https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access

The 4DEnVar-based land coupled 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)

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access

Abstract. A new land coupled data assimilation (LCDA) 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. Monthly mean soil moisture and temperature analyses from a global land reanalysis product are assimilated into the land component of E3SMv2 with a one-month assimilation window along the coupled model trajectory from 1980 to 2016. The coupled assimilation experiment is evaluated using multiple metrics, including the cost function, assimilation efficiency index, correlation, root mean square error and bias, and compared with a control simulation without land data assimilation. The LCDA 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. 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 LCDA system is intended to be a foundational resource to investigate land-derived climate predictability for future prediction research by the E3SM community.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions

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 forcing (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 models for practicable Earth system prediction, including uncoupled and coupled data assimilation (CDA) methods. Some modeling centers employ uncoupled initialization methods that directly utilize reanalysis data or stand-alone model states driven by observations as initial conditions (ICs) (Du et al., 2012; Prodhomme et al., 2016). However, ICs derived from uncoupled methods often exhibit poor consistency between model components (Balmaseda et al., 2009). Initializing a coupled model with data obtained from another model may result in initial shocks due to inconsistencies and eventually produce low prediction skills (Boer et al., 2016). A more effective initialization would involve performing a CDA with observations for each coupled model individually (Ardilouze et al., 2017). The CDA methods incorporate observations into one or several components of the coupled model through data assimilation techniques, with long-term assimilation cycles executed under the coupled modeling framework (He et al., 2020a). The CDA method outperforms the uncoupled method due to the constraint of the coupled model, leading to better consistency of the ICs with the coupled model (He et al., 2020b).

The CDA approaches for initializing coupled models are becoming increasingly prevalent, using a diverse range of data assimilation techniques. Most of these methods utilize simple nudging or nudging-based Incremental Analysis Update (IAU) approaches where analysis increments into a model integration are incorporated in a gradual manner (Bloom et al., 1996; Shaffrey et al., 2017; Smith et al., 2013). Both techniques restore the model states to observations by introducing new terms that are proportional to the discrepancy between observations and model states in the prognostic equations (Hoke and Anthes, 1976). These techniques are time-saving and easy to implement, but the principal disadvantages of these

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions Open Access EGU

methods are the necessity to interpolate observations at every time step and the reliance on experience and experimentation to determine the nudging coefficients (He et al., 2017; Wei et al., 2017). Some modeling centers have developed more advanced CDA systems using variational and filtering approaches, such as the three-dimensional variational data assimilation (3DVar) (Lin et al., 2017; Yao et al., 2021) and ensemble-based techniques like the ensemble Kalman filter (EnKF) (Santanello et al., 2016) or ensemble optimal interpolation (EnOI) (Wu et al., 2018). The former generally utilizes the stationary 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). The objective of four-dimensional variational data assimilation (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 the 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, few modeling centers utilize 4DVar-based initialization methods because of the challenge of adjoint calculation and its high computational cost.

To capitalize on the strengths of both ensemble and variational techniques, there has been a growing interest in developing new hybrid data assimilation methods. One notable example is the hybrid EnKF-3DVar method introduced by Hamill and Snyder (2000), which combines stationary covariances from 3DVar with flow-dependent covariances obtained from short-range forecasts. Another hybrid approach is the Ensemble Transform Kalman Filter (ETKF)-3DVar, proposed by Wang et al. (2008). This method merges ensemble covariances with stationary covariances using the extended control variable technique and preserves ensemble perturbations through the ETKF. Lastly, Liu et al. (2008) developed the four-dimensional ensemble-variational (4DEnVar) algorithm as an additional hybrid method. This technique 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).

In this study, we introduce the development of the 4DEnVar-based land coupled data assimilation

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

(LCDA) system for the Energy Exascale Earth System Model version 2 (E3SMv2) (Golaz et al., 2022). The 4DEnVar method in this LCDA 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 LCDA system, monthly mean soil moisture and temperature data from a global land reanalysis product are assimilated to constrain the land fields of a coupled climate model with a one-month assimilation window. The primary goal of the LCDA 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. This LCDA 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 LCDA system for the land component of E3SMv2. In Sect. 2, we provide an overview of the E3SMv2 model, describe the 4DEnVar methodology in detail and outline the framework of the 4DEnVar-based LCDA system. Preliminary evaluation of the LCDA system is presented in Sect. 3. Finally, major conclusions are discussed in Sect. 4.

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

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

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 Observational Dataset

Monthly mean soil moisture and soil temperature data used in this study are produced by the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). The GLDAS products generate 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 ELM, we utilize GLDAS data produced by CLM.

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).

Following Wang et al. (2010), the original 4DVar can be implemented to produce the optimal analysis in the sample space by minimizing a new cost function:

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions

$$x_a = x_b + x'_a = x_b + P_x \alpha_a \tag{1}$$

$$\tilde{J}(\alpha_a) = \min_{\alpha \epsilon E_m} \tilde{J}(\alpha) \tag{2}$$

$$\tilde{J}(\alpha) = \frac{1}{2}\alpha^T B_\alpha^{-1} \alpha + \frac{1}{2}\left(P_y \alpha - \tilde{y}'_{obs}\right)^T \left(P_y \alpha - \tilde{y}'_{obs}\right) \tag{3}$$

The optimal solution to the aforementioned minimization problem is formulated as:

$$\alpha_a = (B_\alpha^{-1} + P_y^T P_y)^{-1} P_y^T \tilde{y}'_{obs} \tag{4}$$

Here, x_a, x_b, and x'_a represent the optimal analysis, background, and analysis increment, respectively; P_x is the projection matrix comprised of initial perturbation samples; α is the weight coefficients; the superscript T represents the transpose; B denotes the background error covariance matrix; P_y is the projection matrix consisting of observational perturbation samples; $\tilde{y}'_{obs}$ represents the weighted observational innovation.

2.4 4DEnVar-based LCDA System

The 4DEnVar-based LCDA system is developed to assimilate the full-field 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 LCDA system. The control simulation (CTRL) is a 36-year freely coupled integration driven by observed external forcing from 1980 to 2016. CTRL provides the benchmark for assessing the performance of the LCDA system. The assimilation experiment (Assim) is conducted from 1980 to 2016 based on the LCDA 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 LCDA 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).

The flowchart of the 4DEnVar-based LCDA 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 monthly initial states before assimilation, and the observational operator (H)

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

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 ($\tilde{y}'_{obs}$) is determined by calculating the differences in soil moisture and temperature between the monthly mean GLDAS data (y_{obs}) and model outputs (y_b). From the 100-year sample database of the E3SMv2 Pre-industrial Control (PI-CTRL) simulation, 30 monthly mean perturbation samples ($\tilde{y}'$) are chosen according to 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. To alleviate the spurious correlations, a localization scheme is implemented in the 4DEnVar-based LCDA system (Wang et al., 2018).

The schematic diagram in Figure 2 outlines the assimilation process of the 4DEnVar-based LCDA system in E3SMv2. The 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 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 cycle. Due to multi-component interactions during the one-month free coupled integration, the observed land information can potentially benefit other components (e.g., atmosphere and ocean) in the coupled modeling framework (Li et al., 2021; Shi et al., 2022).

2.5 Evaluation Metrics

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

$$\frac{\frac{1}{2}(y_{obs}-y_a)^T R^{-1}(y_{obs}-y_a)-\frac{1}{2}(y_{obs}-y_b)^T R^{-1}(y_{obs}-y_b)}{\frac{1}{2}(y_{obs}-y_b)^T R^{-1}(y_{obs}-y_b)} \times 100\% \qquad (5)$$

where 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. Negative value for this metric indicates that observational information has been correctly incorporated into the model variables.

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

$$AE = \frac{RMSE_{Assim}}{RMSE_{CTRL}} - 1 \tag{6}$$

In this equation, $RMSE_{Assim}$ is the root mean square error (RMSE) between Assim and GLDAS data, while $RMSE_{CTRL}$ represents the RMSE between CTRL and GLDAS data. Negative (positive) AE value indicates improvements (degradations) by the assimilation. In the following sections, we continue to use the GLDAS data as the reference dataset to verify the correctness of the LCDA system.

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 LCDA system. In the first month, the reduction rate reaches approximately 28.8% in Assim. Over the subsequent months, Assim maintains the average reduction rate of 8.5% throughout the entire period. Furthermore, negative reduction rates are observed in 96% of the total months, indicating the effectiveness of the LCDA system. These results suggest that the LCDA system is correctly implemented, with the observational data successfully assimilated into the coupled model.

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 from the second to the eighth layer, suggesting the reduction in RMSE after assimilation. Significant improvements appear over North America, Northern Africa, Europe, and Northern Asia. The largest improvement in these soil layers is observed in the northern part of the Eurasian continent. However, assimilation performance is degraded in South America and monsoon regions (e.g., East Asia and India). This is consistent with the findings in other studies that assimilation updates in monsoon regions are limited due to the dominant impact of monsoon circulations (Timouk et al., 2009; Brocca et al., 2017). The first soil layer, which is highly susceptible to atmospheric forcing, also shows degradation in large areas. Furthermore, some degradations are found in the deep

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

layers, especially the ninth and tenth layers. This may be linked to the quality of assimilation data and other terrestrial factors, as noted 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 first to the ninth layer are dominated by negative AE signals, indicating improved performance 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 Eastern Russia, Europe, North America, Australia, and large parts of Eurasia. In contrast, slight degradations appear over Northwestern Africa, Southern South America and Saudi Arabia. This may be partly related to assimilation uncertainties and possible atmospheric noise, as shown by many past studies (Kwon et al., 2016; Lin et al., 2020). Some locations with degradation are also noted in the tenth layer, which still requires 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 observations across different soil layers. A 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 LCDA system. Enhanced correlations in deep soil layers are more prominent than in shallow layers, which may be attributed to the longer memory of soil processes in the deeper layers (Wang et al., 2010). Improved correlations appear over Northern Africa, North America, Eurasia, and Australia. However, some scattered areas show slight degradations, such as 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, Western Russia, Northern Africa, North America, and Central Eurasia.

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 ninth layer across the global domain, with the exception of the northern region of the Eurasian continent. Furthermore, similar spatial patterns and magnitudes are observed in the performance of different soil layers except for the tenth layer, implying the significant heat transfer from the surface to deep zone that constrains

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

soil temperature across the soil column. Notable improvements are located over South America, North America, Northern Africa, Australia, and Southern Eurasia. Nevertheless, some degradations appear over Central Africa, Eastern Russia, and part of South China. Obvious degradations are also found in the tenth layer. The diminished performance may come from uncertainties in the assimilation data and imbalances between land variables during data assimilation, as supported by the findings of other studies (Park et al., 2018; Zhang et al., 2014). 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, Southern Eurasia, Australia, and North America.

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. Assim shows noticeable improvements with reduced RMSE for soil moisture and temperature at all vertical levels compared with CTRL. For soil moisture, the reduction of RMSE increases with depth from the upper to middle levels, reaching its maximum at the eighth layer. However, this value then decreases as the depth extends further into the tenth layer. This decrease is likely due to the overestimation of observation errors in deep soil layers. For soil temperature, the reduction of RMSE exhibits similar magnitude across shallow layers, which may be explained by the heat transfer process within the soil. From the middle to deep levels, this reduction initially increases with depth, peaking at the eighth layer, and then gradually decreases. In the ninth and tenth layers, there is potential for further improvement in assimilation performance.

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 most months, with 91.7% of the total months in Assim exhibiting lower

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

RMSE than CTRL. In summary, the superior performance for both soil moisture and temperature in Assim demonstrates that land observational information has been effectively incorporated into the model variables through the LCDA 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 in E3SMv1 and v2 during the past decades 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 LCDA system, we preliminarily 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 the observation based on 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.

The free running CTRL experiment fails to simulate the temporal evolution of the 2012 Midwest drought, with a correlation coefficient between CTRL and observation of only 0.27. The onset and peak of the drought are remarkably well captured by Assim, although the drought recovery occurs one month earlier than observed. The correlation coefficient of the Assim time series with observation is 0.61, 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 LCDA system.

We further compare the time series of observed and simulated precipitation anomaly over the Midwest during the 2012 U.S. Midwest drought (Figure 11). As a free running simulation, the precipitation in CTRL is not expected to reproduce the overall dry anomaly in observation. It is noteworthy that the magnitude of the precipitation anomaly is remarkably well captured by Assim. More specifically, Assim can reproduce the positive precipitation anomaly from February 2012 to April 2012 and the dry anomaly from May 2012 to October 2012. The correlation coefficient of the Assim time series with observation is 0.40, much higher than that of CTRL (-0.21). The dramatic increase in the correspondence in precipitation between Assim and observation strongly suggests that the effects of land data assimilation can transmit to the atmosphere through land-atmosphere interactions in the LCDA system, which may improve precipitation simulation. Improvements in the atmosphere states through land data assimilation highlight the important role of the land surface in drought development.

4 Conclusions

In this study, we developed the 4DEnVar-based LCDA system for the E3SMv2 model and evaluated the performance of the LCDA 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, as needed in the nudging method. Within each one-month assimilation window, we assimilate observed land information into the coupled model without breaking the land-atmosphere interaction, which is important for the LCDA system to be used to understand the potential sources of predictability provided by land.

The LCDA system is conducted from 1980 to 2016, 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 observational 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 North America, Northern Africa, Australia, and large parts of Eurasia. However, some degradations are observed in the

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

deep layers, which requires future research to better characterize observation errors in these deep zones. 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. The dramatic increase in the temporal correlations for precipitation anomaly in Assim also demonstrates that the impacts of land data assimilation could potentially contribute to the improvement in the atmospheric states through land-atmosphere interactions, highlighting the importance of the land surface in drought development.

Future improvements in the LCDA system will depend on the use of more observations and improving the quality of the ensemble covariance. It is possible that assimilation performance is restricted in specific domains due to biased atmospheric and oceanic forcing from the coupled model. Hence the continual integration of atmospheric and oceanic assimilations into the LCDA system could be an important way to further enhance its performance, particularly in regions where the land is primarily influenced by other components. Given the independence of the LCDA system from the coupled model, future exploration will focus on its implementation in other model components (e.g., atmosphere, ocean, and sea ice) or different climate models. To this end, the application of the LCDA 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.

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 can be downloaded from the website https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS%20monthly&page=1. The GPCP monthly precipitation data are available online (https://psl.noaa.gov/data/gridded/data.gpcp.html). 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.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Author contributions. LRL initiated this study. PS and LRL designed the experiments. BW provided advice on the data assimilation technique and KZ and SZ provided assistance with the E3SM model. PS developed the data assimilation code and performed the simulations. 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.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

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.

Bloom, S. C., Takacs, L. L., Da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, Monthly Weather Review, 124(6), 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2, 1996.

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.

Brocca, L., Crow, W. T., Ciabatta, L., Massari, C., De Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A. and Wagner, W.: A review of the applications of ASCAT soil moisture products, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5), 2285–2306, https://doi.org/10.1109/JSTARS.2017.2651140, 2017.

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 relatives 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.

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.

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.

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.

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. 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

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

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.

Hamill, T. A. and Snyder, C.: A hybrid ensemble Kalman filter-3D variational analysis scheme, Monthly Weather Review, 128, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000.

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.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

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.

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.

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.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Liu, C., Xiao, Q. and Wang, B.: An ensemble-based four-dimensional variational data assimilation scheme. Part I: technical formulation and preliminary test, Monthly Weather Review, 136, 3363–3373, https://doi.org/10.1175/2008MWR2312.1, 2008.

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.

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.

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.

Park, J. Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., and Zhang, S.: Modeling global ocean biogeochemistry with physical data assimilation: a pragmatic solution to the equatorial instability, Journal of Advances in modeling earth systems, 10(3), 891–906, https://doi.org/10.1002/2017MS001223, 2018.

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.

Prodhomme, C., Doblas-Reyes, F., Bellprat, O., and Dutra, E.: Impact of land-surface initialization on

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions

sub-seasonal to seasonal forecasts over Europe, Climate dynamics, 47, 919–935, https://doi.org/10.1007/s00382-015-2879-4, 2016.

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.

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.

Santanello, J. A., Sujay V. K., Christa D. P., and Patricia M. L.: Impact of soil moisture assimilation on land surface model spin-up and coupled land-atmosphere prediction, Journal of Hydrometeorology, 17, 517–540, https://doi.org/10.1175/JHM-D-15-0072.1, 2016.

Shaffrey, L. C., Hodson, D., Robson, J., Stevens, D. P., Hawkins, E., Polo, I., Stevens, I., Sutton, R. T., Lister, G., Iwi, A. and Smith, D.: Decadal predictions with the HiGEM high resolution global coupled climate model: description and basic evaluation, Climate Dynamics, 48, 297–311, https://doi.org/10.1007/s00382-016-3075-x, 2017.

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.

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(11), 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

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

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.

Timouk, F., Kergoat, L., Mougin, É., Lloyd, C. R., Ceschia, E., Cohard, J. M., De Rosnay, P., Hiernaux, P., Demarez, V. and Taylor, C.M.: Response of surface energy balance to water regime and vegetation development in a Sahelian landscape, Journal of Hydrometeorology, 375, 178–189, https://doi.org/10.1016/j.jhydrol.2009.04.022, 2009.

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.

Wang, X., Baker, D. M., Snyder, C. and Hamill, T. M.: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment, Monthly Weather Review, 136, 5116–5131, https://doi.org/10.1175/2008MWR2444.1, 2008.

Wei, M., Li, Q., Xin, X., Zhou, W., Han, Z., Luo, Y., and Zhao, Z.: Improved decadal climate prediction in the North Atlantic using EnOI-assimilated initial condition, Science Bulletin, 62(16), 1142–1147, https://doi.org/10.1016/j.scib.2017.08.012, 2017.

Wu, B., Zhou, T., and Zheng, F.: EnOI-IAU initialization scheme designed for decadal climate prediction system IAP-DecPreS, Journal of Advances in Modeling Earth Systems, 10(2), 342–356, https://doi.org/10.1002/2017MS001132, 2018.

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.

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, Y. F., Hoar, T. J., Yang, Z. L., Anderson, J. L., Toure, A. M., and Rodell, M.: Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4, Journal of Geophysical Research: Atmospheres, 119(12), 7091–7103, https://doi.org/10.1002/2013JD021329, 2014.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

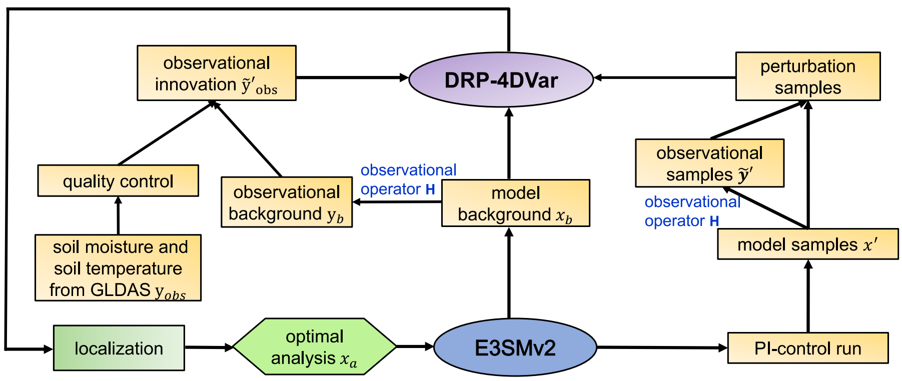


Figure 1. Flowchart of the 4DEnVar-based LCDA system in E3SMv2 based on the DRP-4DVar method.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

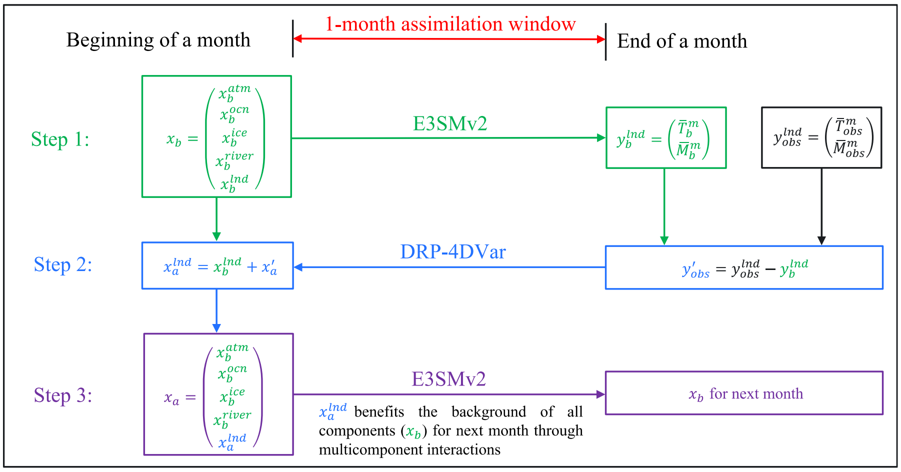


Figure 2. Schematic flowchart of the 4DEnVar-based LCDA system. The beginning of a month is at 0000 UTC on the first day of the month, and the end of the month is at 0000 UTC on the first day of the next month. x_b denotes the background vector including the backgrounds of all E3SMv2 components (atmosphere (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 ($\bar{T}_b^m$) and moisture ($\bar{M}_b^m$) by E3SMv2 using x_b as the initial condition. y_{obs}^{lnd} denotes the monthly mean GLDAS data of soil temperature ($\bar{T}_{obs}^m$) and moisture ($\bar{M}_{obs}^m$). y'_{obs} denotes the observational innovation, which is the difference between the GLDAS data (y_{obs}^{lnd}) and the observational background (y_b^{lnd}).

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access

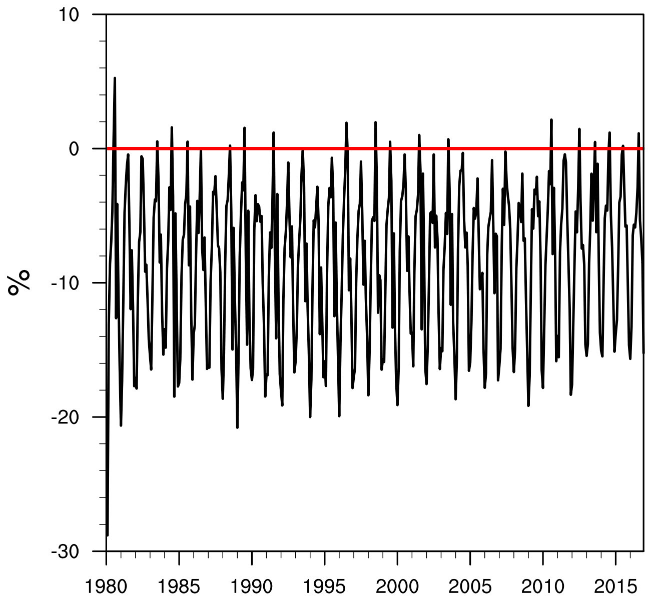


Figure 3. Time series of the reduction rate of the cost function from 1980 to 2016 in the 4DEnVar-based LCDA system.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

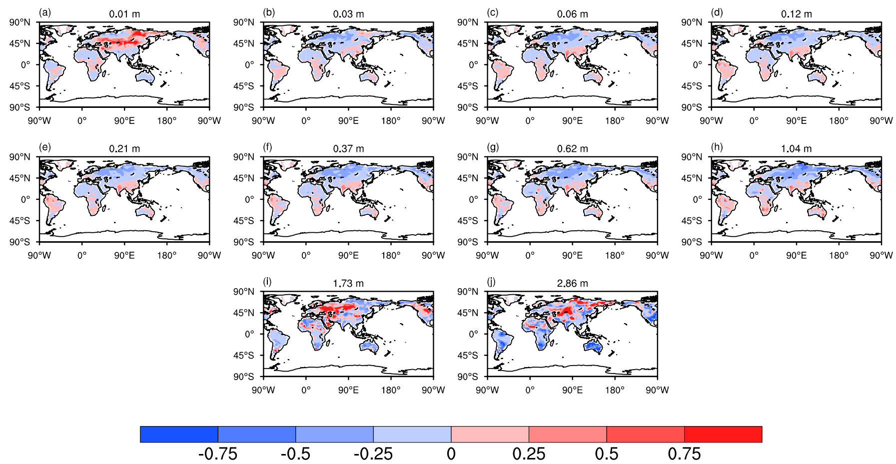


Figure 4. Spatial distribution of the AE index for soil moisture from the surface to deep layers during the 1980-2016 period. The number at the top center denotes the depth of each soil layer.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

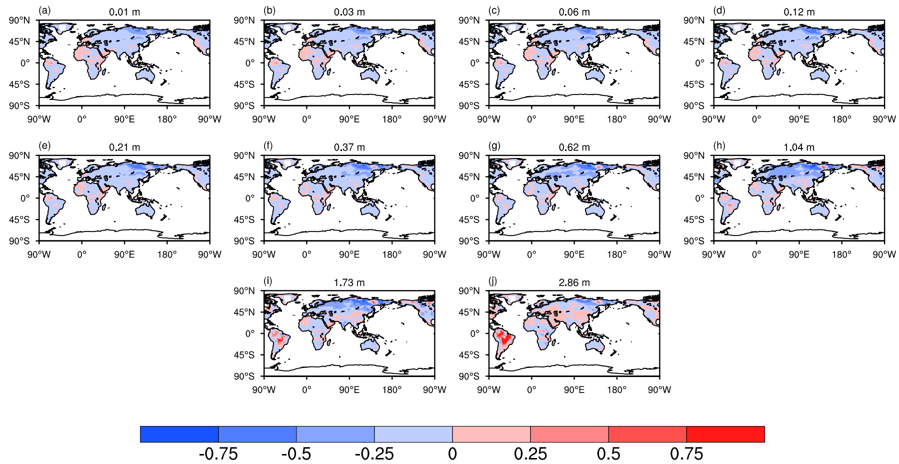


Figure 5. Same as in Figure 4, but for soil temperature.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access

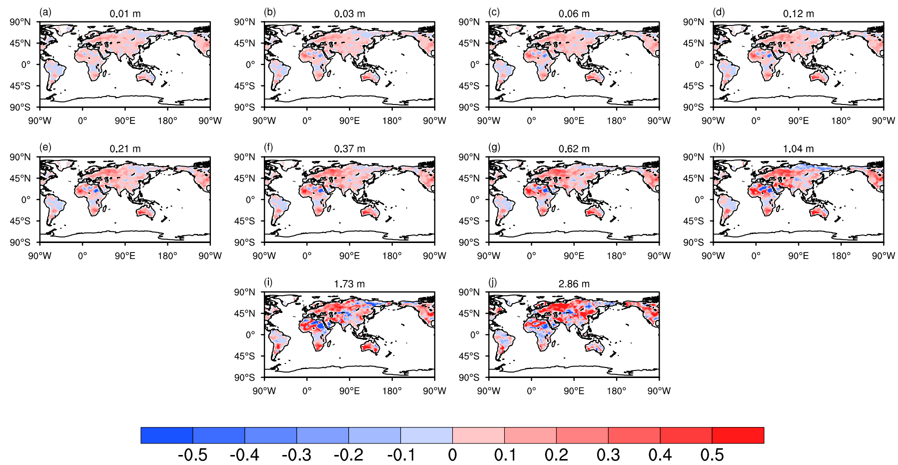
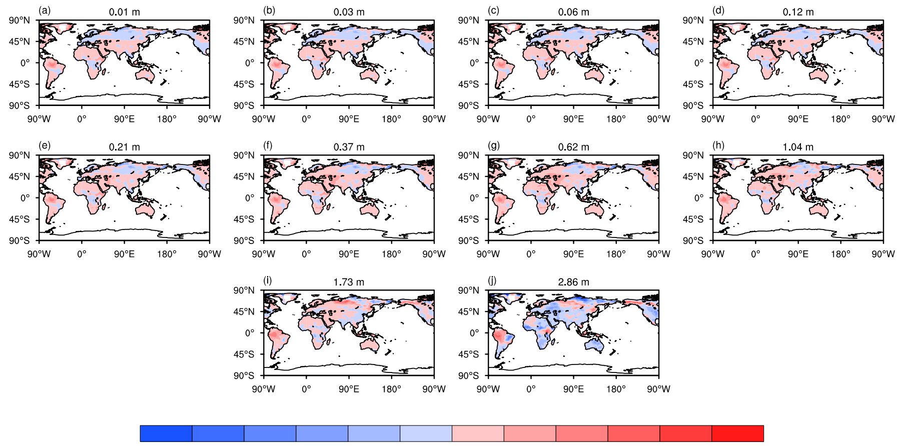


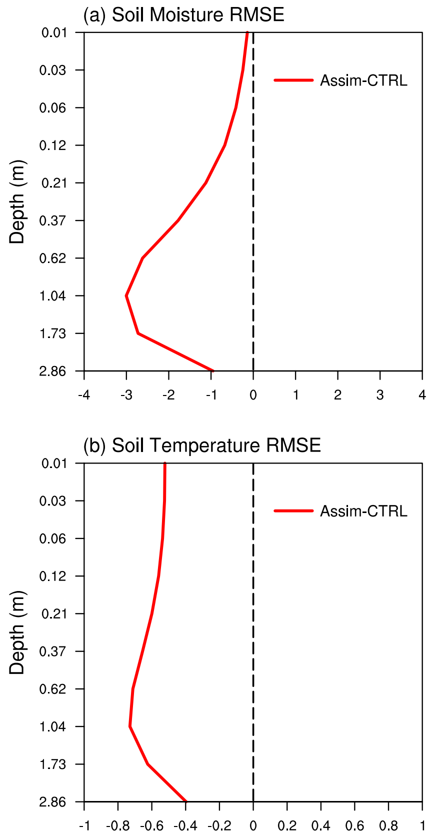
Figure 6. Differences between correlations of soil moisture in Assim and CTRL with the GLDAS data from the surface to deep layers for the period of 1980-2016. The number at the top center denotes the depth of each soil layer.



613

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

Geoscientific Model Development
Discussions
Open Access



615

616 **Figure 8.** Vertical distributions of RMSE differences (Assim minus CTRL) for (a) soil moisture and (b)

617 soil temperature averaged over the global land and throughout 1980-2016.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access

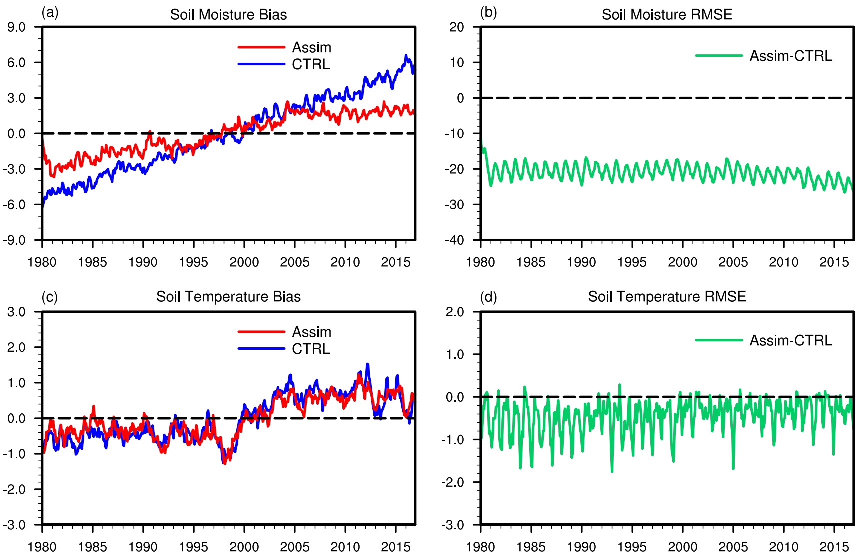


Figure 9. Time series of the vertically averaged global mean soil moisture and temperature bias (left) for Assim (red line) and CTRL (blue line), and RMSE differences (right, green line) between Assim and CTRL from 1980 to 2016.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

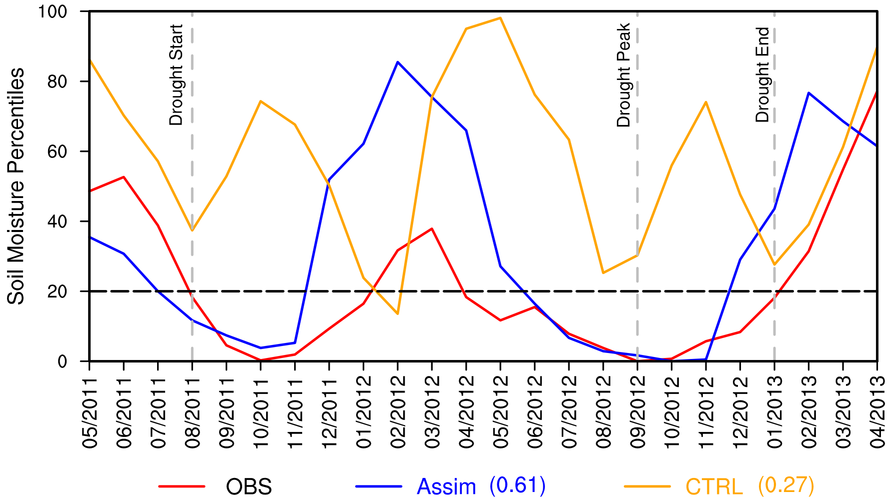


Figure 10. Time series of soil moisture percentiles between May 2011 and April 2013 during the 2012 U.S. Midwest drought. Red line: observation, blue line: Assim, orange line: CTRL. The correlation coefficients between Assim and CTRL with observations are also shown. The three vertical dashed lines mark 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.

https://doi.org/10.5194/gmd-2023-124
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access EGU

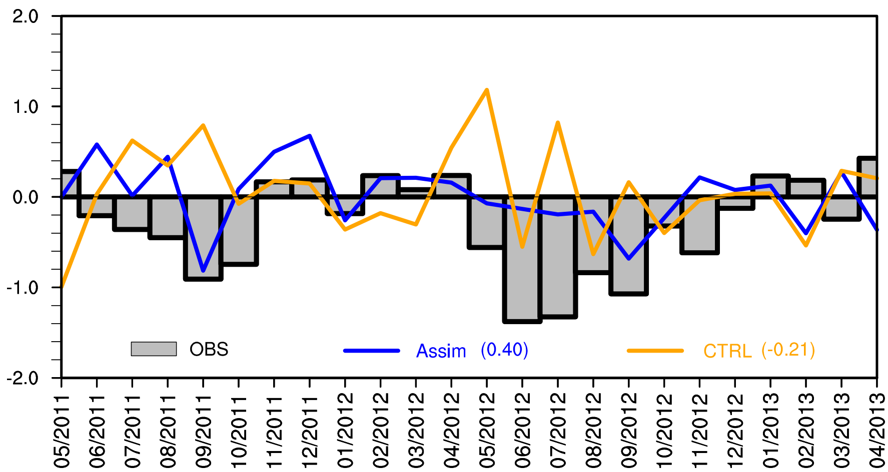


Figure 11. Time series of precipitation anomaly over the Midwest between May 2011 and April 2013 during the 2012 U.S. Midwest drought. Gray bar: observation, blue line: Assim, orange line: CTRL. The precipitation anomalies are calculated by removing the annual cycle and the long-term trend. The correlation coefficients of Assim and CTRL with observation are also shown. The precipitation anomalies are averaged over the U.S. Midwest (36°-50°N, 102°-88°W). The observed precipitation is derived from GPCP monthly precipitation data.