The 4DEnVar-based weakly coupled land data assimilation system for E3SM version 2

Pengfei Shi[1], L. Ruby Leung[1], Bin Wang[2], Kai Zhang[1], Samson M. Hagos[1], and Shixuan Zhang[1]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, USA

[2]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Correspondence: L. Ruby Leung (Ruby.Leung@pnnl.gov) and Pengfei Shi (pengfei.shi@pnnl.gov)

删除了: **weakly coupled land data assimilation**

设置了格式: 英语(英国)

Abstract. A new weakly coupled land data assimilation (WCLDA) system based on the four-dimensional ensemble variational (4DEnVar) method is developed and applied to the fully coupled Energy Exascale Earth System Model version 2 (E3SMv2). The dimension-reduced projection four-dimensional variational (DRP-4DVar) method is employed to implement 4DVar using the ensemble technique instead of the adjoint technique. With an interest in providing initial conditions for decadal climate predictions, monthly mean anomalies of soil moisture and temperature from the Global Land Data Assimilation System (GLDAS) reanalysis from 1980 to 2016 are assimilated into the land component of E3SMv2 within the coupled modeling framework with a one-month assimilation window. The coupled assimilation experiment is evaluated using multiple metrics, including the cost function, assimilation efficiency index, correlation, root mean square error (RMSE) and bias, and compared with a control simulation without land data assimilation. The WCLDA system yields improved simulation of soil moisture and temperature compared with the control simulation, with improvements found throughout the soil layers and in many regions of the global land. In terms of both soil moisture and temperature, the assimilation experiment outperforms the control simulation with reduced RMSE and higher temporal correlation in many regions, especially in South America, Central Africa, Australia, and large parts of Eurasia. Furthermore, significant improvements are also found in reproducing the time evolution of the 2012 U.S. Midwest drought, highlighting the crucial role of land surface in drought lifecycle. The WCLDA system is intended to be a foundational resource for research to investigate land-derived climate predictability.

删除了： initial

1 Introduction

The intrinsic chaos of the atmosphere limits traditional weather forecasting to roughly two weeks (Simmons and Hollingsworth, 2002). The feasibility of atmospheric predictability beyond two weeks lies with the interactions of the atmosphere with slowly varying components of the Earth system such as the ocean or land surface, or from predictable external forcings (Guo et al., 2012). Climate prediction can therefore be conceptually divided into both an initial value and a forced boundary value problem (Collins and Allen, 2002; Conil et al., 2007). One of the biggest technical challenges for improving the quality of climate predictions is the initialization of coupled models from observations (Taylor et al., 2012).

Much work has been devoted to initializing climate system models for practicable decadal climate predictions (DCPs). These models couple various components, such as models of the atmosphere, ocean, sea ice, land and river. Due to their complexity, coupled models are often more susceptible to initial conditions (ICs) than their individual model components, underscoring the importance of data assimilation (DA) (Sakaguchi et al., 2012). The application of DA methods is essential to incorporate reanalysis data into the components of coupled model and produce the optimal ICs to improve DCPs. The initialization for DCPs uses both uncoupled DA and coupled data assimilation (CDA) methods. Uncoupled DA performs DA under the framework of an individual component model (e.g., standalone land surface model forced by atmospheric observations or reanalysis data rather than coupled with an atmospheric model), and then the uncoupled DA analyses from different individual components are combined to form the ICs of a coupled model (Zhang et al., 2020). For example, most existing reanalysis data were produced using uncoupled DA approaches, and these reanalysis datasets are then directly used to initialize DCPs in some studies (Du et al., 2012; Bellucci et al., 2013). However, such uncoupled DA often exhibits poor consistency among the ICs of different component models, and eventually produces low prediction skills (Balmaseda et al., 2009; Boer et al., 2016; Ardilouze et al., 2017).

删除了： land surface, ocean, sea ice, and so on.

删除了： much higher

删除了： dedicated

删除了： capability

删除了： available observations

删除了： estimate of ICs to improve DCPs.

To obtain balanced multi-component ICs in coupled models, recent studies focus on the development of CDA methods under the coupled modeling framework (Penny and Hamill, 2017; He et al., 2020a). The purpose of CDA is to produce balanced and coherent ICs for all components within the climate system by incorporating reanalysis information from one or more components in the coupled model, providing great potential for improving seamless climate predictions (Dee et al., 2014). Some

删除了： observational

studies underscore the superior advantages of CDA over traditional uncoupled DA methods (Lea et al., 2015; Zhang et al., 2005). CDA methods are categorized into two main types: weakly coupled data assimilation (WCDA) and strongly coupled data assimilation (SCDA). WCDA assimilates the observations or existing reanalysis into the respective component of the coupled model and then transfers reanalysis information to the other components through the coupled model integration (He et al., 2020b; Zhang et al., 2020). Considering that sequential DA encompasses both the analysis and the forecast steps, WCDA allows no direct influence of reanalysis information from a single component to other components in the analysis step as the cross-component background error covariances are not used, but coupling in the forecast step allows interactions across different components during the model integration (Browne et al., 2019) and propagates reanalysis information to other components. In contrast, SCDA utilizes cross-component background error covariances to directly assimilate reanalysis information from one component into all components, treating the entire Earth system model as one unified system (Penny et al., 2019). Furthermore, similar to WCDA, SCDA also allows coupling in the forecast step to propagate reanalysis information from one component to the other components (Yoshida and Kalnay, 2018). Several studies indicate that SCDA typically exhibits more pronounced improvements in assimilation performance relative to WCDA (Smith et al., 2015; Sluka et al., 2016). However, the application of SCDA poses substantial technical challenges, particularly in the establishment of effective cross-component background error covariances. Consequently, the majority of contemporary CDA systems still utilize the WCDA framework.

Recent research efforts have started to implement the CDA system to initialize DCPs, using a diverse range of DA techniques from simple to complex. The simplest method is nudging which adjusts the model states towards the observations or existing reanalysis (Hoke and Anthes, 1976; Zhang et al., 2020). Although the nudging method is time-saving and easy to implement, its application in CDA is restricted primarily due to the limited types of observations and the required interpolation of observations at every time step of model integration (He et al., 2017). Previous studies have developed advanced CDA systems using variational and filtering approaches, such as the three-dimensional variational data assimilation (3DVar) (Fujii et al., 2009; Yao et al., 2021), and ensemble-based techniques like the ensemble Kalman filter (EnKF) (Zhang et al., 2007). The former generally utilizes the stationary

删除了： the

删除了： observational

删除了： observations

删除了： the observational

删除了： the observational

删除了：

删除了： the observations

删除了： Laloyaux et al., 2016

background error covariance and assimilates observations sequentially (Lin et al., 2017). In contrast, the latter uses the flow-dependent forecast error covariance and recursively integrates observations into the model (Lei and Hacker, 2015). Several studies also show encouraging progress in constructing CDA systems using four-dimensional variational data assimilation (4DVar) method (Smith et al., 2015; Fowler and Lawless, 2016). The objective of 4DVar is to optimize four-dimensional model states and provide a compatible temporal trajectory that matches observational records across each assimilation window (Mochizuki et al., 2016). The 4DVar method is an advanced assimilation technique that exhibits superiority over other assimilation techniques like nudging and 3DVar in multiple aspects. Initial shocks that influence prediction skills can be significantly minimized by the 4DVar approach due to the dynamical consistency between the model and ICs (Sugiura et al., 2008). However, it is difficult to apply the 4DVar method for CDA systems in the fully coupled model because of the challenge in adjoint integration of the coupled model and its high computational cost in the analysis step. Finally, to capitalize on the strengths of both ensemble and variational techniques, recent studies focus on developing new hybrid data assimilation methods (Wang et al., 2010; Buehner et al., 2018). The hybrid approach utilizes an ensemble forecast to generate flow-dependent forecast error covariances and presents a way to perform 4DVar optimization without the need for tangent linear and adjoint models (Lorenc et al., 2015). However, most studies on CDA have focused on assimilating observations or reanalysis data of ocean, atmosphere and even sea ice (He et al., 2017; Li et al., 2021; Kimmritz et al., 2018). There have been relatively few instances of CDA studies assimilating land observations or land reanalysis data.

In this study, we introduce the development of the 4DEnVar-based weakly coupled land data assimilation (WCLDA) system for the Energy Exascale Earth System Model version 2 (E3SMv2) (Golaz et al., 2022). The 4DEnVar method in this WCLDA system is the dimension-reduced projection 4DVar (DRP-4DVar; Wang et al., 2010) which utilizes the ensemble technique as an alternative to the adjoint technique for implementing 4DVar. In this WCLDA system, monthly mean anomalies of soil moisture and temperature from a global land reanalysis product are assimilated into the land component of a coupled climate model in the analysis step, and subsequently during the forecast step, the land reanalysis information incorporated into the ICs of the land component is propagated to the other components (e.g., atmosphere and ocean) through the fully coupled model integration and influences the ICs of all components for the next assimilation window. The primary goal of the WCLDA system is intended to be a foundational resource for exploring predictability of the Earth system by the E3SM community, specifically focusing on understanding the sources of predictability provided by land versus ocean, with an initial focus on DCPs. This WCLDA system also provides the groundwork for future actionable predictions of Earth system variability using E3SM.

The objective of this paper is to introduce the implementation of the 4DEnVar-based WCLDA system for the land component of E3SMv2. In Section 2, we provide an overview of the E3SMv2 model, describe the 4DEnVar methodology in detail and outline the framework of the 4DEnVar-based WCLDA system. Preliminary evaluation of the WCLDA system is presented in Section 3. Finally, conclusions are discussed in Section 4.

删除了： major

2 Methods

2.1 Model Description

The model used in this study is a relatively new state-of-the-art Earth system model known as Energy Exascale Earth System Model version 2 (E3SMv2), supported by the U.S. Department of Energy (DOE) to improve actionable Earth system predictions and projections (Leung et al., 2020). The atmospheric component is the E3SM Atmosphere Model version 2 (EAMv2), which is built on the spectral-element atmospheric dynamical core with 72 vertical levels (Dennis et al., 2012; Taylor et al., 2020). At the standard resolution, EAMv2 is applied on a cubed sphere with a grid spacing of ~100 km for the dynamics. The ocean component is the Model for Prediction Across Scales-Ocean (MPAS-O), which applies the underlying spatial discretization to the primitive equations with 60 layers using a z-star vertical coordinate (Petersen et al., 2018; Reckinger et al., 2015). The sea ice component is MPAS-SI, which shares the same Voronoi mesh with MPAS-O, with mesh spacing varying between 60km in the mid-latitudes and 30 km at the equator and poles (Golaz et al., 2022). The land component is the E3SM Land Model version 2 (ELMv2), which is based on the Community Land Model version 4.5 (CLM4.5) (Oleson et al. 2013). Simulations are run in a satellite phenology mode with prescribed leaf area index, and the prescribed vegetation distribution has been updated for better consistency between land use and changes in plant functional types described by Golaz et al. (2022). The river transport component is the

Model for Scale Adaptive River Transport version 2 (MOSARTv2), which provides detailed representation of riverine hydrologic variables (Li et al., 2013). These five components exchange fluxes through the top-level coupling driver version 7 (CPL7) (Craig et al., 2012). Further details on the E3SMv2 model are described in Golaz et al. (2022).

2.2 Datasets

Monthly mean soil moisture and soil temperature data in a total of ten soil layers are produced by the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). The GLDAS product generates optimal fields of land surface states and fluxes in near-real time by forcing multiple offline land surface models with observation-based data fields. These reliable and high-resolution global land surface datasets from GLDAS are extensively utilized in weather and climate studies, hydrometeorological investigations and water cycle research (Chen et al., 2021; Zhang et al., 2018). The GLDAS datasets have been available globally at high spatial resolution since January 1979 and can be accessed through the Goddard Earth Science Data and Information Service Center. For more consistency with ELMv2, we utilize GLDAS data produced by CLM. In contrast to decadal timescales, data signals with temporal resolutions shorter than one month can potentially introduce undesirable noise, which can adversely affect DCPs when high temporal resolution data are assimilated into the ICs. Moreover, it is very computationally demanding to assimilate complex actual observations in the initialization for DCPs that requires long-term DA cycles. Therefore, similar to most existing initialization approaches for DCPs that assimilate reanalysis data, this study describes the implementation of a data assimilation approach for initializing DCPs by assimilating monthly mean GLDAS data within the one-month assimilation window.

移动了(插入) [1]

Monthly mean surface soil moisture data from the Advanced Microwave Scanning Radiometer (AMSR) and land surface temperature data from the Moderate Resolution Imaging Spectrometer (MODIS) are utilized for validation. (1) The AMSR data provides surface soil moisture estimations by measuring the microwave emission from the Earth's surface (Njoku et al., 2003). The soil moisture data from AMSR are widely used in scientific research to study surface water cycles, drought conditions and hydrologic modeling (Du et al., 2019; McCabe et al., 2008). (2) MODIS is an essential instrument onboard the Terra and Aqua satellite platforms (Remer et al., 2005). The MODIS datasets provide comprehensive global observations describing atmospheric, terrestrial and oceanic conditions, including land surface temperature, vegetation indices and cloud properties (Justice et al., 2002). The MODIS products are extensively utilized for monitoring environmental changes and supporting climate change research (Gao et al., 2015; Mertes et al., 2015).

Current initialization techniques are broadly classified into two categories: full-field assimilation with reanalysis values, and anomaly assimilation with reanalysis anomalies (Hu et al., 2020; Polkova et al., 2019). The full-field assimilation is commonly performed to reduce the influence of systematic model biases by replacing the initial model state with the optimal available estimate of the reanalysis state (Volpi et al., 2017). However, the model trajectory tends to drift away from the observations and revert to the model's inherent preferred state because of model deficiencies (Smith et al., 2013). This problem is partially addressed with the anomaly assimilation by assimilating the reanalysis anomalies added to the model climatology (Carrassi et al., 2014). In this study, we conduct the anomaly assimilation for the WCLDA system with bias correction applied to GLDAS data before assimilation. For bias correction, the difference between GLDAS data and its long-term average is calculated as anomalies and then added to the simulated model climatology.

删除了： observed values

删除了： observed anomalies

删除了： observed state

删除了： observed anomalies

2.3 Data Assimilation Scheme

The 4DEnVar algorithm in this study is based on the DRP-4DVar technique, which is an efficient pathway for applying 4DVar through using the ensemble method rather than the adjoint technique (Wang et al., 2010). The DRP-4DVar method generates the optimal estimation in the sample space through aligning the observations with ensemble samples along the coupled model trajectory (Liu et al., 2011).

DRP-4DVar is an economical approach that minimizes the cost function of the standard 4DVar by using the ensemble technique instead of the adjoint technique (Wang et al., 2010). The background error covariance matrix B is estimated using the pure ensemble covariance. The ensemble members originate from historical or ensemble forecasts. Considering the high computational cost of ensemble forecasts for the coupled model in our study, we utilize outputs from the pre-industrial control (PI-CTRL) experiment of E3SMv2 to generate ensemble members. The instantaneous state at the beginning of each month and the corresponding monthly mean state of this month from the 100-year balanced PI-CTRL simulation are used as the samples of initial condition (x_i) and forecast samples (y_i). The corresponding perturbation samples are calculated as $x_i' = x_i - x$ and $y_i' = y_i - y$, where x and y are the 100-year average values of x_i and y_i at the same month, respectively. Then, m pairs of perturbation samples $(x_1', x_2', x_3', \cdots, x_m')$ and $(y_1', y_2', y_3', \cdots, y_m')$ are selected at each DA analysis step according to the correlations between y_i' and the observational innovation $y_{obs}' = y_{obs} - y_b$, ensuring that each y' sample is selected independently of the other samples in the ensemble. In this study, $m = 30$. Then the estimation of the background error covariance matrix B is represented by the formula in Eq. (1), utilizing the selected x' samples. We implement both horizontal and vertical localization to reduce sampling errors due to the finite ensemble size and to alleviate the spurious remote influence from distant grid points. Our approach to horizontal localization is to apply a distance-dependent weighting function to the background error covariance. The vertical localization is employed to limit the influence of reanalysis information on specific soil layers. Please refer to Wang et al. (2018) for more detailed descriptions of the localization methodology in our study.

删除了： and the independence between y' samples.

设置了格式： 字体颜色：文字 1

删除了： To remove the spurious remote correlations in the B matrix, the localization approach is applied to optimize the assimilation performance (Wang et al., 2018).

$$\begin{cases} B = bb^T \\ b = \dfrac{1}{\sqrt{m-1}} \times (x_1' - x', x_2' - x', x_3' - x', \cdots, x_m' - x') \\ x' = \dfrac{1}{m}(x_1' + x_2' + x_3' + \cdots + x_m') \end{cases} \tag{1}$$

According to Wang et al. (2010), DRP-4DVar produces the analysis increment (x_a') by minimizing the 4DVar cost function in the incremental form (Courtier et al., 1994):

$$\begin{cases} J(x_a') = \min\limits_{x'} J(x') \\ J(x') = \frac{1}{2}(x')^T B^{-1} x' + \frac{1}{2}(\mathcal{y}' - \mathcal{y}_{obs}')^T (\mathcal{y}' - \mathcal{y}_{obs}') \end{cases} \tag{2}$$

Here $x' = x - x_b$ represents the increment of model variables relative to the background; $\mathcal{y}_{obs}' = r^{-1} y_{obs}' = r^{-1}(y_{obs} - y_b)$ denotes the weighted observational innovation for monthly mean anomalies of soil moisture and temperature, and $R = rr^T$ is the observational error covariance matrix that is usually diagonal; $\mathcal{y}' = r^{-1} y' = r^{-1}(y - y_b)$ is the weighted projection of the increment (x') onto the observation space; the superscript T represents the transpose.

To simplify the calculation of the minimization, the increment of model state variables x' and the corresponding weighted observation increment $\mathcal{y}'$ are projected onto the dimension-reduced sample space through the following projection transformations:

$$\begin{cases} x' = P_x \alpha \\ y' = P_y \alpha \end{cases} \tag{3}$$

where α is the m-dimension column vector containing the weight coefficients $(\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_m)$; P_x and P_y denote the projection matrices that incorporate the initial condition perturbations and the corresponding monthly mean samples as follows:

$$\begin{cases} P_x = (x'_1, x'_2, x'_3, \cdots, x'_m) \\ P_y = (y'_1, y'_2, y'_3, \cdots, y'_m) \end{cases} \tag{4}$$

where $y'_i = r^{-1} y'_i \; (i = 1, 2, \cdots, m)$. Then the original 4DVar cost function defined in Eq. (2) is transformed into the following new cost function and the analysis can be computed in the sample space by minimizing this new cost function:

$$\begin{cases} J(\alpha_a) = \min_{\alpha} J(\alpha) \\ J(\alpha) = \frac{1}{2}\alpha^T B_\alpha^{-1} \alpha + \frac{1}{2}(P_y \alpha - y'_{obs})^T (P_y \alpha - y'_{obs}) \\ x_a = x_b + x'_a = x_b + P_x \alpha_a \end{cases} \tag{5}$$

The solution to this minimization problem is formulated as:

$$\alpha_a = (B_\alpha^{-1} + P_y^T P_y)^{-1} P_y^T y'_{obs} \tag{6}$$

In this study, the DRP-4DVar-based WCLDA system is used to incorporate the land reanalysis data only. The optimal analysis for the land state variables (x_a^{lnd}) is obtained by adding the analysis increment (${x'}_a^{lnd}$) to the background of land ICs (x_b^{lnd}), as expressed in Eq. (7):

$$x_a^{lnd} = x_b^{lnd} + {x'}_a^{lnd} = x_b^{lnd} + P_x \left(B_\alpha^{-1} + P_y^T P_y\right)^{-1} P_y^T y'_{obs} \tag{7}$$

In the analysis step, only the land state variables are updated to the optimal analysis (x_a^{lnd}). Subsequently, we proceed with a one-month freely coupled integration of the E3SMv2 model during the forecast step. This integration is initialized from the optimal land ICs (x_a^{lnd}) along with the background fields as the ICs of other components (e.g., atmosphere and ocean). Throughout this one-month free integration, the interactions among the model components indirectly enhance the background states of these components (e.g., atmosphere and ocean) for the next assimilation window due to the more realistic land state variables. Moreover, this coupled integration also contributes to the balance between the ICs of different components.

2.4 4DEnVar-based WCLDA System

The 4DEnVar-based WCLDA system is developed to assimilate the monthly mean soil moisture and temperature data from the GLDAS analysis dataset into the land component of E3SMv2 using the DRP-4DVar method. Two sets of numerical experiments are conducted to evaluate the performance of land data assimilation in the WCLDA system. The control simulation (CTRL) is a 37-year freely coupled integration driven by observed external forcings (e.g., solar radiation, greenhouse gas and aerosol concentrations) from 1980 to 2016. In the freely coupled simulation, the various components of the Earth system model, namely the atmosphere, land, river, ocean, and sea ice, interact dynamically without any constraints. The observed external forcing mainly acts on the atmospheric component and then influences other components (e.g., land surface, ocean, and sea ice) through their coupling with the atmosphere. CTRL provides the benchmark for assessing the performance of the WCLDA system. The assimilation experiment (Assim) is conducted from 1980 to 2016 based on the WCLDA system in which the GLDAS data are assimilated into the land state variables from the first to the tenth layer with a one-month assimilation window under the coupled modeling framework. The effectiveness of the WCLDA system is evaluated through the comparison between Assim and CTRL. In both Assim and CTRL, the transient-historical external forcings are prescribed following the CMIP6 protocol (Eyring et al., 2016).

删除了:

删除了： restraints

上移了［1]： In contrast to decadal timescales, data signals with temporal resolutions shorter than one month can potentially introduce undesirable noise, which can adversely affect DCPs when high temporal resolution data are assimilated into the ICs. Moreover, it is very computationally demanding to assimilate complex actual observations in the initialization for DCPs that requires long-term DA cycles. Therefore, similar to most existing initialization approaches for DCPs that assimilate reanalysis data, this study describes the implementation of a data assimilation approach for initializing DCPs by assimilating monthly mean GLDAS data within the one-month assimilation window.

The flowchart of the 4DEnVar-based WCLDA system is illustrated in Figure 1. The DRP-4DVar method incorporates three inputs: model background, observational innovation and 30 perturbation samples. First, the E3SMv2 model is executed for one month, during which state variables such as model background (x_b), observational operator (H) and observational background (y_b) are stored. The model background (x_b) denotes the monthly initial states before assimilation, and the observational operator (H) represents a one-month integration by the coupled model to generate monthly mean model outputs (y_b). Second, upon completion of the one-month coupled run, the observational innovation (y'_{obs}) is determined by calculating the differences in soil moisture and temperature between the monthly mean GLDAS data (y_{obs}) and the model outputs (y_b). From the 100-year sample database of the E3SMv2 PI-CTRL simulation, 30 samples of monthly mean perturbation (y') are chosen with the highest absolute correlation with the observational innovation. The corresponding 30 monthly IC samples (x') are also obtained. Finally, the analysis increment is generated in the sample space and the optimal analysis (x_a) is calculated using the DRP-4DVar algorithm.

删除了： To alleviate spurious correlations, a localization scheme is implemented in the DRP-4DVar-based WCLDA system (Wang et al., 2018).

The schematic diagram in Figure 2 outlines the assimilation process of the 4DEnVar-based WCLDA system in E3SMv2. The incorporation of GLDAS data into the E3SMv2 model consists of the analysis step and the forecast step. In the analysis step, the differences between monthly mean GLDAS data and model outputs are calculated and utilized to produce the optimal assimilation analysis at the beginning of a one-month assimilation window. Subsequently, in the forecast step, this optimal assimilation analysis is used as the land ICs combined with the background ICs for other components to conduct one-month forecast using the E3SMv2 model. This forecast generates the backgrounds of all model components for the next assimilation window. As a result, the forecasted backgrounds for all components are influenced by the land reanalysis information incorporated into the ICs of the land component. In general, when the coupled model is used in the forecast step while the optimal assimilation analysis is updated separately for the respective component, the assimilation approach is identified as WCDA (Penny et al., 2019; Zhang et al., 2020).

删除了： to

The detailed assimilation process mainly consists of three steps within each one-month assimilation window: 1) the E3SMv2 model is initially executed for one month to generate the simulated monthly mean soil moisture and temperature (y_b^{lnd}); 2) the observational innovation (y'_{obs}) is obtained through subtracting the model simulation (y_b^{lnd}) from the monthly mean observation (y_{obs}^{lnd}). This innovation is then applied to formulate the optimal assimilation analysis of land surface (x_a^{lnd}) at the beginning of the assimilation window through the DRP-4DVar method; 3) the E3SMv2 model is rewound to the start of the month and the second one-month model run is executed using the optimal ICs (x_a) to generate the background for the next assimilation window. Due to multi-component interactions during the one-month freely coupled integration, the land reanalysis information can potentially benefit other components (e.g., atmosphere and ocean) in the coupled modeling framework (Li et al., 2021; Shi et al., 2022). To assimilate the monthly mean GLDAS product, fully coupled integration by the E3SMv2 model is performed twice within each one-month assimilation window: first to generate the observational innovation by computing the differences between the GLDAS data and model outputs for analysis, and second to forecast the backgrounds of all components for the next assimilation window. When the fully coupled model is executed for the second one-month run, the land reanalysis information is transferred to the other components through multi-component interactions. This approach is similar to previous studies that employed the "two-step" scheme in which the land model integration is performed twice within the same month to assimilate the monthly GRACE-based TWS observations (Houborg et al., 2012; Girotto et al., 2016).

2.5 Evaluation Metrics

The reduction rate of the cost function is a significant metric for verifying the effectiveness of the WCLDA system and evaluating the extent of reanalysis information assimilated by the coupled model, which is formulated as:

删除了： observational

$$\begin{cases} \frac{J_1 - J_0}{J_0} \times 100\% \\ J_0 = \frac{1}{2}(y_{obs} - y_b)^T R^{-1}(y_{obs} - y_b) \\ J_1 = \frac{1}{2}(y_{obs} - y_a)^T R^{-1}(y_{obs} - y_a) \end{cases} \tag{8}$$

where J_0 and J_1 denote the cost function before and after assimilation respectively, y_{obs} represents the GLDAS data, y_a denotes the monthly mean analyses, y_b is the observation-space background, and R is defined as the observation error covariance matrix. The observation error covariance matrix R can be determined statistically by estimating the variance and covariance of the GLDAS data. Negative value for this metric indicates that reanalysis information has been correctly incorporated into the model variables.

删除了： observational

删除了： observational

设置了格式： 字体：(默认) Times New Roman

Following Yin et al. (2014), the assimilation efficiency (AE) index is defined to evaluate the efficiency of the WCLDA system as follows:

$$AE = \frac{RMSE_{Assim}}{RMSE_{CTRL}} - 1 \tag{9}$$

In this equation, $RMSE_{Assim}$ is the root mean square error (RMSE) between the analysis value from Assim and the reference data, while $RMSE_{CTRL}$ represents the RMSE between CTRL and the reference data. Negative (positive) AE value indicates improvements (degradations) by the assimilation. In the following sections, we use the GLDAS data as the main reference data to verify the correctness of the WCLDA system, but some analyses are also performed using AMSR surface soil moisture and MODIS land surface temperature as the reference data.

删除了： Assim

3 Results

3.1 Evaluation of the cost function

Figure 3 displays the time series of the monthly reduction rate of the cost function in the 4DEnVar-based WCLDA system. In the first month, the reduction rate reaches approximately 26.06% in Assim. Over the subsequent months, Assim maintains the average reduction rate of 7.73% throughout the entire 37-year period. Furthermore, negative reduction rates are observed in 98.65% of the total months, indicating the effectiveness of the WCLDA system. These results suggest that the WCLDA system is correctly implemented, with GLDAS data successfully assimilated into the coupled model.

删除了： the observational

3.2 Evaluation of the AE index

The spatial pattern of the AE index for soil moisture at different depths is depicted in Figure 4. The AE value exhibits negative signal in most areas for total ten soil layers, suggesting the reduction in RMSE for soil moisture after assimilation. Significant improvements appear over North America, South America, southern Africa, Europe, and Asia. However, assimilation performance is degraded in the northern part of Russia and northern Africa. This is consistent with the findings in other studies that assimilation updates in northern Russia are limited due to the complexities of accurately representing frozen ground and snow processes in high latitudes (Edwards et al., 2007; Ireson et al., 2013). As surface soil moisture is highly susceptible to atmospheric conditions, assimilation performance of surface soil moisture is limited by the accuracy of atmospheric forcing. Furthermore, some degradations found in the deep layers could be attributed to the substantial influence of various terrestrial factors, such as subsurface runoff and interactions with groundwater, similar to the findings in previous studies (Liu and Mishra, 2017; Zeng and Decker, 2009).

Figure 5 shows the spatial distribution of the AE index for soil temperature from surface to deep layers. Most grid cells from the ten soil layers are dominated by negative AE signals, indicating improved performance for soil temperature after assimilation. Moreover, the spatial patterns across different soil layers are highly consistent with each other and exhibit similar magnitudes in most areas. Notable improvements are observed in central Europe, South America, eastern Russia, and large parts of Eurasia and North America. In contrast, slight degradations appear over Southeast Asia and along the northern fringes of Africa. This may be partly related to model uncertainties and possible atmospheric noise, as shown by many past studies (Kwon et al., 2016; Lin et al., 2020).

We further perform an analysis of the spatial pattern of the AE index for surface soil moisture and land surface temperature between satellite data and model simulations (Figure A1). For surface soil moisture, the comparison with AMSR data suggests that the majority of global regions exhibit reduced RMSE after assimilation. The reduction of RMSE is pronounced in central North America, South America, southern Africa, Australia, and Europe. However, in high-latitude areas, significant degradations are observed in northern Russia, which may be possibly related to model deficiencies in simulating the complex frozen ground and snow processes (Edwards et al., 2007; Ireson et al., 2013). Regarding land surface temperature, improved performances are evident over South America, Australia, southern Africa, and large parts of Eurasia when compared to MODIS data. In contrast, some degradations appear over parts of North America and central Asia, which still require further improvement.

3.3 Evaluation of the correlation

Figure 6 displays the spatial patterns of the differences in temporal correlations for soil moisture between Assim and CTRL with GLDAS data across different soil layers. The majority of global regions in Assim exhibit higher correlations from the first to the tenth layer compared with CTRL, suggesting the overall good performance of the WCLDA system. Enhanced correlations in deep soil layers are more pronounced than in shallow layers, which may be attributed to the longer memory of soil processes in the deeper soil layers (Wang et al., 2010). Improved correlations appear over North America, central Europe, Asia, and parts of Africa. However, some scattered areas show slight degradations, such as northern South America, central Africa, and eastern Russia. Overall, Assim outperforms CTRL with higher correlation (Figure 6) and lower RMSE (Figure 4) in many regions, such as Europe, North America, southern South America, and South Asia.

删除了： observations

The correlation differences in soil temperature between Assim and CTRL from surface to deep layers are shown in Figure 7. Assim yields improved correlations from the first to the tenth layer across the majority of global regions. Furthermore, similar spatial patterns and magnitudes are observed in the performance of different soil layers, implying the significant heat transfer from the surface to deep zone that constrains soil temperature across the soil column. Notable improvements are located over South America, central Africa, Australia, central Europe, and East Asia. Nevertheless, some degradations appear over North America, western Europe, and Northeast China. Assim shows superior performance over CTRL for soil temperature with higher correlation (Figure 7) and lower RMSE (Figure 5) in many regions, including South America, central Europe, Australia, and central Africa.

3.4 RMSE and bias of the global mean soil moisture and temperature

The vertical distributions of RMSE differences between Assim and CTRL for soil moisture and temperature are evaluated in Figure 8. Compared with CTRL, Assim shows noticeable improvements with reduced RMSE for both soil moisture and temperature in all ten soil layers. For soil moisture, the reduction of RMSE increases with depth from the upper to deep soil layers, reaching its maximum at the tenth layer. This could be attributed to the longer soil memory in deep layers than shallow layers. For soil temperature, the reduction of RMSE exhibits similar magnitude from the surface to deep soil layers, which may be explained by the significant heat transfer across different soil layers in regulating soil temperature throughout the soil column.

Figure 9 shows the time evolutions of the vertically averaged global mean soil moisture and temperature bias and RMSE differences. For soil moisture bias (Figure 9a), CTRL exhibits dry biases during the first twenty years and wet biases afterwards. In contrast, Assim shows smaller biases during both periods by reducing the dry bias prior to ~2000 and the wet bias thereafter. Assim also exhibits reduced RMSE (Figure 9b) for soil moisture throughout the entire 37-year period. For soil temperature bias (Figure 9c), CTRL and Assim display comparable performances, possibly due to the small magnitude of model deviation in soil temperature. The RMSE differences (Figure 9d) suggest that Assim decreases the RMSE for soil temperature in the majority of months, with 74.10% of the total months in Assim exhibiting lower RMSE than CTRL. In summary, the superior performance for both soil moisture and temperature in Assim demonstrates that land reanalysis information has been effectively incorporated into the model variables through the WCLDA system.

Noticeably, the simulated soil temperature and soil moisture display similar long-term trends, with cold and dry biases before ~2000 and warm and wet biases afterwards. The soil temperature biases may be related to the global surface air temperature simulated in E3SMv2, which is notably too cold compared to the observed record during the 1970s and 1980s while the model warms up quickly after ~year 2000 (see Figure 23 of Golaz et al., 2022). The global surface air temperature biases during the past decades in E3SMv1 and v2 have been attributed to the strong aerosol forcing in the model (Golaz et al., 2019; 2022). As the global mean precipitation scales with the surface temperature at ~2% per degree (Allen and Ingram, 2002), model biases in surface temperature are reflected in biases in precipitation and hence soil moisture, resulting in similar long-term trends between soil temperature and soil moisture biases in the simulations.

3.5 2012 U.S. Midwest Drought

To further evaluate the performance of the WCLDA system, we briefly investigate the impact of land data assimilation on simulating the temporal evolution of the U.S. Midwest drought in 2012. Time series of soil moisture percentiles over the Midwest (36°-50°N, 102°-88°W) demonstrate significant improvements by Assim in reproducing the time evolution of agricultural drought in 2012 compared with CTRL (Figure 10). From ERA-Interim data, the agricultural drought starts in August 2011, follows by a brief relief in early spring of 2012, peaks in September 2012, and recovers by January 2013. The drought develops rapidly between May and July 2012 over a wide-spread area including the central and midwestern U.S. This flash drought caused significant agricultural damages and economic losses.

删除了： the observation based on

The free running CTRL experiment fails to simulate the temporal evolution of the 2012 Midwest drought, with a correlation coefficient between CTRL and ERA-Interim of only 0.27. The onset and peak of the drought are remarkably well captured by Assim, although the drought recovery occurs two months later than observed. The correlation coefficient of the Assim time series with ERA-Interim is 0.56, which is statistically significant at the 95% confidence level. Our results highlight the importance of land surface states for drought lifecycle, with the potential to improve future drought predictions through the implementation of the WCLDA system.

删除了： observation

删除了： observation

4 Conclusions

In this study, we developed the 4DEnVar-based WCLDA system for the E3SMv2 model and evaluated the performance of this WCLDA system. The DRP-4DVar method was employed for implementing 4DVar using the ensemble method rather than the adjoint technique. Special attention is paid to directly assimilating monthly mean land reanalysis data in this system without interpolating to every time step. Within each one-month assimilation window, we assimilate land reanalysis information into the coupled model without breaking the land-atmosphere interaction, which is important for the WCLDA system to be used to understand the potential sources of predictability provided by land.

Monthly mean anomalies of soil moisture and temperature from the GLDAS reanalysis are assimilated from 1980 to 2016 through the WCLDA system, and its performance is evaluated using multiple metrics, including the cost function, AE index, correlation, RMSE and bias. Compared with CTRL, the cost function is reduced by Assim in most months, suggesting that the GLDAS reanalysis data has been effectively incorporated into the model. In terms of both soil moisture and temperature, Assim outperforms CTRL with lower RMSE and higher temporal correlation in many regions, especially in South America, central Africa, Australia, and large parts of Eurasia. For soil moisture bias, Assim further decreases the dry bias during the first twenty years and the wet bias thereafter. It is noteworthy that the subseasonal-to-seasonal time evolution of soil moisture percentiles during the 2012 U.S. Midwest drought can be quite well captured in Assim, underscoring the significant role of land surface states in drought propagation.

删除了： The WCLDA system is conducted from 1980 to 2016

Our current WCLDA system has some limitations and requires future improvements. Future enhancements of our WCLDA system will explore the assimilation of additional land products, particularly those derived from satellite observations. The incorporation of such satellite-based datasets is expected to further improve the performance of the WCLDA system. It is possible that the influence of the WCLDA system on atmospheric processes may be limited in some domains due to uncertainties of the model parameterizations, particularly in representing land-atmosphere interactions (Zhou et al., 2023). For example, in humid regions where the evaporation process is predominantly energy-limited, the assimilation of soil moisture tends to exert limited influence. Instead, the assimilation of soil temperature may yield more substantial improvements. This underscores the importance of the unique characteristics and constraints presented by complicated regional conditions in the application of assimilation processes. To this end, the application of the WCLDA system would motivate future work to better understand the roles of the land surface in climate variability and provide a foundational resource for future predictability studies by the E3SM community.

删除了： such as the lack of an observation operator to integrate actual observations (e.g., satellite and station data). An observation operator is crucial in providing the linkage between the model variables and actual observations, which differ in spatial and temporal resolutions. Hence future exploration will focus on developing observation operators suitable for assimilating various satellite data, such as the AMSR-E and GRACE data.

设置了格式： 字体：(中文) Times New Roman, 字体颜色：自动设置

Code and data availability. The E3SMv2 source codes used in this study can be accessed on Zenodo at https://zenodo.org/record/8194050. The GLDAS monthly soil moisture and soil temperature data are available online (https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS%20monthly&page=1). The MODIS monthly land surface temperature data can be downloaded from the website (https://disc.gsfc.nasa.gov/datasets/MOD11CM1D_005/summary). The AMSR monthly surface soil moisture data are available from https://doi.org/10.11888/Soil.tpdc.270960. The ERA-Interim monthly soil moisture data are available at https://apps.ecmwf.int/archive-catalogue/?levtype=sfc&type=an&class=ei&stream=moda&expver=1. The model data used in this study can be found on Zenodo at https://zenodo.org/record/8148737.

Author contributions. LRL initiated this study. PS and LRL designed the experiments. PS developed the data assimilation code and performed the simulations. BW provided advice on the data assimilation technique and KZ and SZ provided assistance with the E3SM model. PS and LRL analyzed and interpreted the data. PS and LRL wrote the paper. BW, KZ, SMH, and SZ contributed to the revision.

Competing interests. The authors declare no competing interests.

Acknowledgements. This research was supported by the Office of Science, Department of Energy Biological and Environmental Research as part of the Regional and Global Model Analysis program area. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RL01830.

References

Allen, M. R., and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, Nature, 419, 224–232, https://doi.org/10.1038/nature01092, 2002.

Ardilouze, C., Batté, L., Bunzel, F., Decremer, D., Déqué, M., Doblas-Reyes, F. J., Douville, H., Fereday, D., Guemas, V., MacLachlan, C. and Müller, W.: Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability, Climate Dynamics, 49, 3959–3974, https://doi.org/10.1007/s00382-017-3555-7, 2017.

Balmaseda, M. A., Alves, O. J., Arribas, A., Awaji, T., Behringer, D. W., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T. and Stammer, D.: Ocean initialization for seasonal forecasts, Oceanography, 22(3), 154–159, https://doi.org/10.5670/oceanog.2009.73, 2009.

Bellucci, A., Gualdi, S., Masina, S., Storto, A., Scoccimarro, E., Cagnazzo, C., Fogli, P., Manzini, E., and Navarra, A.: Decadal climate predictions with a coupled OAGCM initialized with oceanic reanalyses, Climate Dynamics, 40, 1483–1497, https://doi.org/10.1007/s00382-012-1468-z, 2013.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R. and Mueller, W. A.: The decadal climate prediction project (DCPP) contribution to CMIP6, Geoscientific Model Development, 9(10), 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016.

Browne, P. A., De Rosnay, P., Zuo, H., Bennett, A., and Dawson, A.: Weakly coupled ocean-atmosphere data assimilation in the ECMWF NWP system, Remote Sensing, 11(3), 234, https://doi.org/10.3390/rs11030234, 2019.

Buehner, M., Du, P., and Bédard, J.: A new approach for estimating the observation impact in ensemble–variational data assimilation, Monthly Weather Review, 146(2), 447-465, https://doi.org/10.1175/MWR-D-17-0252.1, 2018.

Carrassi, A., Weber, R. J. T., Guemas, V., Doblas-Reyes, F. J., Asif, M., and Volpi, D.: Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations, Nonlinear Processes in Geophysics, 21, 521-537, https://doi.org/10.5194/npg-21-521-2014, 2014.

Chen, Z., Zeng, Y., Shen, G., Xiao, C., Xu, L., and Chen, N. C.: Spatiotemporal characteristics and estimates of extreme precipitation in the Yangtze River Basin using GLDAS data, International Journal of Climatology, 41, 1812–1830, https://doi.org/10.1002/joc.6813, 2021.

Collins, M. and Allen, M. R.: Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability, Journal of Climate, 15, 3104–3109, https://doi.org/10.1175/1520-0442(2002)015<3104:ATRROI> 2.0.CO;2, 2002.

Conil, S., Douville, H., and Tyteca, S.: The relative influence of soil moisture and SST in climate predictability explored within ensembles of AMIP type experiments, Climate Dynamics, 28, 125–145, https://doi.org/10.1007/s00382-006-0172-2, 2007.

Courtier, P., Thépaut, J. M., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society, 120, 1367–1387, https://doi.org/10.1002/qj.49712051912, 1994.

Craig, A. P., Vertenstein, M., and Jacob, R.: A new flexible coupler for Earth system modeling developed for CCSM4 and CESM1, International Journal of High Performance Computing Applications, 26(1), 31–42, https://doi.org/10.1177/1094342011428141, 2012.

Dee, D. P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A. J., and Thépaut, J. N.: Toward a consistent reanalysis of the climate system, Bulletin of the American Meteorological Society, 95(8), 1235-1248, https://doi.org/10.1175/BAMS-D-13-00043.1, 2014.

Dennis, J. M., Edwards, J., Loy, R., Jacob, R., Mirin, A. A., Craig, A. P., and Vertenstein, M.: An application-level parallel I/O library for Earth system models, International Journal of High Performance Computing Applications, 26(1), 43–53, https://doi.org/10.1177/1094342011428143, 2012.

Du, H., Doblas-Reyes, F. J., García-Serrano, J., Guemas, V., Soufflet, Y., and Wouters, B.: Sensitivity of decadal predictions to the initial atmospheric and oceanic perturbations, Climate Dynamics, 39(7), 2013–2023, https://doi.org/10.1007/s00382-011-1285-9, 2012.

Du, J., Kimball, J. S., Velicogna, I., Zhao, M., Jones, L. A., Watts, J. D., and Kim, Y.: Multicomponent satellite assessment of drought severity in the contiguous United States from 2002 to 2017 using AMSR-E and AMSR2, Water Resources Research, 55(7), 5394–5412, https://doi.org/10.1029/2018WR024633, 2019.

Edwards, A. C., Scalenghe, R., and Freppaz, M.: Changes in the seasonal snow cover of alpine regions and its effect on soil processes: a review, Quaternary International, 162, 172–181, https://doi.org/10.1016/j.quaint.2006.10.027, 2007.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fowler, A. M., and Lawless, A. S.: An idealized study of coupled atmosphere–ocean 4D-Var in the presence of model error, Monthly Weather Review, 144(10), 4007–4030, https://doi.org/10.1175/MWR-D-15-0420.1, 2016.

Fujii, Y., Nakaegawa, T., Matsumoto, S., Yasuda, T., Yamanaka, G., and Kamachi, M.: Coupled climate simulation by constraining ocean fields in a coupled model with ocean data, Journal of Climate, 22, 5541–5557, https://doi.org/10.1175/2009JCLI2814.1, 2009.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., and Yang, Y.: Fusing Landsat and MODIS data for vegetation monitoring, IEEE Geoscience and Remote Sensing Magazine, 3(3), 47–60, https://doi.org/10.1109/MGRS.2015.2434351, 2015.

Girotto, M., De Lannoy, G. J., Reichle, R. H., and Rodell, M.: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, Water Resources Research, 52(5), 4164–4183, https://doi.org/10.1002/2015WR018417, 2016.

Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-Smith, P. J., Donahue, A. S., Deakin, M., Easter, R. C., Evans, K. J., Feng, Y., Flanner, M., Foucar, J. G., Fyke, J. G., Griffin, B. M., Hannay, C., Harrop, B. E., Hoffman, M. J., Hunke, E. C., Jacob, R. L., Jacobsen, D. W., Jeffery, N., Jones, P. W., Keen, N. D., Klein, S. A., Larson, V. E., Leung, L. R., Li, H. Y., Lin, W., Lipscomb, W. H., Ma, P. L., Mahajan, S., Maltrud, M. E., Mametjanov, A., McClean, J. L., McCoy, R. B., Neale, R. B., Price, S. F., Qian, Y., Rasch, P. J., Reeves Eyre, J. E. J., Riley, W. J., Ringler, T. D., Roberts, A. F., Roesler, E. L., Salinger, A. G., Shaheen, Z., Shi, X., Singh, B., Tang, J., Taylor, M.

删除了:

A., Thornton, P. E., Turner, A. K., Veneziani, M., Wan, H., Wang, H., Wang, S., Williams, D. N., Wolfram, P. J., Worley, P. H., Xie, S., Yang, Y., Yoon, J.-H., Zelinka, M. D., Zender, C. S., Zeng, X., Zhang, C., Zhang, K., Zhang, Y., Zheng, X., Zhou, T., and Zhu, Q.: The DOE E3SM Coupled Model Version 1: Overview and Evaluation at Standard Resolution, Journal of Advances in Modeling Earth Systems, 11, 2089–2129, https://doi.org/https://doi.org/10.1029/2018MS001603, 2019.

Golaz, J. C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W. Y., Bradley, A. M., Tang, Q., Maltrud, M. E., Forsyth, R. M., Zhang, C. Z., Zhou, T., Zhang, K., Zender, C. S., Wu, M. X., Wang, H. L., Turner, A. K., Singh, B., Richter, J. H., Qin, Y., Petersen, M. R., Mametjanov, A., Ma, P., Larson, V. E., Krishna, J., Keen, N. D., Jeffery, N., Hunke, E. C., Hannah, W. M., Guba, O., Griffin, B. M., Feng, Y., Engwirda, D., Vittorio, A. V., Cheng, D., Conlon, L. M., Chen, C., Brunke, M. A., Bisht, G., Benedict, J. J., Asay-Davis, X. S., Zhang, Y. Y., Zhang, M., Zeng, X. B., Xie, S. C., Wolfram, P. J., Vo, T., Veneziani, M., Tesfa, T. K., Sreepathi, S., Salinger, A. G., Jack Reeves Eyre, J. E., Prather, M. J., Mahajan, S., Li, Q., Jones, P. W., Jacob, R. L., Huebler, G. W., Huang, X. L., Hillman, B. R., Harrop, B. E., Foucar, J. G., Fang, Y. L., Comeau, D. S., Caldwell, P. M., Bartoletti, T., Balaguru, K., Taylor, M. A., McCoy, R. B., Leung, L. R., and Bader, D. C.: The DOE E3SM Model version 2: Overview of the physical model and initial model evaluation, Journal of Advances in Modeling Earth Systems, 14, e2022MS003156, https://doi.org/10.1029/2022MS003156, 2022.

Guo, Z., Dirmeyer, P. A., Delsole, T., and Koster, R. D.: Rebound in atmospheric predictability and the role of the land surface, Journal of Climate, 25(13), 4744–4749, https://doi.org/10.1175/JCLI-D-11-00651.1, 2012.

He, Y., Wang, B., Liu, M., Liu, L., Yu, Y., Liu, J., Li, R., Zhang, C., Xu, S., Huang, W., Liu, Q., Wang, Y., and Li, F.: Reduction of initial shock in decadal predictions using a new initialization strategy, Geophysical Research Letters, 44(16), 8538–8547, https://doi.org/10.1002/2017GL074028, 2017.

He, Y., Wang, B., Liu, L., Huang, W., Xu, S., Liu, J., Wang, Y., Li, L., Huang, X., Peng, Y., Lin, Y., and Yu, Y.: A DRP-4DVar-based coupled data assimilation system with a simplified off-line localization technique for decadal predictions, Journal of Advances in Modeling Earth Systems, 12(4), e2019MS001768, https://doi.org/10.1029/2019MS001768, 2020a.

He, Y., Wang, B., Huang, W., Xu, S., Wang, Y., Liu, L., Li, L., Liu, J., Yu, Y., Lin, Y., Huang, X., and Peng, Y.: A new DRP-4DVar-based coupled data assimilation system for decadal predictions using a fast online localization technique, Climate Dynamics, 54, 3541–3559, https://doi.org/10.1007/s00382-020-05190-w, 2020b.

Hoke, J. E. and Anthes, R. A.: The initialization of numerical models by a dynamic-initialization technique, Monthly Weather Review, 104(12), 1551–1556, https://doi.org/10.1175/1520-0493(1976)104<1551:TIONMB>2.0.CO;2, 1976.

Houborg, R., Rodell, M., Li, B., Reichle, R., and Zaitchik, B. F.: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, Water Resources Research, 48, W07525, https://doi.org/10.1029/2011WR011291, 2012.

Hu, S., Zhou, T., and Wu, B.: Improved ENSO prediction skill resulting from reduced climate drift in IAP-DecPreS: A comparison of full-field and anomaly initializations, Journal of Advances in Modeling Earth Systems, 12, e2019MS001759, https://doi.org/10.1029/2019MS001759, 2020.

Ireson, A. M., Van Der Kamp, G., Ferguson, G., Nachshon, U., and Wheater, H. S.: Hydrogeological processes in seasonally frozen northern latitudes: understanding, gaps and challenges, Hydrogeology Journal, 21, 53–66, https://doi.org/10.1007/s10040-012-0916-5, 2013.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., and Morisette, J. T.: An overview of MODIS Land data processing and product status, Remote Sensing of Environment, 83, 3–15, https://doi.org/10.1016/S0034-4257(02)00084-6, 2002.

Kimmritz, M., Counillon, F., Bitz, C. M., Massonnet, F., Bethke, I., and Gao, Y.: Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model, Tellus, 70A, 1435945, https://doi.org/10.1080/16000870.2018.1435945, 2018.

Kwon, Y., Yang, Z. L., Zhao, L., Hoar, T. J., Toure, A. M., and Rodell, M.: Estimating snow water storage in North America using CLM4, DART, and snow radiance data assimilation, Journal of Hydrometeorology, 17(11), 2853–2874, https://doi.org/10.1175/JHM-D-16-0028.1, 2016.

Lea, D. J., Mirouze, I., Martin, M. J., King, R. R., Hines, A., Walters, D., and Thurlow, M.: Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere–land–ocean–

删除了：

删除了： Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P.: A coupled data assimilation system for climate reanalysis, Quarterly Journal of the Royal Meteorological Society, 142(694), 65-78, https://doi.org/10.1002/qj.2629, 2016.

sea ice model, Monthly Weather Review, 143(11), 4678-4694, https://doi.org/10.1175/MWR-D-15-0174.1, 2015.

Lei, L. L. and Hacker, J. P.: Nudging, ensemble, and nudging ensembles for data assimilation in the presence of model error, Monthly Weather Review, 143(7), 2600–2610, https://doi.org/10.1175/MWR-D-14-00295.1, 2015.

Leung, L. R., Bader, D. C., Taylor, M. A., and McCoy, R. B.: An introduction to the E3SM special collection: Goals, science drivers, development, and analysis, Journal of Advances in Modeling Earth Systems, 12(11), e2019MS001821, https://doi.org/10.1029/2019MS001821, 2020.

Li, F., Wang, B., He, Y., Huang, W., Xu, S., Liu, L., Liu, J. and Li, L.: Important role of North Atlantic air–sea coupling in the interannual predictability of summer precipitation over the eastern Tibetan Plateau, Climate Dynamics, 56, 1433–1448, https://doi.org/10.1007/s00382-020-05542-6, 2021.

Li, H. Y., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., and Leung, L. R.: A physically based runoff routing model for land surface and Earth system models, Journal of Hydrometeorology, 14, 808–828, https://doi.org/10.1175/JHM-D-12-015.1, 2013.

Lin, L. F., Ebtehaj, A. M., Wang, J., and Bras, R. L.: Soil moisture background error covariance and data assimilation in a coupled land-atmosphere model, Water Resources Research, 53(2), 1309–1335, https://doi.org/10.1002/2015WR017548, 2017.

Lin, P., Yang, Z. L., Wei, J., Dickinson, R. E., Zhang, Y., and Zhao, L.: Assimilating multi-satellite snow data in ungauged Eurasia improves the simulation accuracy of Asian monsoon seasonal anomalies, Environmental Research Letters, 15(6), 064033, https://doi.org/10.1088/1748-9326/ab80ef, 2020.

Liu, D., and Mishra, A. K.: Performance of AMSR_E soil moisture data assimilation in CLM4. 5 model for monitoring hydrologic fluxes at global scale, Journal of Hydrometeorology, 547, 67–79, https://doi.org/10.1016/j.jhydrol.2017.01.036, 2017.

Liu, J. J., Wang, B., and Xiao, Q. N.: An evaluation study of the DRP-4-DVar approach with the Lorenz-96 model, Tellus A: Dynamic Meteorology and Oceanography, 63, 256–262, https://doi.org/10.1111/j.1600-0870.2010.00487.x, 2011.

Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., and Fairbairn, D.: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP, Monthly Weather Review, 143, 212–229, https://doi.org/10.1175/MWR-D-14-00195.1, 2015.

McCabe, M. F., Wood, E. F., Wójcik, R., Pan, M., Sheffield, J., Gao, H., and Su, H.: Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies, Remote Sensing of Environment, 112(2), 430–444, https://doi.org/10.1016/j.rse.2007.03.027, 2008.

Mertes, C. M., Schneider, A., Sulla-Menashe, D., Tatem, A. J., and Tan, B.: Detecting change in urban areas at continental scales with MODIS data, Remote Sensing of Environment, 158, 331–347, https://doi.org/10.1016/j.rse.2014.09.023, 2015.

Mochizuki, T., Masuda, S., Ishikawa, Y., and Awaji, T.: Multiyear climate prediction with initialization based on 4D-Var data assimilation, Geophysical Research Letters, 43(8), 3903–3910, https://doi.org/10.1002/2016GL067895, 2016.

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V.: Soil moisture retrieval from AMSR-E, IEEE Transactions on Geoscience and Remote Sensing, 41(2), 215–229, https://doi.org/10.1109/TGRS.2002.808243, 2003.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J. F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J., and Yang, Z. L.: Technical description of version 4.5 of the Community Land Model (CLM) (Tech. Rep. NCAR/TN-503+STR). Boulder, Colorado, USA: National Center for Atmospheric Research, http://dx.doi.org/10.5065/D6RR1W7M, 2013.

Penny, S. G., and Hamill, T. M.: Coupled data assimilation for integrated earth system analysis and prediction, Bulletin of the American Meteorological Society, 98(7), ES169-ES172, https://doi.org/10.1175/BAMS-D-17-0036.1, 2017.

Penny, S. G., Bach, E., Bhargava, K., Chang, C. C., Da, C., Sun, L., and Yoshida, T.: Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model, Journal of Advances in Modeling Earth Systems, 11(6), 1803-1829, https://doi.org/10.1029/2019MS001652, 2019.

Petersen, M., Asay-Davis, X. S., Jacobsen, D., Maltrud, M., Ringler, T., Van Roekel, L., and Wolfram, P.: MPAS ocean user’s guide V6, Zenodo, https://doi.org/10.5281/zenodo.1246893, 2018.

Polkova, I., Köhl, A., and Stammer, D.: Climate-mode initialization for decadal climate predictions, Climate Dynamics, 53, 7097–7111, https://doi.org/10.1007/s00382-019-04975-y, 2019.

Reckinger, S. M., Petersen, M. R., and Reckinger, S. J.: A study of overflow simulations using MPAS-Ocean: Vertical grids, resolution, and viscosity, Ocean Modeling, 96, 291–313, https://doi.org/10.1016/j.ocemod.2015.09.006, 2015.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS aerosol algorithm, products, and validation, Journal of the Atmospheric Sciences, 62(4), 947–973, https://doi.org/10.1175/JAS3385.1, 2005.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M. and Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bulletin of the American Meteorological society, 85(3), 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Sakaguchi, K., Zeng, X., and Brunke, M. A.: The hindcast skill of the CMIP ensembles for the surface air temperature trend, Journal of Geophysical Research: Atmospheres, 117, D1611, https://doi.org/10.1029/2012JD017765, 2012.

Shi, P. F., Wang, B., He, Y., Lu, H., Yang, K., Xu, S. M., Huang, W. Y., Liu, L., Liu, J. J., Li, L. J., and Wang, Y.: Contributions of weakly coupled data assimilation–based land initialization to interannual predictability of summer climate over Europe, Journal of Climate, 35, 517–535, https://doi.org/10.1175/JCLI-D-20-0506.1, 2022.

Simmons, A. J. and Hollingsworth, A.: Some aspects of the improvement in skill of numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 128(580), 647–677, https://doi.org/10.1256/003590002321042135, 2002.

Sluka, T. C., Penny, S. G., Kalnay, E., and Miyoshi, T.: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation, Geophysical Research Letters, 43(2), 752-759, https://doi.org/10.1002/2015GL067238, 2016.

Smith, P. J., Fowler, A. M., and Lawless, A. S.: Exploring strategies for coupled 4DVar data assimilation using an idealised atmosphere–ocean model, Tellus A, 67(1):27025, https://doi.org/10.3402/tellusa.v67.27025, 2015.

Smith, D. M., Eade, R., and Pohlmann, H.: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction, Climate Dynamics, 41, 3325–3338, https://doi.org/10.1007/s00382-013-1683-2, 2013.

Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., Igarashi, H. and Ishikawa, Y.: Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations, Journal of Geophysical Research: Oceans, 113, C10017, https://doi.org/10.1029/2008JC004741, 2008.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93(4), 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., and Eldrid, C.: An energy consistent discretization of the nonhydrostatic equations in primitive variables, Journal of Advances in Modeling Earth Systems, 12, e2019MS001783, https://doi.org/10.1029/2019MS001783, 2020.

Volpi, D., Guemas, V., and Doblas-Reyes, F. J.: Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state, Climate Dynamics, 49, 1181–1195, https://doi.org/10.1007/s00382-016-3373-3, 2017.

Wang, B., Liu, J., Wang, S., Cheng, W., Liu, J., Liu, C., Xiao, Q., and Kuo, Y. H.: An economical approach to four-dimensional variational data assimilation, Advances in Atmospheric Sciences, 27, 715–727, https://doi.org/10.1007/s00376-009-9122-3, 2010.

Wang, B., Liu, J., Liu, L., Xu, S., and Huang, W.: An approach to localization for ensemble-based data assimilation, PloS one, 13(1), e0191088, https://doi.org/10.1371/journal.pone.0191088, 2018.

Wang, G., Dolman, A. J., Blender, R., and Fraedrich, K.: Fluctuation regimes of soil moisture in ERA-40 reanalysis data, Theoretical and Applied Climatology, 99, 1–8, https://doi.org/10.1007/s00704-009-0111-3, 2010.

Yao, Y., Luo, Y., Huang, J., and Ma, J.: Improving the downscaled springtime temperature in Central Asia through assimilating meteorological and snow cover observations, Atmospheric Research, 258, 105619, https://doi.org/10.1016/j.atmosres.2021.105619, 2021.

Yin, J., Zhan, X., Zheng, Y., Liu, J., Hain, C. R., and Fang, L.: Impact of quality control of satellite soil moisture data on their assimilation into land surface model, Geophysical Research Letters, 41(20), 7159–7166, https://doi.org/10.1002/2014GL060659, 2014.

Yoshida, T., and Kalnay, E.: Correlation-cutoff method for covariance localization in strongly coupled data assimilation, Monthly Weather Review, 146(9), 2881-2889, https://doi.org/10.1175/MWR-D-17-0365.1, 2018.

Zeng, X., and Decker, M.: Improving the numerical solution of soil moisture–based Richards equation for land models with a deep or shallow water table, Journal of Hydrometeorology, 10, 308–319, https://doi.org/10.1175/2008JHM1011.1, 2009.

Zhang, H., Zhang, L. L., Li, J., An, R. D., Deng, Y.: Climate and Hydrological Change Characteristics and Applicability of GLDAS Data in the Yarlung Zangbo River Basin, China, Water, 10, 254, https://doi.org/10.3390/w10030254, 2018.

Zhang, S., Harrison, M. J., Wittenberg, A. T., Rosati, A., Anderson, J. L., and Balaji, V.: Initialization of an ENSO forecast system using a parallelized ensemble filter, Monthly Weather Review, 133(11), 3176-3201, https://doi.org/10.1175/MWR3024.1, 2005.

Zhang, S., Harrison, M. J., Rosati, A., and Wittenberg, A.: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies, Monthly Weather Review, 135(10), 3541-3564, https://doi.org/10.1175/MWR3466.1, 2007.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, S., Lu, F., Li, M., Deng, X.: Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review, Climate Dynamics, 54, 5127-5144, https://doi.org/10.1007/s00382-020-05275-6, 2020.

Zhou, J., Yang, K., Crow, W.T., Dong, J., Zhao, L., Feng, H., Zou, M., Lu, H., Tang, R. and Jiang, Y.: Potential of remote sensing surface temperature-and evapotranspiration-based land-atmosphere coupling metrics for land surface model calibration, Remote Sensing of Environment, 291, 113557, https://doi.org/10.1016/j.rse.2023.113557, 2023.

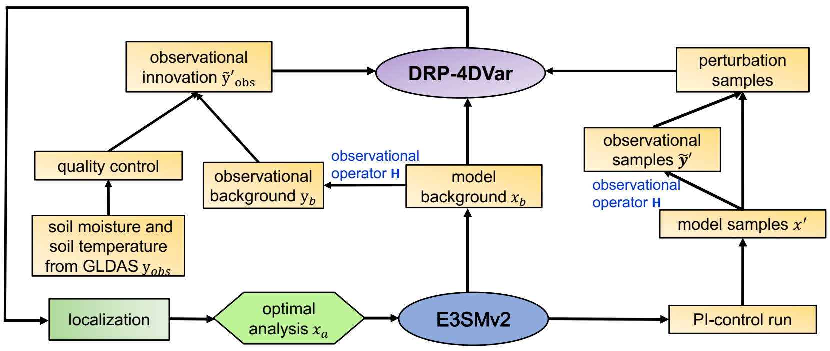

Figure 1. Flowchart of the 4DEnVar-based WCLDA system in E3SMv2 based on the DRP-4DVar method.

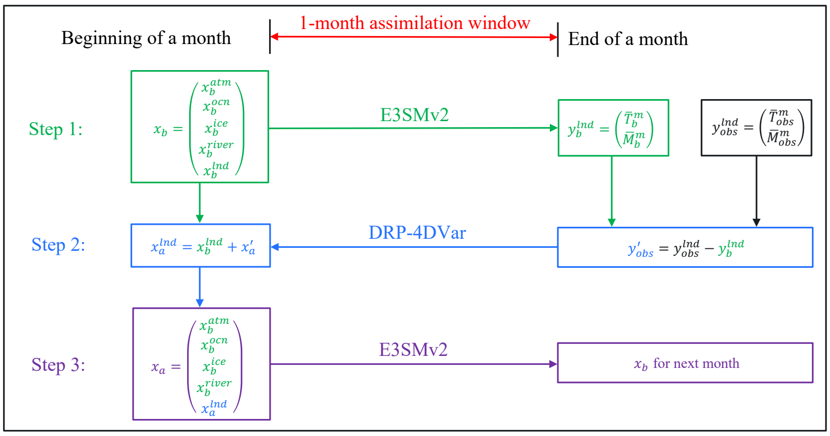

Figure 2. Schematic flowchart of the 4DEnVar-based WCLDA system. The beginning of a month is at 0000 UTC on the first day of the month, and the end of the month is at 0000 UTC on the first day of the next month. x_b denotes the background vector including the backgrounds of all E3SMv2 components (atmosphere (x_b^{atm}), ocean (x_b^{ocn}), sea ice (x_b^{ice}), river transport (x_b^{river}) and land surface (x_b^{lnd})). x_a consists of the assimilation analysis of land surface (x_a^{lnd}) and the backgrounds of other components. y_b^{lnd} represents the simulated monthly mean soil temperature ($\overline{T}_b^m$) and moisture ($\overline{M}_b^m$) by E3SMv2 using x_b as the initial condition. y_{obs}^{lnd} denotes the monthly mean GLDAS data of soil temperature ($\overline{T}_{obs}^m$) and moisture ($\overline{M}_{obs}^m$). y'_{obs} denotes the observational innovation, which is the difference between the GLDAS data (y_{obs}^{lnd}) and the observational background (y_b^{lnd}).

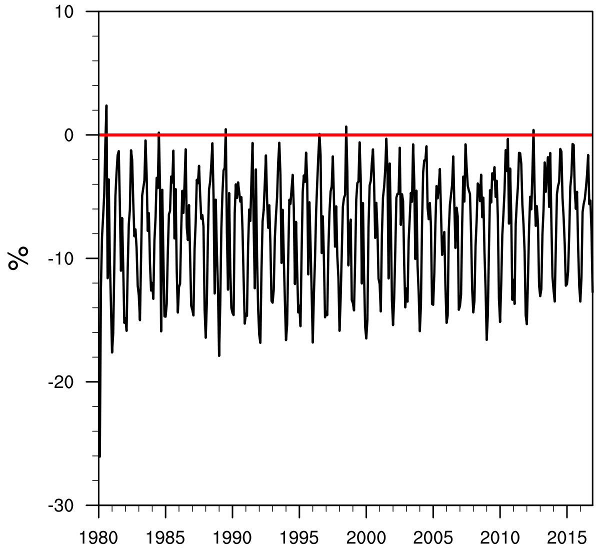

Figure 3. Time series of the reduction rate of the cost function from 1980 to 2016 in the 4DEnVar-based WCLDA system.

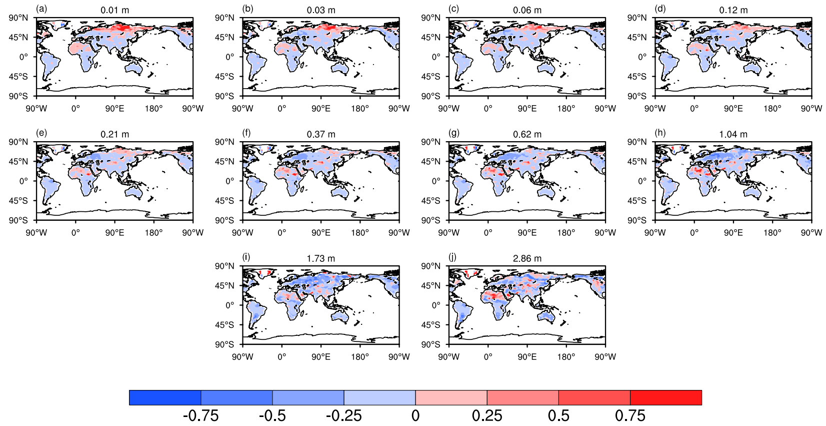

Figure 4. Spatial distribution of the AE index for soil moisture from the surface to deep layers during the 1980-2016 period. The number at the top center denotes the depth of each soil layer.

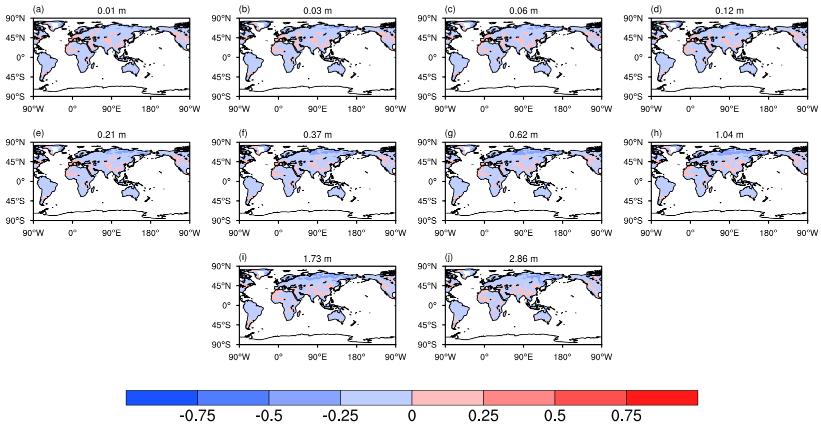

Figure 5. Same as in Figure 4, but for soil temperature.

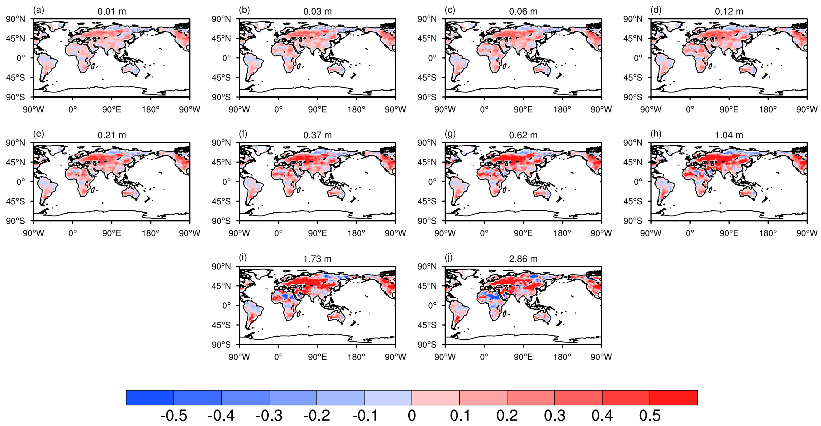

Figure 6. Differences between correlations of soil moisture in Assim and CTRL with the GLDAS data from the surface to deep layers for the period of 1980-2016. The number at the top center denotes the depth of each soil layer.

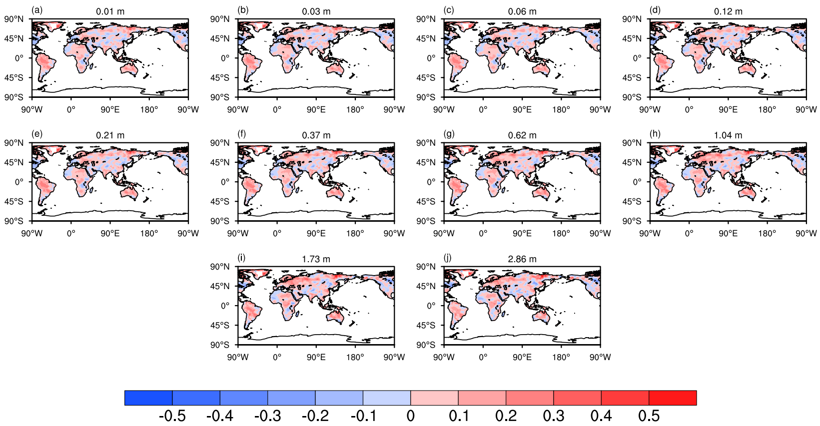

Figure 7. Same as in Figure 6, but for soil temperature.

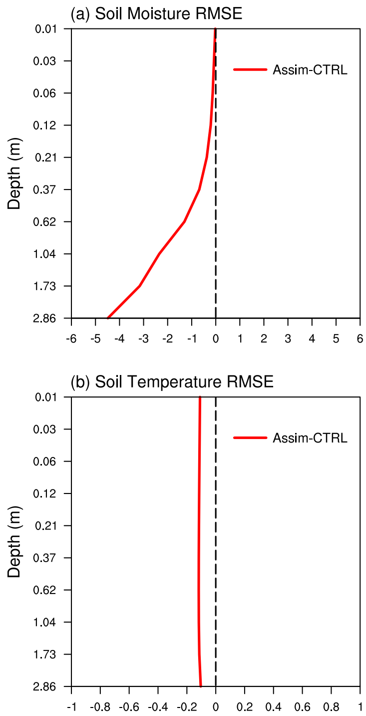

Figure 8. Vertical distributions of RMSE differences (Assim minus CTRL) for (a) soil moisture and (b) soil temperature averaged over the global land during the 1980-2016 period.

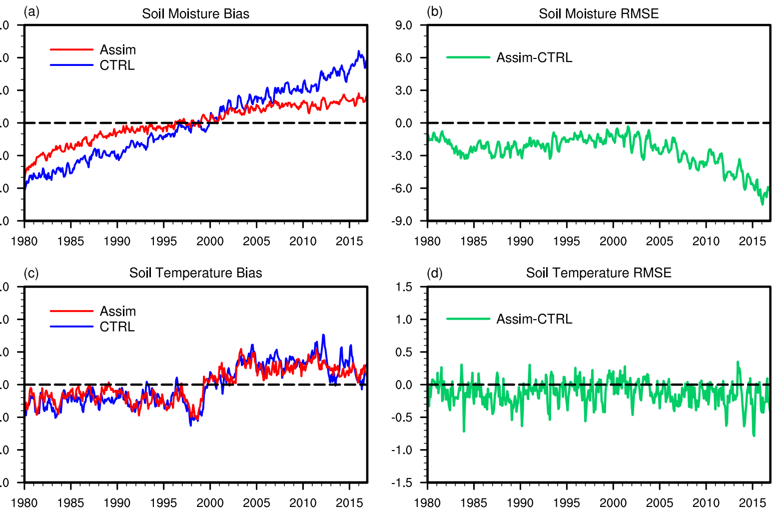

Figure 9. Time series of the vertically averaged global mean soil moisture and temperature bias (left) for Assim (red line) and CTRL (blue line), and RMSE differences (right, green line) between Assim and CTRL from 1980 to 2016.

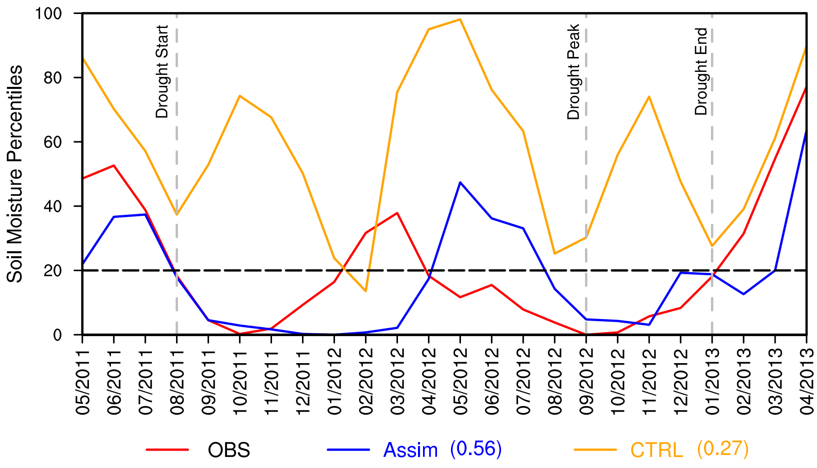

Figure 10. Time series of soil moisture percentiles between May 2011 and April 2013 during the 2012 U.S. Midwest drought. Red line: observation, blue line: Assim, orange line: CTRL. The correlation coefficients of Assim and CTRL with observations are also shown. The three vertical dashed lines mark the timing of drought start, drought peak and drought end, respectively. The start of the agricultural drought is defined as the month when soil moisture falls below the 20th percentile. The soil moisture percentiles are averaged over the U.S. Midwest (36°-50°N, 102°-88°W). The observed soil moisture is derived from ERA-Interim monthly soil moisture data.

Appendix A: Supporting Information

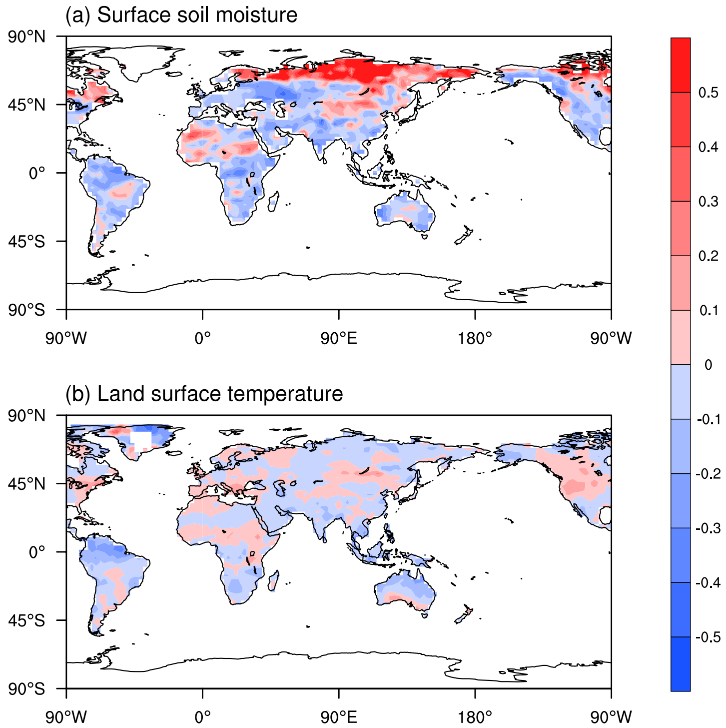

Figure A1. Spatial distribution of the AE index for (a) surface soil moisture and (b) land surface temperature during the 2003-2014 period. The surface soil moisture and land surface temperature are derived from monthly AMSR and MODIS satellite data, respectively.