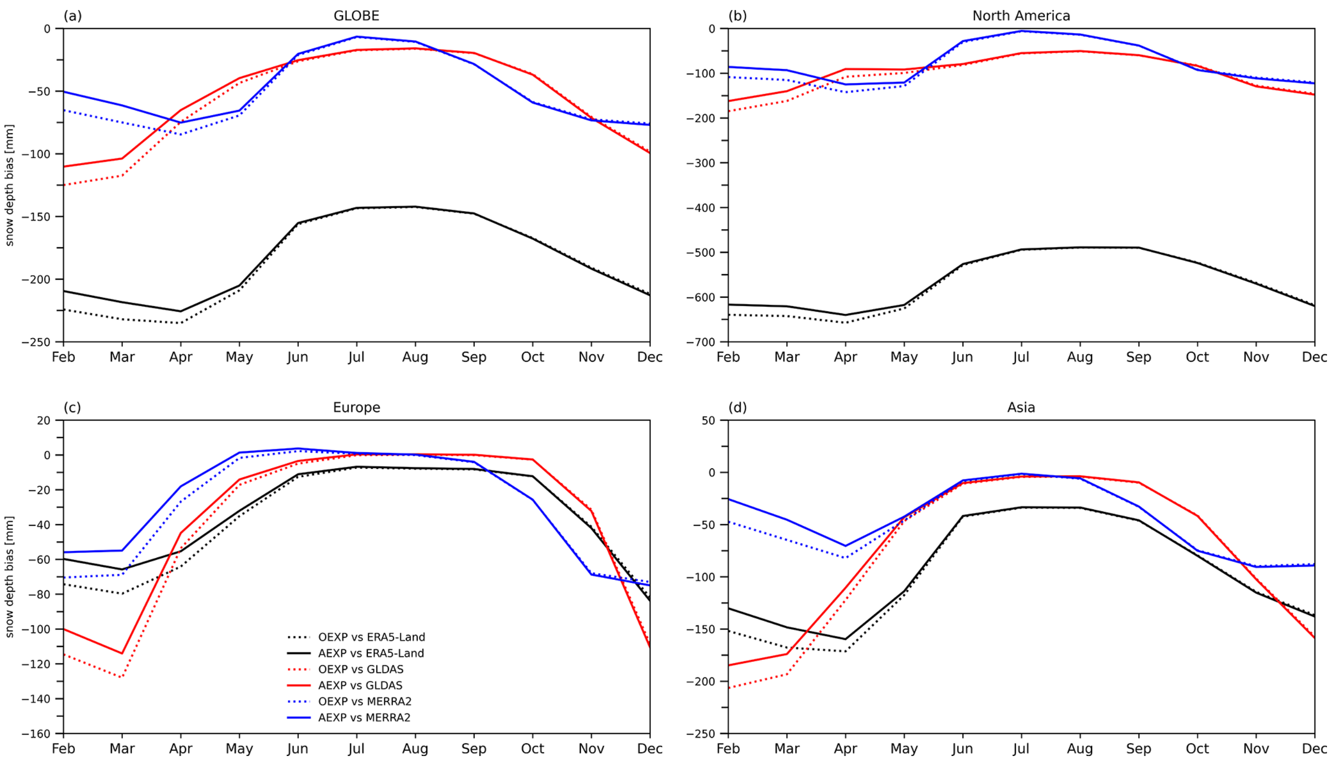

Figure 9. Snow depth (in mm) BIAS of OEXP (dotted) and AEXP (solid) based on ERA5-Land (black), GLDAS (red) and MERRA2 (blue). **(a)** For global excluding Antarctica, **(b)** for North America, **(c)** for Europe, and **(d)** for Asia.

and MERRA2 over the globe (Fig. 9a) and North America (Fig. 9b). This indicates that, over North America, the snow depth simulated by the LSM used in the ERA5-Land is vastly different from that by the CoLM and the LSMs used in the GLDAS and MERRA2. Regardless of the benchmark datasets, the large negative BIAS in simulating snow water content by the CoLM occurs in late winter and early spring, and so does the large improvement by assimilation for snow depth.

Taking into account the seasonal variations in snow depth, $RMSD_d$ is shown in Fig. 10 (summer time is not shown here due to the relatively small magnitude). In general, improvements are observed for most regions in the Northern Hemisphere throughout the year and different seasons, with the largest enhancements (up to 150 mm in February) centered in Northwest Asia. Consistent with snow temperature, the assimilation yields the largest RMSD reductions in winter (RMSD based on GLDAS reduces over 100 mm in northwest Asia as shown in Fig. 10e), with relatively small improvements in other seasons.

After discussing the results of the water content above the soil layers, the remainder of this section focuses on the water content (solid plus liquid phase) within the soil layers, as shown in Figs. 11 and 12. It can be observed from Fig. 11a and c that large uncertainties emerge in the Northeast regions of North America and Northern Eurasia, as most of these regions are considered humid areas with high soil moisture. The minimum uncertainties are observed in Northern Africa, the Middle East, and the Junggar Basin, as most of these regions are regarded as arid areas with low soil moisture. These results show that CoLM has a relatively adequate simulation of soil moisture in regions with low humidity but an insufficient simulation of soil moisture in regions with high humidity.

Improvements in unbiased RMSD occur in regions with relatively large uncertainty in the soil moisture simulation. Three maximum enhancements are observed in North America, Northern Eurasia, and Northeast Asia, as shown in Fig. 11b and d. The largest improvements, centered on northern Asia based on MERRA2 datasets, reach a 40 % reduction. It is also noteworthy that significant improvements in soil moisture within the upper soil layer emerge in high latitudes in the Northern Hemisphere, suggesting the effectiveness of the assimilation scheme, which jointly updates temperature and water content to obtain more realistic and consistent updated model fields for energy and water content. In other words, while the updated soil water content has a direct influence on the mass of water, the temperature change in the upper soil layers also drives the water content to change its phase. This joint modification of soil moisture is most pronounced at mid- to high latitudes.

A comparison of the two experiments in different soil layers based on the GLDAS datasets is shown in Fig. 12. As the depth increases, the absolute value of the (unbiased RMSD)$_d$ increases and reaches its maximum in the deepest soil layer within 100–200 cm, as shown in Fig. 12d. The maximum improvements are observed in northern regions of South America, central Africa, southern Asia, and northern Asia, where most regions are humid. Conversely, large degradations can be found in central areas of North America, eastern South America, and northern Australia, where most regions are relatively dry. Furthermore, localized degradations in Fig. 12 also occur in western North America and parts of northern Asia, where strong nonlinearity during seasonal freeze–thaw transitions makes the exact timing of phase changes difficult to capture. Additionally, performance degradation around open waters, such as the Great Lakes, is likely caused by the artificial alteration of land–lake thermal gradients. Updating only land grids while excluding water bodies disrupts localized secondary circulations, thereby weakening the assimilation benefits in these specific boundary regions.

Increasing improvements are seen as the soil depth increases, which reveals that the soil moisture modified by the assimilation within the upper layers is transferred down to the deep layers through vertical exchange among soil layers (taking the Amazon Rainforest as an example, the largest reduction in unbiased RMSD is about 0.06, 0.12, 0.15, and 6.58 $kg\,m^{-2}$ based on GLDAS datasets for soil layers of 0–10, 10–40, 40–100, and 100–200 cm, respectively). This conclusion aligns with those summarized from the soil temperature comparisons shown in Fig. 4. Unlike the surface layer, which has a strong connection with the above atmospheric conditions, the water stored in the deep layers can be regarded as a contribution from both the internal variation of soil water content within the layer and the accumulative effect of transferred modified signals from upper layers.

Furthermore, we attribute improvements in South America, central Africa, and southern Asia within the deepest soil layer to the high water conductivity in these humid areas, allowing for fast and efficient vertical transfer from the top soil layer to the bottom. In contrast, in central areas of North America, eastern South America, and northern Australia, it is challenging to transfer the modified information within the upper layers to the deep layers through vertical exchange owing to the relatively low water conductivity.

4.2.3 Modulations of Evaporation and Surface Heat Fluxes

Although evaporation and heat flux do not directly feedback on the prediction of land surface variables in offline LSMs, they provide key insights for studies in coupled models. Comparing these crucial fluxes against established datasets allows us to assess the model's performance and sheds light on the effectiveness of assimilation in improving the representation of land-atmosphere interactions.

Large RMSD values are observed in eastern North America, northern South America, central Africa, and central Asia for all three variables (Fig. 13). These regions are known for their high humidity, indicating relatively large uncertainties

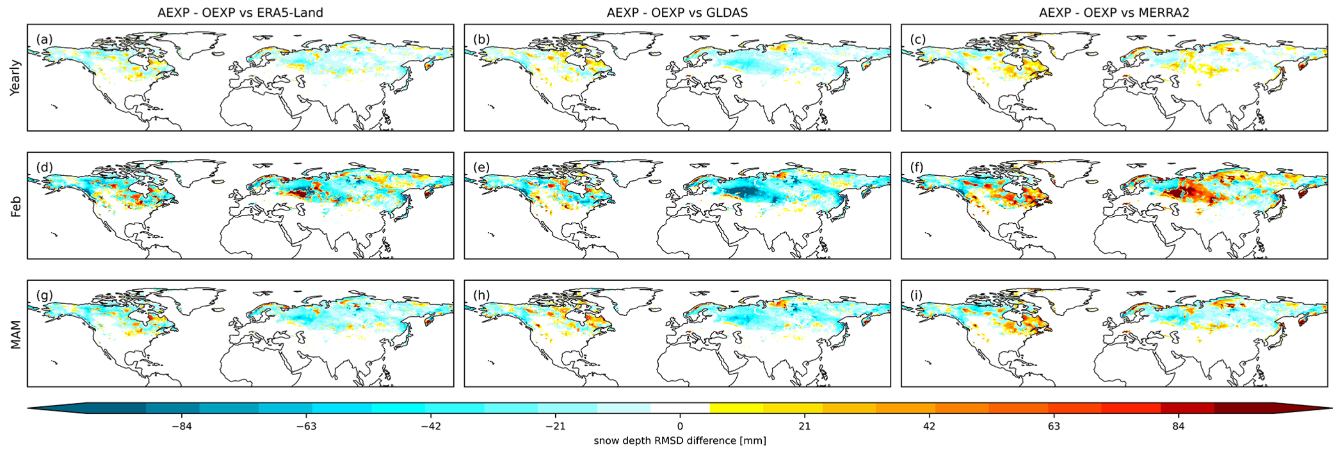

Figure 10. Snow depth (in mm) $RMSD_d$ based on ERA5-Land (first column), GLDAS (second column), and MERRA2 (third column). **(a–c)** For year, **(d–f)** for February and **(g–i)** for MAM (March–April–May).

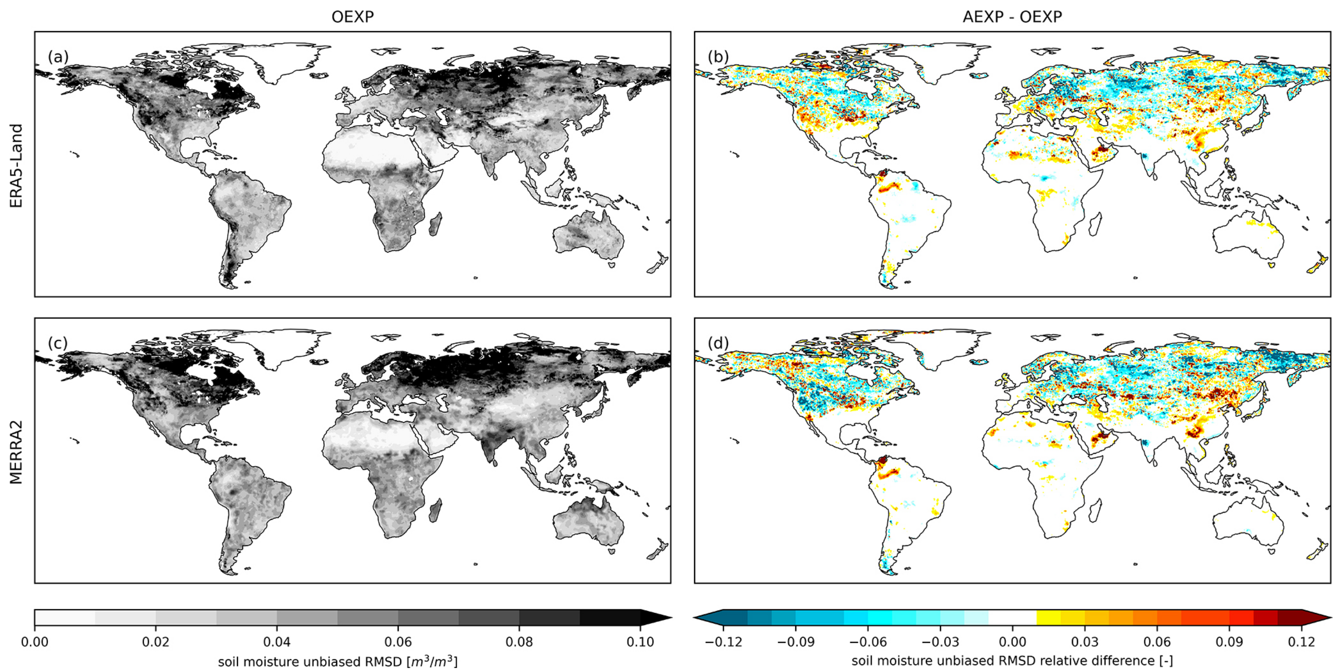

Figure 11. Soil moisture (in $m^3\,m^{-3}$) yearly (31 January to 31 December 2001) unbiased RMSD of OEXP (first column) and relative (unbiased $RMSD)_d$ (second column) in the upper soil layer based on **(a, b)** ERA5-Land and **(c, d)** MERRA2.

in the flux simulated by CoLM in humid regions. Conversely, there is relatively high confidence in areas with low humidity, such as northern Africa and the Middle East, as depicted in Fig. 13a, d, and h.

The Mediterranean region also exhibits a larger RMSD for evaporation from bare soil (Fig. 13a) compared with latent (Fig. 13d) and sensible (Fig. 13h) heat fluxes, which show no similarity of large RMSD. The RMSD in the Amazon Rainforest for evaporation (Fig. 13a) is not as significant as that for the latent heat flux (Fig. 13d). This difference between evaporation and latent heat flux can be attributed to the fact that evaporation from bare soil does not include transpiration from vegetation. There is abundant vegetation in the Mediterranean region and the Amazon Rainforest, where abundant vegetation enhances transpiration, thereby increasing latent heat flux while leaving bare-soil evaporation unchanged.

Obvious improvements can be found for the evaporative heat flux over western North America, the Tibetan Plateau, and Northeast Asia, as can be seen in Fig. 13b and c. The consistent pattern of $RMSD_d$ in soil moisture (Fig. 11b) and in evaporation (Fig. 13b) based on the ERA5-Land dataset can be attributed to the fact that the changes in soil moisture

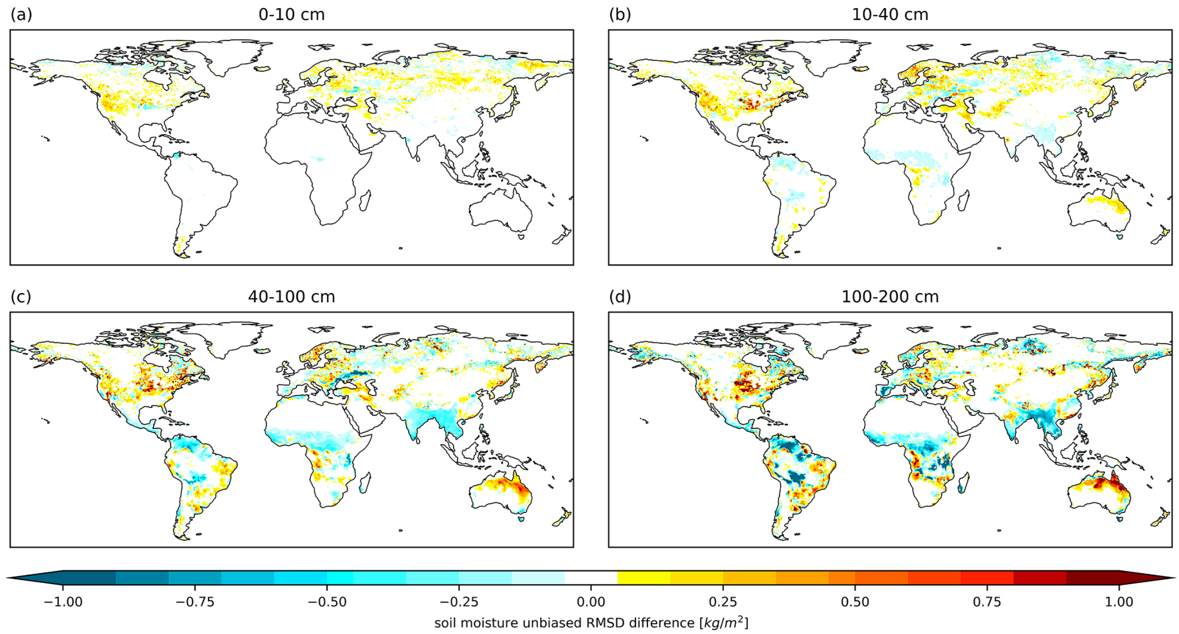

Figure 12. Soil moisture ($kg\,m^{-2}$) yearly (from 31 January to 31 December 2001) (unbiased RMSD)$_d$ based on GLDAS: **(a)** for 0–10 cm, **(b)** for 10–40 cm, **(c)** for 40–100 cm and **(d)** for 100–200 cm.

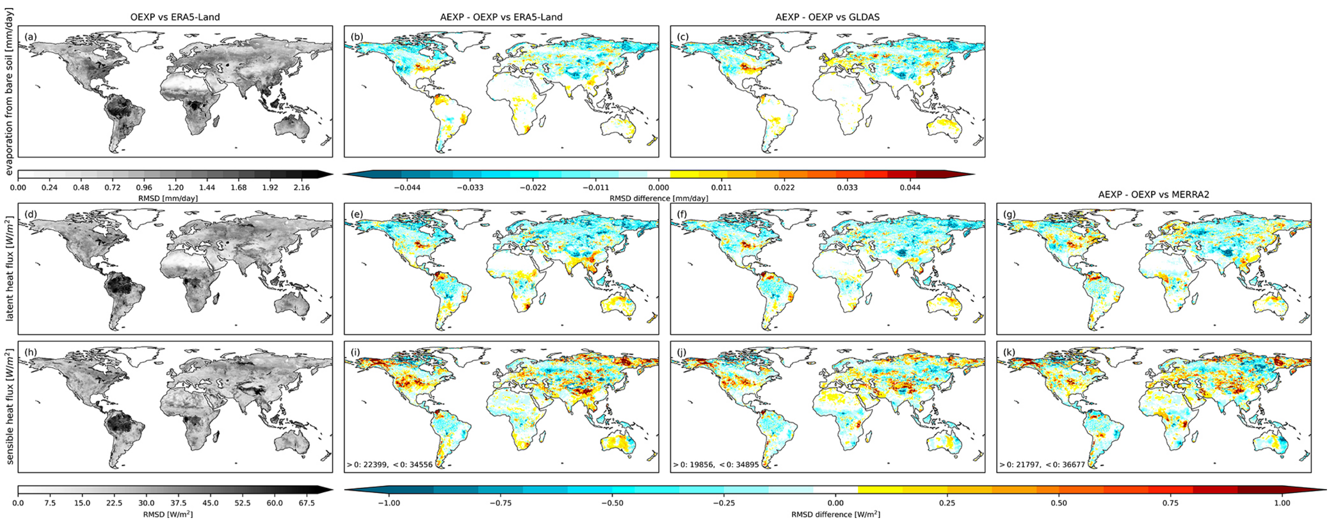

Figure 13. Yearly (from 31 January to 31 December 2001) RMSD based on ERA5-Land of OEXP (first column) and RMSD$_d$ based on ERA5-Land (second column), GLDAS (third column) and MERRA2 (fourth column) for evaporation from bare soil ($mm\,d^{-1}$, first row), latent heat flux ($W\,m^{-2}$, second row), and sensible heat flux ($W\,m^{-2}$, third row). MERRA2 does not provide evaporation from bare soil, so no corresponding panel is shown in the first row. The number of the positive and negative values of RMSD$_d$ for sensible heat flux is shown in **(i)**, **(j)**, and **(k)**.

within the upper soil layers have a direct influence on the amount of evaporation. This consistency is also found in the comparison between Fig. 12a and Fig. 13c based on GLDAS datasets.

The large improvements in Fig. 13b and c for evaporation flux from the bare soil, are also observed over similar regions in Fig. 13e–g for latent heat flux regardless of the benchmark datasets. Additionally, considerable improvements in the Amazon Rainforest are consistent with the findings aforementioned, and highlight the importance of considering transpiration from vegetation.

Large degradations are observed in Alaska, western regions of North America, the Tibetan Plateau, and Northeast Asia, while considerable improvements are shown in the Amazon Rainforest, central Africa, and eastern Asia for sensible heat flux. The partition of the available energy between latent and sensible heat fluxes depends strongly on the soil moisture. Taking Fig. 13e and i based on the ERA5-Land dataset as an example, improvements are observed for both fluxes in regions with high humidity, such as the Amazon Rainforest and central Africa. In contrast, opposite patterns are observed in regions with relatively low humidity, such as western North America and Northeast Asia. This highlights that assimilation-induced changes in soil moisture can simultaneously modulate both fluxes, depending on the prevailing surface energy partitioning.

To further validate the assimilation impact on surface energy partitioning against independent ground truth, in-situ observations of sensible and latent heat fluxes from the FLUXNET2015 and AmeriFlux networks are incorporated. The corresponding site-specific evaluation is presented in Fig. 14. As illustrated in the spatial distribution (Fig. 14a), the assimilation scheme yields clear improvements at the majority of the evaluated sites globally. Most locations exhibit enhanced performance in both heat fluxes (green circles) or at least one flux variable (orange triangles). Quantitatively, the site-specific $\mathrm{RMSD_d}$ (Fig. 14b) reveals consistent and often substantial error reductions for the latent heat flux across most stations. In contrast, the response of the sensible heat flux exhibits greater site-to-site variability, which aligns with the mixed regional patterns observed in the gridded comparisons.

5 Discussion

The most notable soil temperature difference between the assimilation and the open loop occurs in permafrost regions of Northeast Asia (Fig. 4b). Within this region the three reanalyses also differ markedly. GLDAS agrees well with MERRA2, whereas ERA5-Land yields substantially different soil temperature magnitudes. The $\mathrm{BIAS_d}$ based on ERA5-Land is positive globally (figures similar to Fig. 4b, e, h, c, f, and i are omitted). These positive values suggest degradation in soil temperature estimation worldwide. Given that the soil temperature from the ERA5-Land dataset is known to carry a relatively large bias in permafrost regions (Cao et al., 2020), we exclude it from the evaluation process, as indicated in Table 2.

Wide discrepancies among the third-party datasets remain, most notably for snow depth, as shown in Fig. 9a. While the BIAS global mean based on GLDAS and MERRA2 is around −75 mm, that based on ERA5-Land datasets is approximately −200 mm. While the $\mathrm{RMSD_d}$ based on ERA5-Land is nearly identical to that based on GLDAS (Fig. 10d and e), it is opposite to that based on MERRA2, which indicates degradation in central Eurasia and eastern regions of North America (Fig. 10f).

Despite the notable differences observed for certain variables among different third-party datasets, it is reassuring to observe relatively consistent results across different datasets for most variables on a global scale. This consistency underscores the importance of utilizing diverse datasets in the evaluation process. By incorporating multiple datasets, the evaluation results become more robust, providing a confident understanding of the performance of both the CoLM and DA method and reducing the impact of biases or uncertainties inherent in any single dataset. Furthermore, the performance enhancements yielded by this study are generally comparable to relevant previous studies. Huang et al. (2008) assimilated LST to update soil temperature, reporting an error reduction of approximately 1 K at specific sites. This aligns with our results, which show a bias reduction of 1.0–2.0 K with increasing depth over Northeast Asia (Fig. 4). Furthermore, while Chen et al. (2021) assimilated both LST and brightness temperature to achieve a 19 %–71 % RMSD reduction for shallow soil moisture on the Tibetan Plateau, the LST-only assimilation in this study yields a comparable impact, reducing the unbiased RMSD for upper-layer soil moisture by up to 40 % in Northern Asia (Fig. 11).

While the assimilation scheme proposed in this study does not take the bias correction (Bosilovich et al., 2007; Reichle et al., 2010) into account, it is noteworthy that one of the biggest challenges faced in this study is not the bias in either model forecasts or observations. Since the land surface processes are strongly influenced by the overlying atmosphere, the forecasts of an offline LSM are almost determined by the atmospheric forcing rather than the initial conditions. Several studies have shown unrealistically small ensemble spread in the land surface variables (Muñoz-Sabater et al., 2021), such as the soil moisture content (Draper, 2021), in an offline LSM. Despite this inherent limitation, diagnostics of spread indicate that the ensemble members dynamically evolve without collapsing. Taking the Northern Hemisphere as an example, the temperature spread in the first soil layer is relatively low, ranging from 0.3 to 1.4 K, owing to the strong atmospheric forcing. This expected limitation further justifies the joint update of soil temperature and soil moisture. Meanwhile, the soil moisture spread increases from approximately $0.5\,\mathrm{kg\,m^{-2}}$ at the surface to $10.0\,\mathrm{kg\,m^{-2}}$ in the deepest layer, accompanied by physically consistent seasonal variations.

Direct inflation of the analysis covariance is not used in this study. In the forecast phase, the CoLM perturbs its tunable parameters to generate ensemble forecasts as the background for next cycle, with the purpose of not only accounting for the model error but also inflating the covariance. Physically, this tuning strategy modulates six core land surface processes: surface turbulent exchange, canopy and snow hydrology, runoff generation, alongside soil thermal and water dynamics. Perturbing these specific mechanisms is essential for LST assimilation, as the simulated LST is predominantly sensitive to parameters governing the surface energy balance and outgoing radiation. Furthermore, parameters regulating energy storage and water retention directly dictate the hydro-thermal cross-covariance, which ensures a

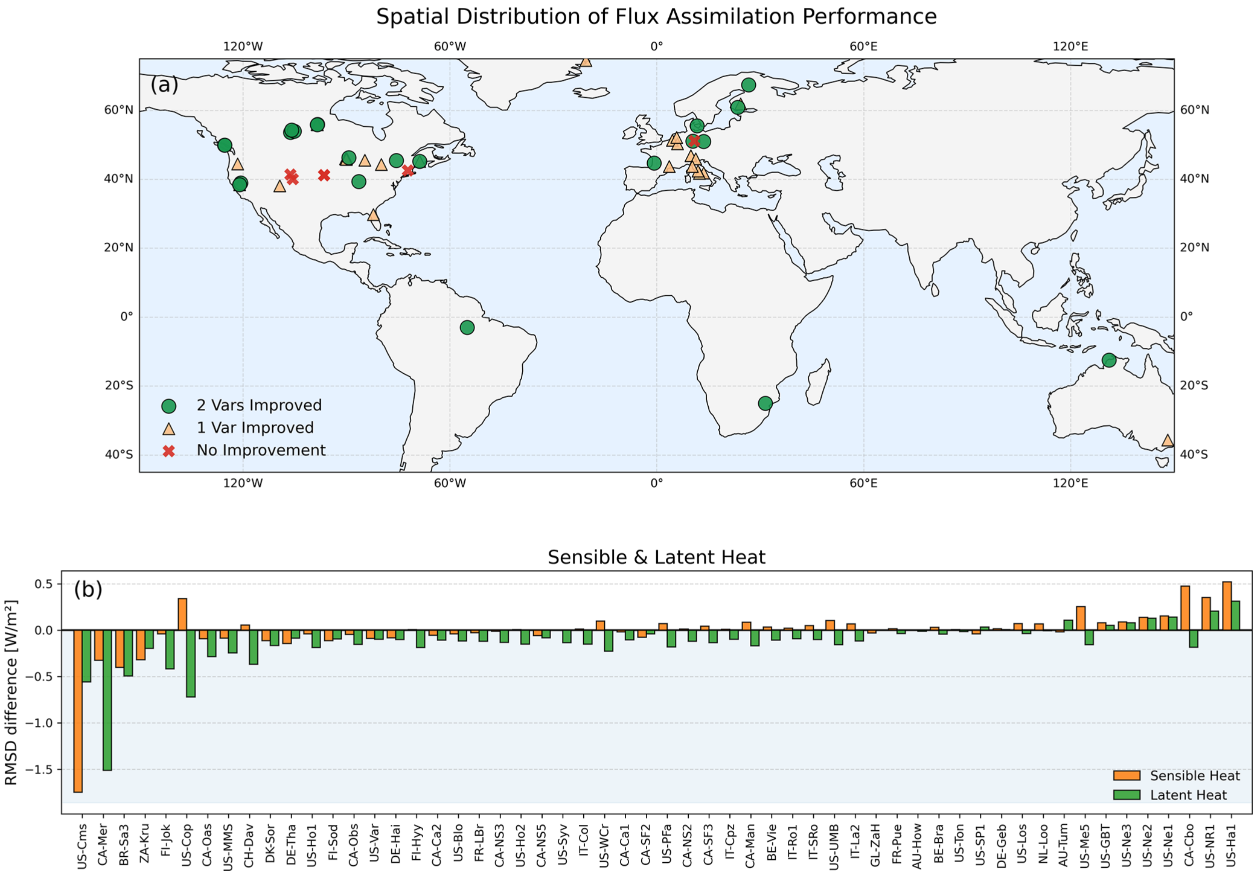

Figure 14. Spatial and quantitative evaluation of the impact of assimilation on surface heat fluxes across global sites. **(a)** Distribution of assimilation performance based on sensible heat and latent heat. Markers categorize sites by the number of improved flux variables: green circles indicate improvement in both heat fluxes, orange triangles represent improvement in a single variable, and red crosses denote no improvement. **(b)** Site-specific $\mathrm{RMSD_d}$ for sensible heat (orange bars) and latent heat (green bars) across the year (31 January to 31 December 2001).

physically consistent joint update of temperature and moisture. The parameters are perturbed up to 15 % of the corresponding parameter's absolute value (see Sect. 2.1), which is neither too small for the aforementioned reason nor too large for model stability. Although the improvements during the free prediction time (Fig. 3a), it is noticeable that the difference between the assimilation and the open loop tends to decrease as time increases, as shown in Figs. 5b–d and 9.

6 Conclusions

In this study, the LETKF is applied to the CoLM for the first time. MODIS-LST from the Terra satellite is assimilated throughout 2001, followed by a 2-month free forecast. Unlike previous studies that used LST to update either the temperature or the water content within the soil layers, the assimilation scheme proposed in this study jointly updates both soil temperature and water content in the upper layers to prolong the influence of the LST assimilation. On the one hand, the jointly updated temperature and water content by assimilating MODIS-LST have a relatively larger capacity for storing and retaining the influence from LST. On the other hand, it is theoretically more likely to produce both energy and water fields in a consistent manner, especially for middle to high latitudes where the surface temperature is around the freezing or melting point, which enables the updated temperature within the soil and snow layer to change the phase of water content. Additional experiments demonstrate that this joint update outperforms updating only soil temperature, improving both the soil water and energy states (Fig. 1) and also snow-related variables (results not shown here). The evaluation based on the MODIS-LST shows a relatively small enhancement, consistent with the nature of LST which is strongly determined by the atmospheric forcing in an offline LSM. Considerable improvements are seen for other variables based on three benchmark datasets, such as the soil temperature, with the largest BIAS reduction centered in Northeast Asia and its magnitude increasing as the depth increases (about 1.0, 1.5, and 2.0 K for 0–10, 40–100, and 100–200 cm, respectively), indicating the signals induced by assimilating LST can be transferred to and persist longer in deeper soil layers. Similar to soil temperature, the maximum improvements in snow temperature and snow depth also occur over Northeast Asia, with a reduction of RMSD of about 4 K and 150 mm, respectively. This improvement over middle to high latitudes indicates the effectiveness of the joint update on the temperature and water content, which is allowed to change its phase depending on the changed temperature. Furthermore, improvements in soil water content are seen, especially in the humid regions within the tropics; taking the Amazon Rainforest as an example, the largest reduction in unbiased RMSD is about 0.06, 0.12, 0.15, and 6.58 $kg\,m^{-2}$ for depths of 0–10, 10–40, 40–100, and 100–200 cm, respectively, indicating the high water conductivity in these humid areas, which allows for fast and efficient vertical transfer from the top soil layer to the bottom.

The challenge of maintaining a realistic ensemble spread for ensemble assimilation into offline LSMs remains open. Thus, it might be more effective for ensemble-based land surface DA to use a coupled model with two-way coupling between the land and atmosphere. Assimilating satellite radiance observations in the atmospheric window into a land-atmosphere two-way coupling model has been undergoing investigation and will be presented in another paper.

Code and data availability. The WFDE5 atmospheric forcing data are available at https://doi.org/10.24381/cds.20d54e34 (Copernicus Climate Change Service, Climate Data Store, 2025b). The MODIS-LST dataset is available at https://doi.org/10.5285/58a01734f841466daa1837353aee5ff8 (ESA Land Surface Temperature Climate Change Initiative, 2025). The ERA5-Land, GLDAS, and MERRA2 datasets for evaluation are available at https://doi.org/10.24381/cds.e2161bac (Copernicus Climate Change Service, Climate Data Store, 2025a), https://doi.org/10.5067/E7TYRXPJKWOQ (NASA Global Land Data Assimilation System, 2025), and https://doi.org/10.5067/RKPHT8KC1Y1T (NASA Global Modeling and Assimilation Office (GMAO), 2025), respectively. The Common Land Model (CoLM) version used in this study is archived at Zenodo: https://doi.org/10.5281/zenodo.18649912 (The Common Land Model (CoLM), 2025). The LETKF code developed for the CoLM DA is hosted at Zenodo: https://doi.org/10.5281/zenodo.18649772 (Fu and Zheng, 2025a). The datasets and plotting scripts for generating the figures in this study are available at Zenodo: https://doi.org/10.5281/zenodo.17284395 (Fu and Zheng, 2025b).

Author contributions. Conceptualization, Y.Z.; methodology, Y.Z.; software, Y.F. and Y.Z.; validation, Y.F. and Y.Z.; formal analysis, Y.F.; investigation, Y.F. and Y.Z.; resources, Y.Z. and J.L.; data curation, Y.F. and Y.Z.; writing – original draft preparation, Y.F.; writing – review and editing, Y.Z. and Y.F.; visualization, Y.F.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z.; all authors have read and agreed to the published version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This research was funded by the National Natural Science Foundation of China (grant nos. 12241104 and 42275161) and the State Key Laboratory of Climate System Prediction and Risk Management (CPRM) initiative project (grant no. CPRM-2025-NUIST-012). The authors sincerely thank the editor Dr. Nathaniel Chaney, reviewer Dr. Lu Li, and the two anonymous reviewers for their insightful feedback and constructive suggestions, which significantly improved the manuscript.

Financial support. This research has been supported by the National Natural Science Foundation of China (grant nos. 12241104 and 42275161) and the State Key Laboratory of Climate System Prediction and Risk Management (CPRM) initiative project (grant no. CPRM-2025-NUIST-012).

Review statement. This paper was edited by Nathaniel Chaney and reviewed by Lu Li and two anonymous referees.

References

Anderson, J. L.: An Ensemble Adjustment Kalman Filter for Data Assimilation, Mon. Weather Rev., 129, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2, 2001.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, Mon. Weather Rev., 127, 2741–2758, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, 1999.

Avissar, R.: Conceptual Aspects of a Statistical-dynamical Approach to Represent Landscape Subgrid-scale Heterogeneities in Atmospheric Models, J. Geophys. Res.-Atmos., 97, 2729–2742, https://doi.org/10.1029/91JD01751, 1992.

Balsamo, G., Mahfouf, J.-F., Bélair, S., and Deblonde, G.: A Land Data Assimilation System for Soil Moisture and Temperature: An Information Content Study, J. Hydrometeorol., 8, 1225–1242, https://doi.org/10.1175/2007JHM819.1, 2007.

Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Noah Land Surface Model L4 3 Hourly 0.25 × 0.25 Degree, Version 2.1, https://doi.org/10.5067/E7TYRXPJKWOQ, 2020.

Bishop, C., Whitaker, J., and Lei, L.: Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization, Mon Weather Rev., 145, 4575–4592, https://doi.org/10.1175/MWR-D-17-0102.1, 2017.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Weather Rev., 129, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Bonan, B., Albergel, C., Zheng, Y., Barbu, A. L., Fairbairn, D., Munier, S., and Calvet, J.-C.: An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region, Hydrol. Earth Syst. Sci., 24, 325–347, https://doi.org/10.5194/hess-24-325-2020, 2020.

Bonan, G.: A Land Surface Model (LSM Version 1.0) for Ecological, Hydrological, and Atmospheric Studies: Technical Description and User's Guide, Tech. rep., UCAR/NCAR, https://doi.org/10.5065/D6DF6P5X, 1996.

Bonavita, M., Torrisi, L., and Marcucci, F.: Ensemble Data Assimilation with the CNMCA Regional Forecasting System, Q. J. Roy. Meteor. Soc., 136, 132–145, https://doi.org/10.1002/qj.553, 2010.

Bosilovich, M. G., Radakovich, J. D., Da Silva, A., Todling, R., and Verter, F.: Skin Temperature Analysis and Bias Correction in a Coupled Land-Atmosphere Data Assimilation System, J. Meteorol. Soc. Jpn. Ser. II, 85A, 205–228, https://doi.org/10.2151/jmsj.85A.205, 2007.

Browne, P., De Rosnay, P., Zuo, H., Bennett, A., and Dawson, A.: Weakly Coupled Ocean–Atmosphere Data Assimilation in the ECMWF NWP System, Remote Sens.-Basel, 11, 234, https://doi.org/10.3390/rs11030234, 2019.

Buehner, M.: Local Ensemble Transform Kalman Filter with Cross Validation, Mon. Weather Rev., 148, 2265–2282, https://doi.org/10.1175/MWR-D-19-0402.1, 2020.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, Mon. Weather Rev., 126, 1719–1724, https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Cao, B., Gruber, S., Zheng, D., and Li, X.: The ERA5-Land soil temperature bias in permafrost regions, The Cryosphere, 14, 2581–2595, https://doi.org/10.5194/tc-14-2581-2020, 2020.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M., and Betts, A.: Modeling of Land Surface Evaporation by Four Schemes and Comparison with FIFE Observations, J. Geophys. Res.-Atmos., 101, 7251–7268, https://doi.org/10.1029/95JD02165, 1996.

Chen, W., Huang, C., Yang, Z.-L., and Zhang, Y.: Retrieving Accurate Soil Moisture over the Tibetan Plateau Using Multisource Remote Sensing Data Assimilation with Simultaneous State and Parameter Estimations, J. Hydrometeorol., 22, 2751–2766, https://doi.org/10.1175/JHM-D-20-0298.1, 2021.

Copernicus Climate Change Service: ERA5-Land Hourly Data from 1950 to Present, Copernicus Climate Change Service [data set], https://doi.org/10.24381/CDS.E2161BAC, 2019.

Copernicus Climate Change Service, Climate Data Store: ERA5-Land hourly data from 1950 to present, Copernicus Climate Change Service, Climate Data Store [data set], https://doi.org/10.24381/cds.e2161bac, 2025a.

Copernicus Climate Change Service, Climate Data Store: Near surface meteorological variables from 1979 to 2019 derived from bias-corrected reanalysis, Copernicus Climate Change Service, Climate Data Store [data set], https://doi.org/10.24381/cds.20d54e34, 2025b.

Crow, W. T. and Wood, E. F.: The Assimilation of Remotely Sensed Soil Brightness Temperature Imagery into a Land Surface Model Using Ensemble Kalman Filtering: A Case Study Based on ESTAR Measurements during SGP97, Adv. Water Resour., 26, 137–149, https://doi.org/10.1016/S0309-1708(02)00088-X, 2003.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data, 12, 2097–2120, https://doi.org/10.5194/essd-12-2097-2020, 2020.

Dai, Y. and Zeng, Q.: A Land Surface Model (IAP94) for Climate Studies Part I: Formulation and Validation in off-Line Experiments, Adv. Atmos. Sci., 14, 433–460, https://doi.org/10.1007/s00376-997-0063-4, 1997.

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.-L.: The Common Land Model, B. Am. Meteorol. Soc., 84, 1013–1024, https://doi.org/10.1175/BAMS-84-8-1013, 2003.

Dai, Y., Dickinson, R. E., and Wang, Y.-P.: A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance, J. Climate, 17, 2281–2299, https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2, 2004.

De Lannoy, G. J. M., Bechtold, M., Albergel, C., Brocca, L., Calvet, J.-C., Carrassi, A., Crow, W. T., De Rosnay, P., Durand, M., Forman, B., Geppert, G., Girotto, M., Hendricks Franssen, H.-J., Jonas, T., Kumar, S., Lievens, H., Lu, Y., Massari, C., Pauwels, V. R. N., Reichle, R. H., and Steele-Dunne, S.: Perspective on Satellite-Based Land Data Assimilation to Estimate Water Cycle Components in an Era of Advanced Data Availability and Model Sophistication, Frontiers in Water, 4, 981745, https://doi.org/10.3389/frwa.2022.981745, 2022.

De Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L.: A Simplified Extended Kalman Filter for the Global Operational Soil Moisture Analysis at ECMWF, Q. J. Roy. Meteor. Soc., 139, 1199–1213, https://doi.org/10.1002/qj.2023, 2013.

De Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J., and Isaksen, L.: Initialisation of Land Surface Variables for Numerical Weather Prediction, Surv. Geophys., 35, 607–621, https://doi.org/10.1007/s10712-012-9207-x, 2014.

Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P. J.: Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. Technical Note [NCAR (National Center for Atmospheric Research)], Tech. rep., National Center for Atmospheric Research, Scientific Computing Div., Boulder, CO (United States), https://doi.org/10.5065/D67W6959, 1993.

Draper, C. S.: Accounting for Land Model Error in Numerical Weather Prediction Ensemble Systems: Toward Ensemble-Based Coupled Land/Atmosphere Data Assimilation, J. Hydrometeorol., https://doi.org/10.1175/JHM-D-21-0016.1, 2021.

Drusch, M.: Initializing Numerical Weather Prediction Models with Satellite-derived Surface Soil Moisture: Data Assimilation Experiments with ECMWF's Integrated Forecast System and the TMI Soil Moisture Data Set, J. Geophys. Res.-Atmos., 112, 2006JD007478, https://doi.org/10.1029/2006JD007478, 2007.

Entekhabi, D., Nakamura, H., and Njoku, E.: Solving the Inverse Problem for Soil Moisture and Temperature Profiles by Sequential Assimilation of Multifrequency Remotely Sensed Observations, IEEE T. Geosci. Remote, 32, 438–448, https://doi.org/10.1109/36.295058, 1994.

ESA Land Surface Temperature Climate Change Initiative: Land Surface Temperature from MODIS (Moderate resolution Infra-red Spectroradiometer) on Terra, level 3 collated (L3C) global product (2000–2018), version 3.00, https://doi.org/10.5285/58a01734f841466daa1837353aee5ff8, data set, 2025.

Evensen, G.: Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics, J. Geophys. Res.-Oceans, 99, 10143–10162, https://doi.org/10.1029/94JC00572, 1994.

Evensen, G.: The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dynam., 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Farchi, A. and Bocquet, M.: On the Efficiency of Covariance Localisation of the Ensemble Kalman Filter Using Augmented Ensembles, Frontiers in Applied Mathematics and Statistics, 5, 3, https://doi.org/10.3389/fams.2019.00003, 2019.

Fu, S., Nie, S., Luo, Y., and Chen, X.: Implications of Diurnal Variations in Land Surface Temperature to Data Assimilation Using MODIS LST Data, J. Geogr. Sci., 30, 18–36, https://doi.org/10.1007/s11442-020-1712-0, 2020.

Fu, Y. and Zheng, Y.: Local Ensemble Transform Kalman Filter (LETKF) adapted to Common Land Model (CoLM), inherited from UMD-LETKF, Zenodo [source code], https://doi.org/10.5281/zenodo.18649772, 2025a.

Fu, Y. and Zheng, Y.: Results from "Effectively Assimilate Satellite Land Surface Temperature into Offline Land Surface Models within Ensemble-based Assimilation Frameworks", Zenodo [data set], https://doi.org/10.5281/zenodo.17284395, 2025b.

Gaspari, G. and Cohn, S. E.: Construction of Correlation Functions in Two and Three Dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Ghent, D., Ermida, S., Jimenez, C., and Dodd, E.: LST_cci Product User Guide v2.0, European Space Agency, https://admin.climate.esa.int/documents/1557/LST-CCI-D4.3-PUG_-_i2r0_-_Product_User_Guide.pdf (last access: 25 September 2026), 2021.

Ghent, D., Veal, K., and Perry, M.: ESA Land Surface Temperature Climate Change Initiative (LST_cci): Land Surface Temperature from MODIS (Moderate Resolution Infra-red Spectroradiometer) on Terra, Level 3 Collated (L3C) Global Product (2000–2018), Version 3.00, https://doi.org/10.5285/58A01734F841466DAA1837353AEE5FF8, 2022.

Ghent, D., Dodd, E., Veal, K., Perry, M., Jimenez, C., and Ermida, S.: LST_cci Algorithm Theoretical Basis Document v1.2, European Space Agency, https://admin.climate.esa.int/documents/1711/LST-CCI-D2.2-ATBD_-_i3r0_-_Algorithm_Theoretical_Basis_Document.pdf (last access: 25 September 2026), 2023.

Global Climate Observing System: Essential Climate Variables, https://gcos.wmo.int/site/global-climate-observing-system-gcos/essential-climate-variables/land-surface-temperature (last access: 25 September 2026), 2022.

Global Modeling and Assimilation Office and Pawson, S.: MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://doi.org/10.5067/RKPHT8KC1Y1T, 2015.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, Mon. Weather Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hamrud, M., Bonavita, M., and Isaksen, L.: EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation, Mon. Weather Rev., 143, 4847–4864, https://doi.org/10.1175/MWR-D-14-00333.1, 2015.

Han, X., Franssen, H.-J. H., Montzka, C., and Vereecken, H.: Soil Moisture and Soil Properties Estimation in the Community Land Model with Synthetic Brightness Temperature Observations, Water Resour. Res., 50, 6081–6105, https://doi.org/10.1002/2013WR014586, 2014.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, Mon. Weather Rev., 126, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Mon. Weather Rev., 144, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1, 2016.

Huang, C., Li, X., and Lu, L.: Retrieving Soil Temperature Profile by Assimilating MODIS LST Products with Ensemble Kalman Filter, Remote Sens. Environ., 112, 1320–1336, https://doi.org/10.1016/j.rse.2007.03.028, 2008.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter, Physica D, 230, 112–126, https://doi.org/10.1016/j.physd.2006.11.008, 2007.

Ji, D. and Dai, Y.: The Common Land Model (CoLM) Technical Guide, http://globalchange.bnu.edu.cn/download/doc/CoLM/CoLM_Technical_Guide.pdf (last access: 25 September 2026), 2010.

Jin, M. and Dickinson, R. E.: New Observational Evidence for Global Warming from Satellite, Geophys. Res. Lett., 29, https://doi.org/10.1029/2001GL013833, 2002.

Jin, M. and Dickinson, R. E.: Land Surface Skin Temperature Climatology: Benefitting from the Strengths of

Satellite Observations, Environ. Res. Lett., 5, 044004, https://doi.org/10.1088/1748-9326/5/4/044004, 2010.

Kalnay, E. and Cai, M.: Impact of Urbanization and Land-Use Change on Climate, Nature, 423, 528–531, https://doi.org/10.1038/nature01675, 2003.

Khaki, M., Hendricks Franssen, H.-J., and Han, S. C.: Multi-Mission Satellite Remote Sensing Data for Improving Land Hydrological Models via Data Assimilation, Sci. Rep.-UK, 10, 18791, https://doi.org/10.1038/s41598-020-75710-5, 2020.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J. M.: A Parameterization of Snowpack and Frozen Ground Intended for NCEP Weather and Climate Models, J. Geophys. Res.-Atmos., 104, 19569–19585, https://doi.org/10.1029/1999JD900232, 1999.

Koster, R. D. and Suarez, M. J.: Modeling the Land Surface Boundary in Climate Models as a Composite of Independent Vegetation Stands, J. Geophys. Res.-Atmos., 97, 2697–2715, https://doi.org/10.1029/91JD01696, 1992.

Koster, R. D. and Suarez, M. J.: Soil Moisture Memory in Climate Models, J. Hydrometeorol., 2, 558–570, https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2, 2001.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A Catchment-based Approach to Modeling Land Surface Processes in a General Circulation Model: 1. Model Structure, J. Geophys. Res.-Atmos., 105, 24809–24822, https://doi.org/10.1029/2000JD900327, 2000.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of Strong Coupling Between Soil Moisture and Precipitation, Science, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

Kurosawa, K., Kotsuki, S., and Miyoshi, T.: Comparative study of strongly and weakly coupled data assimilation with a global land–atmosphere coupled model, Nonlin. Processes Geophys., 30, 457–479, https://doi.org/10.5194/npg-30-457-2023, 2023.

Lakshmi, V.: A Simple Surface Temperature Assimilation Scheme for Use in Land Surface Models, Water Resour. Res., 36, 3687–3700, https://doi.org/10.1029/2000WR900204, 2000.

Mahfouf, J.-F.: Analysis of Soil Moisture from Near-Surface Parameters: A Feasibility Study, J. Appl. Meteorol. Clim., 30, 1534–1547, https://doi.org/10.1175/1520-0450(1991)030<1534:AOSMFN>2.0.CO;2, 1991.

McLaughlin, D.: Recent Developments in Hydrologic Data Assimilation, Rev. Geophys., 33, 977–984, https://doi.org/10.1029/95RG00740, 1995.

Meng, C. L., Li, Z.-L., Zhan, X., Shi, J. C., and Liu, C. Y.: Land Surface Temperature Data Assimilation and Its Impact on Evapotranspiration Estimates from the Common Land Model, Water Resour. Res., 45, 2008WR006971, https://doi.org/10.1029/2008WR006971, 2009.

Miyoshi, T., Yamane, S., and Enomoto, T.: Localizing the Error Covariance by Physical Distances within a Local Ensemble Transform Kalman Filter (LETKF), SOLA, 3, 89–92, https://doi.org/10.2151/sola.2007-023, 2007.

Miyoshi, T., Sato, Y., and Kadowaki, T.: Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System, Mon. Weather Rev., 138, 2846–2866, https://doi.org/10.1175/2010MWR3209.1, 2010.

Muñoz-Sabater, J., Lawrence, H., Albergel, C., Rosnay, P., Isaksen, L., Mecklenburg, S., Kerr, Y., and Drusch, M.: Assimilation of SMOS Brightness Temperatures in the ECMWF Integrated Forecasting System, Q. J. Roy. Meteor. Soc., 145, 2524–2548, https://doi.org/10.1002/qj.3577, 2019.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Naegeli, K., Neuhaus, C., Salberg, A.-B., Schwaizer, G., Weber, H., Wiesmann, A., Wunderle, S., and Nagler, T.: ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction – Snow on Ground (SCFG) from AVHRR (1982–2018), Version 2.0, NERC EDS Centre for Environmental Data Analysis [data set], https://doi.org/10.5285/3F034F4A08854EB59D58E1FA92D207B6, 2022.

Nagler, T., Schwaizer, G., Mölg, N., Keuris, L., Hetzenecker, M., and Metsämäki, S.: ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction – Snow on Ground (SCFG) from MODIS (2000–2020), Version 2.0, NERC EDS Centre for Environmental Data Analysis [data set], https://doi.org/10.5285/8847A05EEDA646A29DA58B42BDF2A87C, 2022.

NASA Global Land Data Assimilation System: GLDAS Noah Land Surface Model L4 3-hourly 0.25×0.25 degree V2.1, NASA Global Land Data Assimilation System [data set], https://doi.org/10.5067/E7TYRXPJKWOQ, 2025.

NASA Global Modeling and Assimilation Office (GMAO): MERRA-2 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, NASA Global Modeling and Assimilation Office (GMAO) [data set], https://doi.org/10.5067/RKPHT8KC1Y1T, 2025.

Novick, K., Biederman, J., Desai, A., Litvak, M., Moore, D., Scott, R., and Torn, M.: The AmeriFlux Network: A Coalition of the Willing, Agr. Forest Meteorol., 249, 444–456, https://doi.org/10.1016/j.agrformet.2017.10.009, 2018.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation, Tellus A, 56, 415, https://doi.org/10.3402/tellusa.v56i5.14462, 2004.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., Van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., Da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. D., Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. D., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, U., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V. R. D., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., Van Der Molen, M., Van Gorsel, E., Van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data, Sci. Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Pérez-Planells, L. and Martin, M.: LST_cci Product Validation and Intercomparison Report (PVIR), European Space Agency, https://admin.climate.esa.int/documents/2242/LST-CCI-D4.1-PVIR_-_i3r2_-_Product_Validation_and_Intercomparison_Report.pdf (last access: 25 September 2026), 2020.

Reichle, R. H.: Data Assimilation Methods in the Earth Sciences, Adv. Water Resour., 31, 1411–1418, https://doi.org/10.1016/j.advwatres.2008.01.001, 2008.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, Mon. Weather Rev., 130, 103–114, https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002.

Reichle, R. H., Kumar, S. V., Mahanama, S. P. P., Koster, R. D., and Liu, Q.: Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models, J. Hydrometeorol., 11, 1103–1122, https://doi.org/10.1175/2010JHM1262.1, 2010.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review, Earth-Sci. Rev., 99, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.

Seo, E., Lee, M.-I., and Reichle, R. H.: Assimilation of SMAP and ASCAT Soil Moisture Retrievals into the JULES Land Surface Model Using the Local Ensemble Transform Kalman Filter, Remote Sens. Environ., 253, 112222, https://doi.org/10.1016/j.rse.2020.112222, 2021.

Solberg, R., Reksten, J. H., Salberg, A.-B., Naegeli, K., Wunderle, S., Schwaizer, G., and Nagler, T.: ESA Snow Climate Change Initiative (Snow_cci): Daily Global Snow Cover Fraction – Snow on Ground (SCFG) from ATSR-2 (1995–2003), Version 1.0, NERC EDS Centre for Environmental Data Analysis [data set], https://doi.org/10.5285/0AEBA0C203C2447B9553A78F99D3A276, 2023.

Swinbank, R., Shutyaev, V., and Lahoz, W. A. (Eds.): Data Assimilation for the Earth System, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-010-0029-1, 2003.

The Common Land Model (CoLM): CoLM MPI Version 2010, Zenodo [code], https://doi.org/10.5281/zenodo.18649912, 2025.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble Square Root Filters, Mon. Weather Rev., 131, 1485–1490, https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2, 2003.

Vinnikov, K. Y., Robock, A., Speranskaya, N. A., and Schlosser, C. A.: Scales of Temporal and Spatial Variability of Midlatitude Soil Moisture, J. Geophys. Res.-Atmos., 101, 7163–7174, https://doi.org/10.1029/95JD02753, 1996.

Whitaker, J. S. and Hamill, T. M.: Ensemble Data Assimilation without Perturbed Observations, Mon. Weather Rev., 130, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002.

Yu, Z., Fu, X., Luo, L., Lü, H., Ju, Q., Liu, D., Kalin, D. A., Huang, D., Yang, C., and Zhao, L.: One-dimensional soil temperature simulation with Common Land Model by Assimilating in Situ Observations and MODIS LST with the Ensemble Particle Filter, Water Resour. Res., 50, 6950–6965, https://doi.org/10.1002/2012WR013473, 2014.

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 Resour. Res., 60, e2023WR036080, https://doi.org/10.1029/2023WR036080, 2024.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, S., Lu, F., Li, M., and Deng, X.: Coupled Data Assimilation and Parameter Estimation in Coupled Ocean–Atmosphere Models: A Review, Clim. Dynam., 54, 5127–5144, https://doi.org/10.1007/s00382-020-05275-6, 2020.