Assimilation of Remotely Sensed Leaf Area Index for Improving Land Surface Simulation Performance at a Global Scale

Xiaolu Ling, Jian Gao, Zeyu Tang, and Wenhao Liu

***Abstract*—The Community Land Model version 4 with carbon and nitrogen components is coupled with data assimilation research testbed to assimilate remotely sensed leaf area index (LAI), to analyze the improvement in model performance for simulating land surface variables and land–atmospheric exchange fluxes. The results demonstrate that assimilation effectively addresses the issue of significant overestimation of LAI values, particularly noticeable in regions characterized by low latitudes and dense vegetation coverage. At a global scale, the disparities between simulated and assimilated LAI relative to observational data, are measured at 0.90 and −0.07, representing 54.1% and 3.9% of the observed values, respectively. The root mean square difference (RMSD) for assimilated LAI is 1.61 comparing with the simulated LAI of 1.85. Assimilating LAI globally leads to a noteworthy 1% reduction in the mean relative difference of the global average 2-m air temperature (T_{2m}) and a concurrent decrease of 0.15 °C in RMSD. However, at the global level, the assimilation of LAI does not yield a significant enhancement in the modeling capability of heat fluxes, although modeling capability of sensible heat (HS) slightly outperforms latent heat. Improvements in land surface variables after assimilation show significant variations at regional scales due to factors such as vegetation coverage and climatic conditions. Overall, in regions characterized by periodic changes in vegetation, such as forested areas in Western Eurasian Continent (region 5), the enhancements in T_{2m} and HS after assimilating LAI are particularly notable, with mean relative difference reduced by 7% and 20%, respectively.**

***Index Terms*—2-m air temperature, land data assimilation (DA), leaf area index (LAI), surface fluxes.**

I. Introduction

VEGETATION, as a crucial component of land surface variables, plays a significant role in the water, energy, and carbon cycles within the Earth's ecosystem [1], [2]. The leaf area index (LAI), which is defined as the total one-sided area of all the leaves within a canopy in a given region [3], influences the water, mass, and energy exchanges. It does so by participating in biophysical, hydrological, and biogeochemical processes [4], a phenomenon observed in both natural environments and simulation models.

LAI can be quantified through various methods, including in situ observation, satellite-retrieved products, and model simulations. In situ observation is often considered the benchmark, serving as a reference for estimating the accuracy of remotely sensed or simulated values. However, expanding in situ observations to regional scales presents significant challenges. Over the past decades, a variety of satellite-based LAI products have been developed, each offering different temporal and spatial resolutions. Notable examples include the moderate resolution imaging spectroradiometer (MODIS) LAI product [5], the global land surface satellite (GLASS) LAI product [6], and the satellite pour l'observation de la terre (SPOT) LAI dataset [7], among others [8]. These satellite-derived LAI products typically offer spatial resolutions ranging from 500 m to 1° and temporal resolutions varying from eight days to one month. Model simulations can provide LAI data at finer tempo-spatial resolutions [9], yet the accuracy of simulated LAI is contingent upon the specific module settings and parameterization of land surface physical and biogeochemical processes.

LAI was originally prescribed as an input parameter or a boundary condition in earlier generations of land surface models (LSMs), where the inputted LAI values exhibited only seasonal variability without annual variation [10], [11]. However, with advancements in modeling technology, contemporary LSMs and ecological models increasingly incorporate dynamic LAI variations, or even predict LAI as an output variable [9], [12]. Despite these advancements, significant uncertainties remain, particularly in terms of amplitude and phase [13]. Moreover, the accuracy of LAI is critical, as its uncertainties can substantially impact the model's effectiveness in simulating various land surface variables.

Previous studies have demonstrated that the integration of satellite-based products can enhance the simulation accuracy of related land surface variables at both global and regional scales [14], [15], [16], [17], [18], [19]. For instance, replacing default LAI values with field observations or remote sensing products has been shown to improve simulations of momentum and trace

Manuscript received 30 September 2023; revised 31 January 2024 and 7 April 2024; accepted 8 April 2024. Date of publication 12 April 2024; date of current version 1 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42075114 and Grant 41705101, in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions under Grant 140119001, and in part by the General Project of Modern Agriculture from the Primary R&D Program of Xuzhou under Grant KC21132. *(Corresponding author: Xiaolu Ling.)*

Xiaolu Ling is with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China., and also with the Jiangsu Key Laboratory of Coal-based Greenhouse Gas Control and Utilization, China University of Mining and Technology, Xuzhou 221008, China. (e-mail: lingxl@cumt.edu.cn).

Jian Gao, Zeyu Tang, and Wenhao Liu are with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China. (e-mail: ts21160122p31@cumt.edu.cn; ts22160044a31@cumt.edu.cn; tb21160015b1@cumt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3388006

gas exchanges [20], [21], [22], [23]. Additionally, assimilating other variables, such as brightness temperature, radar backscatter, and normalized difference vegetation index (NDVI) for grass and trees, has also been effective in refining LAI [24], [25]. Furthermore, the assimilation of canopy parameters, including LAI, and/or soil moisture (SM), has led to improved model performance in several areas. These areas include the simulation of the carbon cycle [26], [27], [28], [29], evapotranspiration (ET) [13], [30], hydrological processes [31], [32], [33], agriculture dynamics [34], [35], [36], ecosystem functions [24], [37], and seasonal temperature predictions [38].

Internationally recognized land surface data assimilation (DA) systems include the North American Land Data Assimilation System (NLDAS [39], NLDAS-2 [40], [41]), the Global Land Data Assimilation System (GLDAS [42]), the European Land Data Assimilation System (ELDAS [43]), the Western China Land Data Assimilation System (WCLDAS [44]), and the Canadian Land Data Assimilation System (CaLDAS [45]).

Variational and sequential algorithms represent two fundamental mathematical approaches for assimilating diverse variables [46]. The most prevalent variational methods for DA are the 3DVAR and 4DVAR methods [47], [48]. 3DVAR is characterized by its simplicity and low computational cost, capable of handling nonlinear observation operators and achieving globally optimal analysis values. However, a limitation of 3DVAR is isotropic and unchanging background covariance. On the other hand, 4DVAR addresses this issue to certain degree, but it demands significantly higher computational resources for both implementation and maintenance [49], [50]. Sequential algorithms, which are based on the Kalman Filter [51], have undergone extensive development. This includes the ensemble Kalman filter (EnKF) [52], extended Kalman filter (EKF) [53], ensemble adjusted Kalman filter (EAKF) [54], maximum likelihood ensemble filter (MLEF) [55], particle filter (PF), and others. In recent decades, there has been an innovative trend of combining variational and sequential algorithms. This has led to the emergence of hybrid methods, such as the four-dimensional (4-D) EnKF (4-DEnKF) [56] and the 4-D local ensemble transform Kalman filter [57], among others.

The data assimilation research testbed (DART), developed by the National Center for Atmospheric Research (NCAR) [58], has been successfully coupled with the Community Earth System Model (CESM) [59] and its various components. These components include the community atmosphere model (CAM) [60], community land model (CLM) [61], [62], and ocean [63]. This integration, specifically the coupled DART/CLM system, has demonstrated that the assimilation of remote sensing datasets can significantly enhance model performance. Notable improvements have been observed in the simulations of SM [64], snow water equivalent [62], gross primary productivity (GPP), and ET [13].

In previous study, the authors have adeptly assimilated remotely sensed LAI using the coupled DART/CLM4 with carbon and nitrogen components (CLM4CN) system, leading to significant improvements in the model's performance for simulating global GPP and ET. Previous research in this area did not extensively explore the impact of LAI assimilation on the accuracy of land state variables and land-atmosphere fluxes, and comparisons were generally limited to global scales. To address this gap, our research utilizes the GLDAS reanalysis data to validate the modeled land surface variables and land–atmospheric fluxes, both with and without LAI assimilation. Furthermore, this article aims to elucidate the underlying physical mechanisms between land and atmosphere influenced by LAI, particularly in the context of energy and water balances.

TABLE I
EXPERIMENT DESIGN

Name	Ensemble No.	Assimilated variables	Updated variables	Spatial resolution	Simulation Periods
CTL	40	No	No	0.9° × 1.25°	2002
CN	40	GLASS LAI	LAI, Leaf C, Leaf N	0.9° × 1.25°	2002

II. DATA AND METHODOLOGY

Based on the DART and its EAKF algorithm, the assimilation of globally remotely sensed LAI products into the CLM4CN was undertaken to investigate the significance of LAI in enhancing the performance of land surface simulations at a global scale. The experimental design is delineated in Section II-A, whereas Section II-B elaborates on CLM4CN, encompassing ensemble forcing, ensemble initialization, as well as model output and restart files, along with DART, which encompasses the EAKF DA methodology. The datasets for assimilation and evaluation are expounded upon in Section II-C. Validation and evaluation criteria are explicated in Section II-D.

A. Experimental Design

In this article, two experiments were carried out: the assimilation experiment (referred to as the CN Experiment) and the control simulation experiment (referred to as the CTL Experiment), as given in Table I. The CTL Experiment was conducted by using the CLM4CN model independently. The CN Experiment was conducted by using the coupling DART/CLM4CN system to assimilate LAI values.

Both the CTL and CN Experiments were conducted with a spatial resolution of 0.9° latitude by 1.25° longitude and a temporal resolution of 6 h, utilizing 40 ensemble members. In the CN Experiment, the GLASS LAI data were assimilated into the CLM4CN model at eight-day intervals, providing regular updates to the model based on observed vegetation characteristics through remote sensing. Taking into account factors such as computational resources, we focused our experiments and analysis on the year 2002.

Fig. 1 delineates the integrated coupling workflow of the DART/CLM4CN framework. In scenarios where the CLM4CN model operates independently of an atmospheric model, it necessitates an ensemble of meteorological datasets for driving simulations, in conjunction with a set of ensemble initial conditions. Throughout the simulation phase of CLM4CN, ensemble-based output and restart files are systematically generated in accordance with predefined user configurations. These restart

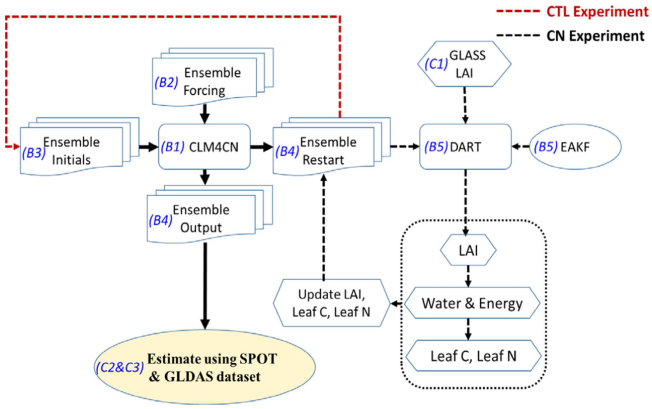

Fig. 1. Coupling workflow of the DART/CLM4CN framework. The subindices in the figure correspond to the subsections described within Section II.

files are then utilized as initial inputs for subsequent temporal iterations, a process depicted by the red dashed lines in Fig. 1.

In the framework of the CN experiment, the model intricately transmits ensemble-based restart file data corresponding to each discrete time step to the DART. Within this process, DART utilizes the sophisticated EAKF algorithm to effectively assimilate the GLASS LAI data with the ensemble LAI extracted from the restart files. Following this assimilation, the model's inherent physical equations are applied to update the LAI, leaf carbon (Leaf C), and leaf nitrogen (Leaf N) values. These updated variables are then methodically reintegrated into the restart files, thereby establishing the initial conditions for the subsequent time step.

The coupling of DART with the CLM4CN has been effectively achieved, as detailed in [13]. For further insights into this coupling process, readers are directed to this reference. During the assimilation phase, GLASS LAI data is assimilated into the CLM4CN model employing the EAKF algorithm [54], [65]. This assimilation process simultaneously updates LAI, Leaf C, and Leaf N.

It has been observed that the assimilation results in an improved representation of LAI, particularly in low-latitude regions. This enhanced LAI subsequently leads to more accurate simulations of GPP and ET [13]. This outcome underlines the effectiveness of incorporating remotely sensed LAI into the CLM4CN model using the EAKF algorithm, thereby enhancing the model's ability to simulate key ecological processes.

B. Methodology

1) CLM4CN and its Calculation Process for LAI: The CLM4CN is an extension of the CLM4 model, integrating carbon and nitrogen cycles to study the interactions in terrestrial ecosystems and their responses to climate change.

The model simulates vegetation growth, death, competition, and succession processes based on climate conditions and soil characteristics. Plant growth is constrained by factors, such as photosynthesis, temperature, moisture, and nutrients. The model calculates vegetation carbon balance through photosynthesis and respiration processes. Soil carbon calculations include organic matter decomposition and biogeochemical transformations of carbon. The model simulates nitrogen inputs (such as atmospheric deposition and biological nitrogen fixation), storage, transformations, and losses. The nitrogen cycle is closely linked to the carbon cycle, affecting plant growth and soil organic matter decomposition. CLM4CN also simulates water and energy balance, including processes, such as ET, SM movement, precipitation, and energy exchanges.

In the CLM4CN model, LAI is a prognostic variable and is determined by combining the specific leaf area index (SLA) at the canopy's top with leaf carbon. This leaf carbon is in turn calculated based on the leaf carbon-to-nitrogen ratio, which varies across different plant functional types (PFTs). Additionally, the model calculates radiation by incorporating climatological monthly SM and the vegetation parameters, which include the fraction of PFTs, LAI, and stem area index.

Essentially, LAI has a significant influence on the GPP as it affects the canopy radiation environment. This influence is a function of the canopy—level pool of leaf carbon and nitrogen. It is noteworthy that the availability of leaf carbon, which is critical for LAI determination, is dependent on the GPP itself.

2) Ensemble Forcing: CLM4CN requires time-series meteorological data to drive the model, typically including temperature, precipitation, wind speed, humidity, solar radiation, etc.

An 80-member reanalysis was conducted using DART in conjunction with the CAM Version 4 with the finite volume core (DART/CAM4 FV [60]). This reanalysis was performed at a resolution of 1.9° by 2.5° and is accessible via the UCAR Research Data Archive (https://rda.ucar.edu/datasets/ds199-1/). For this reanalysis, 40 ensemble meteorological fields were randomly selected to serve as the driving forces for the CLM4CN model. This selection was made with careful consideration of both computational feasibility and the performance of EAKF.

3) Ensemble Initial Conditions: Before the simulation starts, it's necessary to initialize the state of soil, vegetation, and atmosphere. This includes soil temperature, moisture, vegetation type, cover, soil carbon and nitrogen content, etc. In establishing the initial steady-state condition for the ensemble in this article, the researchers employed a three-step process, each crucial for ensuring the model's reliability and accuracy.

1) *Long-duration Single Model Run Using Qian's Forcing:* The first step involved driving a single instance of the CLM4CN model with Qian's atmospheric forcing data [66] for an extensive period of 4000 years [67]. This long duration is essential for the model to reach a steady-state condition, where it adequately simulates the Earth's climate system over a prolonged period. This step ensures that the model's output is not overly influenced by initial conditions or short-term variability.
2) *Ensemble Mean Forcing for 1000 Years:* The second step used the ensemble mean of 40 atmospheric forcing from the DART/CAM4 FV dataset [60] to drive the CLM4CN model for 1000 years. This step helps the model adjust to a variety of atmospheric conditions, averaged over the 40 different scenarios, providing a more generalized and stable model state that is still reflective of diverse climatic conditions.

3) *Specific Ensemble Forcing From 1998 to 2001:* In the final step, the researchers used the 40 individual ensemble atmospheric forcing from the DART/CAM4 FV dataset [60] to drive the ensemble CLM4CN from 1998 to 2001 [13]. This phase is crucial for aligning the model with recent climatic conditions, ensuring that the study's outcomes are relevant to current and near-future climate scenarios.

4) Model Output and Restart Files: The model outputs include vegetation cover, biomass, soil carbon and nitrogen content, gas exchanges (such as CO_2 and N_2O emissions), and water and energy fluxes. These outputs are used to analyze the ecosystem's response to environmental changes. In practical operations, the CLM4CN also generates restart files, including output data from the model simulations and intermediate data which represents the state of various model components at the end of a simulation period. Moreover, the restart files serve as the initial condition for the next run of the model, which is crucial for multiyear or multidecadal simulations where the model is run in segments over time. By using the restart file from the end of one simulation as the starting point for the next, the model ensures continuity in the simulation process, accurately reflecting the progression of ecological and environmental changes over time.

5) DART and EAKF DA Method: DART is an open-source platform created to aid in DA research, development, and education across various fields such as atmospheric science, oceanography, and hydrology. Initiated in 2001 and formalized under the data assimilation research section at NSF NCAR in 2003, DART has been providing sophisticated ensemble DA algorithms and diagnostic tools for over two decades. The first officially supported version of DART was released in 2004, with a consistent version control history available from 2005. This marks DART as a long-standing, well-supported software project in the scientific community. Over the years, it has seen more than a dozen releases, with each major release uniquely named after islands in alphabetical order, starting with the first release named Easter. The current latest version of DART is Manhattan, while the version used in this article is Lanai.

DART provides a range of filter algorithms to update the posterior distribution using observations and the prior ensemble. These include EnKF, EAKF, Kernel Filter, PF, among others. Each of these filters has its strengths and is suitable for different types of DA problems. The choice of filter algorithm in DART depends on the specific requirements of the model system, including its linearity, size, and the nature of the observations.

In this article, the algorithm of EAKF is utilized. EAKF is a variant of the EnKF and is widely used in DA for fields like meteorology, oceanography, and hydrology. According to Ling et al. [65], EAKF is the optimal algorithm for LAI assimilation using DART/CLM4CN. As an ensemble-based DA method, it uses a collection of model states (ensemble) to estimate the state of a physical system and update this estimate with new observational data.

The equations of EAKF are as follows.

1) *Forecast Step:* This is the first step in the EAKF algorithm, where the current state of the model is used to generate predictions for future states. This step involves the dynamic equations of the model. For each ensemble member *i*, the forecast update of the state vector X_i^f is given by

$$X_i^f = M\left(X_i^{a,t-1}\right) \tag{1}$$

where X_i^f is the forecast state, M is the dynamic equation of the model, and $X_i^{a,t-1}$ is the analysis state from the previous time step.

2) *Analysis Step:* In this step, observations are used to update the state of each ensemble member. First, calculate the ensemble mean and covariance matrix in observation space, then use the Kalman gain matrix to update the state of each ensemble member.

a) Calculate the ensemble mean Y^f and the covariance matrix P_{xy} in observation space

$$Y^f = \frac{1}{N}\sum_{i=1}^{N} H\left(X_i^f\right) \tag{2}$$

$$P_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i^f - \bar{X}^f\right)\left(H\left(X_i^f\right) - Y^f\right)^T \tag{3}$$

where *N* is the number of ensemble members (*N* is 40 in this article), H is the observation operator mapping from model space to observation space.

b) Calculate the Kalman gain matrix

$$\mathrm{K} = P_{xy}\left(HP_{xy}^T + R\right)^{-1} \tag{4}$$

where R is the observation error covariance matrix.

c) Update the state of each ensemble member

$$X_i^a = X_i^f + K\left(Y_{\text{obs}} - H\left(X_i^f\right)\right) \tag{5}$$

where Y_{obs} is the actual observed value, and X_i^a is the analysis (updated) state.

C. Dataset

1) GLASS LAI Products: GLASS LAI data, known for its reasonably representation of global vegetation characteristics [6], are assimilated into the DART/CLM4CN framework. To align with the resolution of the CLM4CN simulation, the original GLASS LAI data, which have a spatial resolution of 0.05°, are spatially averaged to match the CLM4CN model's resolution. The assimilation frequency is consistent with the satellite data, i.e., GLASS LAI is assimilated every eight days.

2) SPOT LAI Products: SPOT series of satellites are a family of Earth-observing satellites operated by Centre National d'Études Spatiales, the French space agency, with the participation of Belgium and Sweden. SPOT stands for "Satellite for Earth Observation" in French. These satellites have been designed to capture high-resolution, optical imaging data of the Earth's surface, making them extremely valuable for a wide range of applications, including mapping, land use planning, agriculture, forestry, and environmental monitoring.

The SPOT satellites provided ten-daily observations at global scale in the spatial resolution of 1/112° (approximately equivalent to 1 km) over a temporal range from 1999 to 2020. The data

is displayed in a regular latitude/longitude grid using the WGS 1984 projection. Detailed descriptions of the algorithm can be found in the works of [7], [68]. Additionally, the accuracy of the LAI products in capturing the effects of environmental events on vegetation, such as hail storms or large fires, is annually verified, as discussed in [69].

3) Global Land Data Assimilation System: The GLDAS dataset, collaboratively developed by National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA), integrates surface observations, satellite remote sensing data, and terrestrial model outputs to produce comprehensive global reanalysis data. This data encompasses various surface state variables, such as SM and soil temperature, as well as fluxes like ET and sensible heat fluxes (HS) [42].

There are two versions of GLDAS available: GLDAS-1 and GLDAS-2. GLDAS-1 has been providing data from 1979 to the present and includes four different model simulations/assimilation frameworks: CLM, NOAH, MOS, and VIC. These models offer data with a spatial resolution of $1^\circ \times 1^\circ$ and temporal resolutions ranging from 3 h to several months. GLDAS-2 provides data at a finer spatial resolution compared to GLDAS-1, allowing for more detailed and region-specific analysis of land surface conditions. While GLDAS-1 laid the foundation with its global coverage and range of land surface variables, GLDAS-2 advances this with improved model accuracy, higher resolution data, and extended temporal coverage, making it more suitable for in-depth climate research and environmental studies.

One of the key strengths of GLDAS is its comprehensive coverage of data types. It offers a wide array of data, including atmospheric assimilation products or reanalysis information from sources such as Global Earth Observation System (GEOS), Global Data Assimilation System (GDAS), and European Centre for Medium-Range Weather Forecasts (ECMWF), along with direct observations such as driving fields. GLDAS products have gained widespread use among researchers for validation purposes, particularly against satellite remote sensing products. Additionally, these products have been extensively applied in studies related to global and regional climate change [70], [71], [72], [73].

D. Validation and Evaluation

Bias, relative bias (rBias), root mean square deviation (RMSD), correlation coefficients were utilized to validate the model performance with and without assimilation. Bias is a statistical measure that quantifies the systematic difference between the values predicted by simulation/assimilation and the observed values. It indicates the tendency of the model to consistently overestimate or underestimate the observed values.

The formula for calculating bias is

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{\text{model},\,i} - X_{\text{obs},\,i}\right) \tag{6}$$

where n is the total number, $X_{\text{model},\,i}$ represents the model's output from Experiments CTL and CN, $X_{\text{obs},\,i}$ represents the observed data. For LAI, $X_{\text{obs},\,i}$ refers to the SPOT LAI products, and for the other land surface variables, $X_{\text{obs},\,i}$ refers to the GLDAS dataset.

Besides indicates whether the simulation/assimilation tends to overestimate or underestimate the observed values, rBias provides insight into the magnitude of this bias relative to the observed values

$$\text{rBias} = \frac{\sum_{i=1}^{n}\left(X_{\text{model},\,i} - X_{\text{obs},\,i}\right)}{\sum_{i=1}^{n} X_{\text{obs},\,i}} \times 100\% \tag{7}$$

RMSD is a statistical measure used to quantify the average magnitude of the differences between the values predicted by a model or estimator and the observed values. It assesses the overall accuracy of the model's predictions. Lower RMSD values indicate better agreement between the model predictions and the observed values

$$\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{\text{model},\,i} - X_{\text{obs},\,i}\right)^2} \tag{8}$$

Correlation coefficients are statistical measures used to quantify the strength and direction of the linear relationship between two variables. They indicate how closely the two variables are related

$$\text{r} = \frac{\sum_{i=1}^{n}\left(X_{\text{model},\,i} - \overline{X_{\text{model}}}\right)\left(X_{\text{obs},\,i} - \overline{X_{\text{obs}}}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{\text{model},\,i} - \overline{X_{\text{model}}}\right)^2 \sum_{i=1}^{n}\left(X_{\text{obs},\,i} - \overline{X_{\text{obs}}}\right)^2}} \tag{9}$$

By employing the aforementioned criteria, bias, rBias and RMSD for LAI and associated land surface variables were computed for the entirety of 2002. Additionally, correlation coefficients were determined between the latitudinal averages of LAI and land surface variables in July and November 2002.

III. Results

As a physical characteristic parameter of vegetation, LAI had a nonnegligible influence on land surface variables and land–atmospheric fluxes. In this section, we initially examine the spatial distribution characteristics during the summer season in the Northern Hemisphere, followed by an analysis of the seasonal variations in different subregions. The SOPT LAI and GLDAS data are employed for the validation of LAI and land surface variables, as well as land-atmospheric fluxes.

The objectives of this section are twofold: assessing the impact of LAI assimilation on model performance; and analyzing the role of vegetation in weather and climate dynamics.

A. Leaf Area Index

To illustrate the impact of assimilating LAI into CLM4CN, Fig. 2 presents the spatial distribution of LAI for: observation; CTL—observation; CN—observation over improved regions; and CN—CTL over nonimproved regions in July 2002.

In July 2002, a detailed observation revealed two distinct latitudinal belts with high LAI values, primarily concentrated in the tropics and the boreal zones between at 50 °N–65°N. These regions were predominantly covered by evergreen broadleaf and

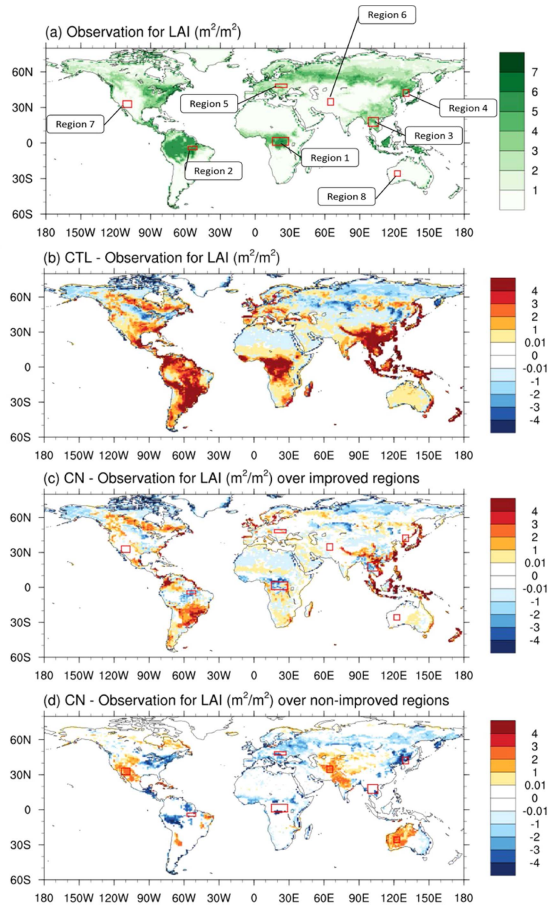

Fig. 2. Spatial distribution of LAI in July 2002. (a) Observation. (b) CTL—observation, (c) CN—observation over improved regions, and (d) CN—CTL over non-improved regions. The red boxes indicate selected typical regions, with detailed information provided in Table II. An "improved region" is defined as an area where the absolute difference between the CN Experiment and GLDAS data is smaller than the absolute difference between the CTL Experiment and GLDAS data.

TABLE II
MEAN DIFFERENCE OF LAI FOR CN—OBS AND CTL—OBS OVER TYPICAL REGIONS

Regions	Distribution	Longitude/ Latitude	CN − OBS	CTL − OBS	Dominant Land Cover Types
Global	Global	180°W–180°E, 90°S–90°N	−0.07	0.90	All
Region 1	Central Africa	15°E–28.75°E, 1.88°S–4.69°N	−1.98	7.52	Evergreen broadleaf forests
Region 2	Eastern Amazonia	57.5°W–50°W, 5.63°S–2.81°S	−2.56	4.13	Evergreen broadleaf forests
Region 3	Southern Eurasia	97.5°E–106.25°E, 14.06°N–21.56°N	−0.48	6.79	Evergreen broadleaf forests, cropland
Region 4	Northeastern China	127.5°E–132.5°E, 39.38°N–45°N	−3.17	2.67	Deciduous broadleaf forests, mixed forests
Region 5	Western Europe	17.5°E–27.5°E, 46.88°N–49.69°N	−1.48	2.27	Mixed forests, cropland
Region 6	Central Eurasia	62.5°E–67.5°E, 31.88°N–37.5°N	2.42	0.06	Cropland, open shrublands
Region 7	Southwestern North America	113.75°W–106.25°W, 30°N–35.63°N	2.68	0.52	Grassland, open shrublands
Region 8	Western Australia	120°E–125°E, 28.13°S–23.44°S	1.94	0.26	Open shrublands

boreal forests, respectively, contributing significantly to their LAI measurements.

When examining the CLM4CN simulations' ability to replicate these global LAI distribution patterns, a notable overestimation was observed, particularly in the low latitude tropical regions and the northern boreal forests. This discrepancy in the simulations reached as high as 5 m^2/m^2, underscoring a need for refinement in the model to more accurately capture the true variation of LAI across these diverse ecological landscapes. In sparsely vegetated areas, such as Northern Africa and Northern Eurasia, which are characterized by deserts and perennial snow and ice cover, the model underestimates the LAI.

DA techniques have shown significant potential in addressing the overestimation issues observed in the LAI estimations. Most notably, the DA-adjusted LAI values exhibit substantial reductions in several key regions, as depicted in Fig. 2(a). These areas include central Africa, eastern Amazonia, southern Eurasia, northeastern China, and western Europe, marked as Regions 1 to 5. In these regions, DA has effectively corrected the previously observed overestimations, bringing the LAI estimates closer to more realistic values. Conversely, in central Eurasia, southwestern North America, and western Australia (region 6 to 8), the DA technique has led to a significant increase in the estimated LAI values. This adjustment suggests an initial underestimation of LAI in these areas by the model before the application of DA.

Table II gives typical regions based on their vegetation characteristics. After assimilation, regions 1 to 5 (with the exception of region 4) exhibit notable improvements in LAI accuracy, showing less deviation from expected values. These regions are characterized by rich vegetation cover type, suggesting that the assimilation process is particularly effective. On the other hand, the assimilated LAI values in regions 6 to 8, which predominantly consist of croplands, open shrublands, and grassland, do not show significant improvement. This indicates that the current assimilation techniques may have limitations or less efficacy in areas dominated by these types of vegetation.

On a global scale, the discrepancies between simulated and assimilated LAI relative to observations are 0.90 and −0.07, constituting 54.1% and 3.9% of the observed values, respectively. Regionally, the most substantial improvement is observed in Region 3, with the relative deviation decreasing from 214.8% to 15.1%. Following this, regions 1 and 2 exhibit reductions in relative deviations from 76.2% and 78.0% to 47.1% and 50.8%, respectively.

Fig. 3 illustrates the temporal evolution of RMSD between observed LAI and LAI simulated in the CN and CTL experiments throughout 2002. In regions 1 to 5, CN consistently outperforms

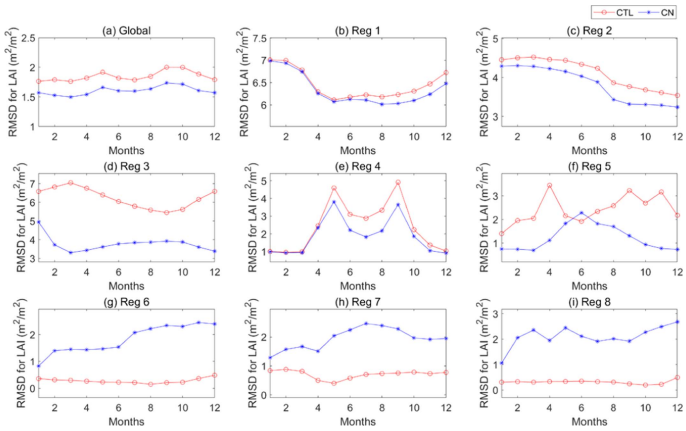

Fig. 3. Evolution of RMSD of the CN/CTL experiment with observed LAI over time in 2002.

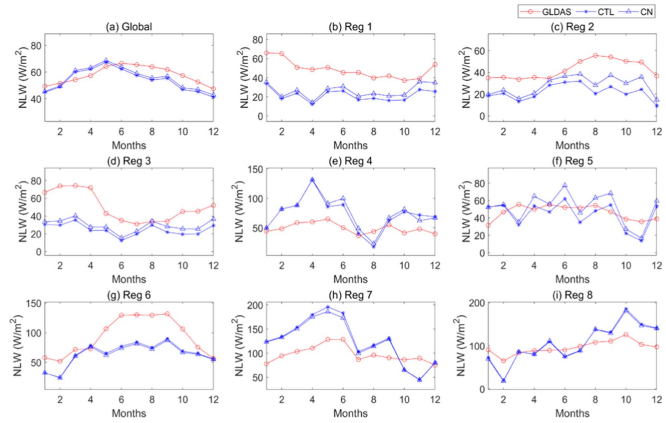

Fig. 5. Evolution of regional averaged NLW over time at global and regional scales in 2002.

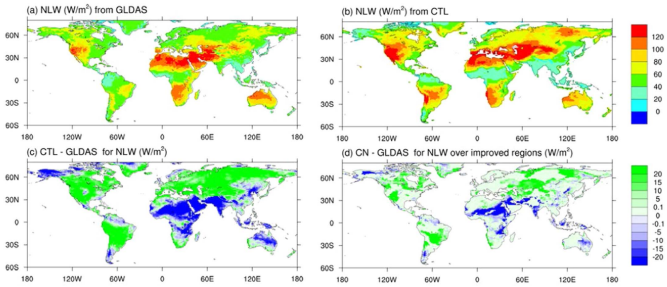

Fig. 4. Spatial distribution of NLW in July 2002 for (a) GLDAS NLW, (b) the CTL experiment, (c) CTL—GLDAS, and (d) CN—GLDAS over improved regions.

CTL, indicating improved model alignment with observed LAI in densely vegetated areas. Conversely, in regions 6 to 8 characterized by croplands and open shrublands, and grasslands, CN's performance is inferior, suggesting LAI assimilation does not enhance model accuracy in these regions.

Globally, CN's annual mean RMSD for LAI is 1.61 compared to CTL's 1.85, suggesting CN's more accurate global estimates. The seasonal variations in RMSD for both CN and CTL experiments are consistent, reaching their peak values in May and September. This implies that both the model and assimilation exhibit weaker performance in reproducing LAI during these two months. Seasonal RMSD patterns vary across regions. In regions characterized by low latitudes and dense vegetation cover (reg1 and 2), both the simulation and assimilation demonstrate analogous seasonal variations in LAI. Notably, the RMSD for LAI exhibits a discernible trend, manifesting higher values at the initial time, followed by a gradual decline over the course of the season.

Significant differences in response are observed over mid-latitude regions with dense vegetation. In Reg3, CN and CTL show opposite trends, with higher simulated LAI RMSD corresponding to reduced assimilated LAI RMSD. A similar pattern is observed in Reg5, suggesting that assimilation tends to yield smaller LAI RMSD when the simulated LAI RMSD is higher. The annual average of LAI RMSD is 6.23 m^2/m^2 in reg 3, which is reduced to 3.77 m^2/m^2 after LAI assimilation. Similar as global average, in Region 4, both CN and CTL show similar trends in LAI RMSD variations, with a larger amplitude. During the northern hemisphere growing season, the simulated LAI RMSD is 3.55, while the assimilated LAI RMSD is 2.67. However, this enhancement is less evident during nongrowing seasons.

B. Net Longwave Radiation (NLW)

Fig. 4 illustrates the spatial distribution of NLW for July 2002, using data from GLDAS, CTL experiment, and differences between CTL and GLDAS, as well as CN and GLDAS in areas of improvement. The CTL experiment can capture the global distribution characteristics of NLW. However, a global overestimation is observed, except for the regions of India, northern and central Africa, and northern Australia.

Fig. 4(d) displays the difference between the CN experiment and GLDAS in areas where NLW has improved after assimilation. The NLW difference between CN experiment and GLDAS indicates a negative discrepancy in low-latitude areas and a positive one in middle- and high-latitude regions, primarily covered by savannas and grasslands. DA-derived NLW tends to be overestimated globally, with a notable correlation between DA-estimated NLW and LAI, especially in regions like central South America, Northern Africa, and Europe. This correlations is attributed to increased LAI affecting light transmission and, consequently, net shortwave radiation to the land surface.

In areas characterized by high value and substantial differences in LAI, such as the Amazon, the corresponding response in comparatively less pronounced. Conversely, regions predominantly covered by woody savanna or savannas, such as central Africa, exhibit a more substantial impact on NLW owing to the prevalence of specific vegetation types. Consequently, the influence of LAI changes on radiation dynamics is less pronounced in high LAI regions when compared to areas with lower LAI values.

Fig. 5 illustrated the temporal evolution of regional averaged NLW during the year 2002, both on global and regional scales.

TABLE III
MEAN RELATIVE DIFFERENCE AND RMSD OF LAND SURFACE VARIABLES FOR CN AND CTL EXPERIMENTS OVER TYPICAL REGIONS

		Global	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8
NLW	CTL	−0.06	−0.41	−0.37	−0.35	0.35	−0.04	−0.22	0.22	0.08
rBias	CN	−0.04	−0.34	−0.27	−0.29	0.39	0.10	−0.23	0.20	0.08
NLW	CTL	6.58	28.24	22.27	28.99	32.21	14.67	34.67	43.91	31.91
RMSD	CN	5.79	24.76	16.22	25.07	33.61	16.48	36.42	40.57	31.03
T2m	CTL	−0.09	−0.17	−0.09	−0.10	0.17	−0.13	−0.09	0.19	−0.03
rBias	CN	−0.07	−0.14	−0.05	−0.07	0.16	−0.06	−0.10	0.19	−0.03
T2m	CTL	2.54	4.96	2.68	3.02	5.75	5.68	3.70	7.12	3.69
RMSD	CN	2.38	4.25	1.52	2.28	5.76	4.40	4.07	7.07	3.69
HS	CTL	−0.50	−0.84	−0.79	−0.64	0.68	−0.78	−0.68	0.08	−0.14
rBias	CN	−0.48	−0.79	−0.74	−0.57	0.79	−0.57	−0.70	0.06	−0.17
HS	CTL	30.11	41.57	56.16	61.66	66.48	34.50	87.91	44.28	50.88
RMSD	CN	28.90	39.25	52.45	56.17	66.62	28.44	90.59	42.85	54.84
LE	CTL	−0.24	−0.85	−0.76	−0.50	0.27	−0.39	−0.27	0.72	0.45
rBias	CN	−0.24	−0.81	−0.74	−0.53	0.06	−0.39	−0.27	0.82	0.60
LE	CTL	13.85	105.79	102.27	52.78	42.71	62.91	15.98	45.66	35.57
RMSD	CN	13.96	101.25	99.14	53.86	27.50	60.30	16.03	49.84	41.62
ST	CTL	0.04	−0.11	−0.03	−0.09	0.15	0.00	−0.12	0.24	−0.06
rBias	CN	0.05	−0.08	0.01	−0.06	0.19	0.02	−0.13	0.23	−0.06
ST	CTL	1.89	3.27	1.19	2.77	4.21	3.39	4.48	9.07	4.27
RMSD	CN	1.95	2.48	0.67	2.06	4.90	3.45	4.54	8.78	4.27
SM	CTL	−0.18	0.14	−0.07	0.03	−0.01	0.06	−0.02	−0.06	−0.08
rBias	CN	−0.16	0.15	−0.07	0.06	0.05	0.11	−0.05	−0.09	−0.14
SM	CTL	0.05	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03
RMSD	CN	0.05	0.06	0.05	0.05	0.03	0.06	0.03	0.03	0.04

The parallelism in seasonal patterns by CN and CTL NLW suggests a strong reliance of the CN NLW on the underlying dynamics of the forwarding model. Notably, the disparity between CN NLW and GLDAS NLW is more pronounces or equivalent in regions 4, 6, 7, and 8. This observation indicates that in these specific regions, the assimilation of LAI into the model does not necessarily enhance the accuracy of NLW estimations. It suggests that the model's capacity to simulate NLW in these regions might be limited or unaffected by the incorporation of LAI data.

On the other hand, both the CN and CTL models demonstrate a commendable ability to replicate the magnitude and seasonality of NLW across most regions. However, exceptions are noted in Regions 5 and 8, where discrepancies are observed. These deviations might be attributable to various factors, including model sensitivity to regional climatic and environmental conditions, the accuracy of input data (like LAI), or inherent limitations in the models' algorithms.

Table III gives the mean relative difference and RMSD of land surface variables for CN and CTL experiments across typical regions. The largest improvements for NLW were found in regions 1, 2, and 3, with mean relative difference improving by 7%, 10%, and 6%, respectively. Correspondingly, the RMSD also decreased by 3.48, 6.05, and 3.93 W/m^2 for the same regions.

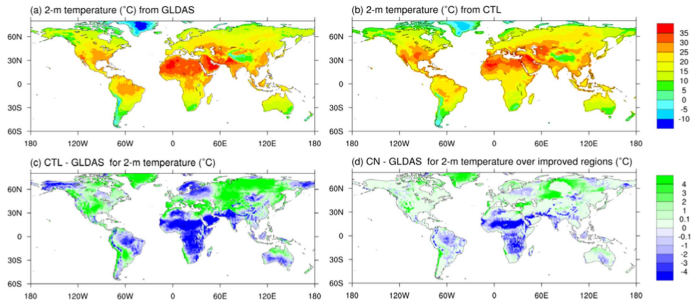

Fig. 6. Spatial distribution of T_{2m} in July 2002 for (a) GLDAS, (b) the CTL experiment, (c) CTL—GLDAS, and (d) CN—GLDAS over improved regions.

C. 2-m Air Temperature (T_{2m})

LAI exert an immediate and direct impact on the physical attributes of the land surface, particularly influencing crucial parameters such as albedo and surface roughness. These modifications in surface properties subsequently induce alterations in the surface energy balance, thereby affecting the 2-m air temperature (T_{2m}).

Fig. 6 illustrates the spatial distribution of T_{2m} in July 2002, presenting data from different sources and experiments: GLDAS; the CTL experiment; the difference between CTL and GLDAS; and the difference between the CN experiment and GLDAS across regions where improvements are observed.

The simulation results indicate a tendency to underestimate T_{2m} in several key regions: central Amazonia, northern North America, northern and central Africa, and most of Eurasia. In contrast, there is a pronounced overestimation of T_{2m} in central and southern North America, eastern and southeastern Eurasia, and most of Australia.

After assimilation, the enhanced LAI data mitigated the previous overestimation of T_{2m} in regions such as western and southern North America, northern Eurasia, and northern Australia. Conversely, for areas like southern Africa, central Amazon, and southwestern Eurasia, the refined LAI data helped in reducing the underestimation of T_{2m}.

Interestingly, the overall efficacy of DA in correcting the LAI estimates appears to be more pronounced in regions where T_{2m} was previously underestimated than in areas where it was overestimated. This distinction suggests that the integration of improved LAI data into the model is particularly beneficial in regions that suffer from a consistent underestimation of air temperature.

When considering both Figs. 4 and 6 together, a coherent pattern emerges regarding the global distribution of T_{2m} and NLW. There is a noticeable trend of decreasing T_{2m} and NLW in regions, such as western North America, central Asia, and Europe. This trend is associated with the increase in DA-estimated LAI, indicating a significant interaction between vegetation density and climatic factors such as temperature and radiation. This

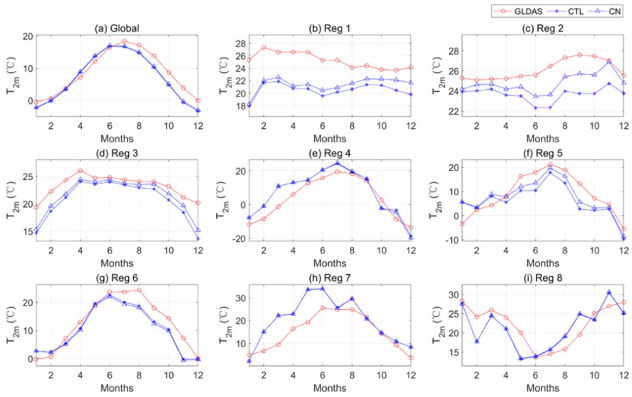

Fig. 7. Evolution of regional averaged T_{2m} over time at global and regional scales in 2002.

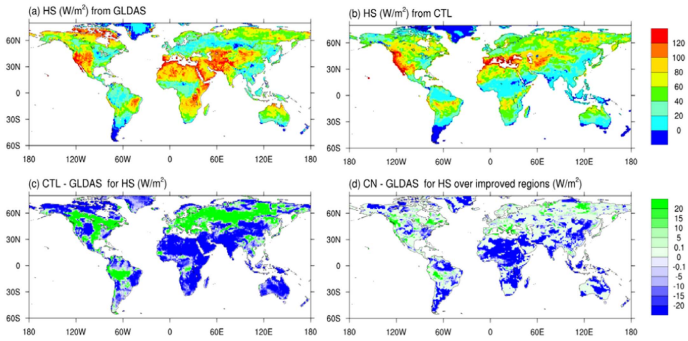

Fig. 8. Spatial distribution of sensible HS in July 2002 for (a) GLDAS, (b) the CTL experiment, (c) CTL—GLDAS, and (d) CN—GLDAS over improved regions.

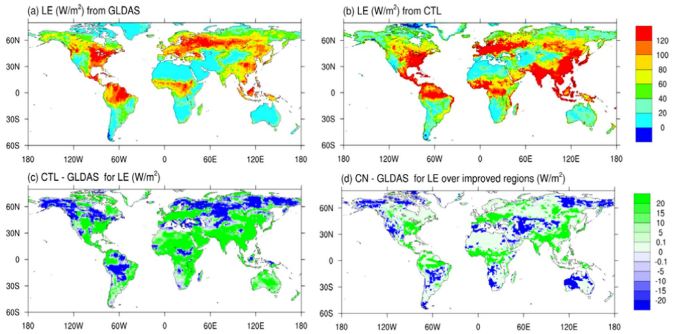

Fig. 9. Spatial distribution of LE in July 2002 for (a) GLDAS, (b) the CTL experiment, (c) CTL—GLDAS, and (d) CN—GLDAS over improved regions.

consistency underscores the importance of accurate vegetation representation in climate models, as it significantly impacts the model's ability to predict key climatic variables accurately.

Fig. 7 illustrates the temporal evolution of regional averaged T_{2m} at global and regional scales in the year 2002. Notably, the variations in T_{2m} across regions 1, 2, 3, and 5 are more pronounced compared to other regions, where T_{2m} remains relatively unchanged. After assimilation of LAI, there is a global improvement (or at least no detrimental change) in T_{2m}, with the exception of region 6.

Globally, assimilating LAI results in a 1% reduction in the mean relative difference of global average T_{2m} and a decrease of 0.15 °C in RMSD. The largest improvement for T_{2m} was found over region 5, with mean relative difference and RMSD improving by 7% and 1.28 °C, respectively. Next is region 2, with Mean Relative Difference improving by 4% and RMSD decreasing by 1.16 °C.

In forested areas (regions 1–3), there is a tendency for T_{2m} to increase as LAI decreases, and the extent of change in T_{2m} is more significant than that in both ground temperature (TG) and vegetation temperature (TV). The graphical representation of TG and TV is illustrated in Appendix, Fig. A1 and A2. In area with sparse vegetation, the impact of reduced LAI on T_{2m} is comparatively muted due to the inherently low LAI values. However, the limited regulatory capacity of the vegetation in these areas means that the variations in T_{2m} attributable to LAI alterations are not effectively captured by either the CTL or CN experiments. In forested areas, the amplification of TV is less pronounced than that of TG, owing to the strong moderating influence of the vegetation. During the growing seasons in region 4, even though the discrepancy between CN and CTL T_{2m} is minor, CN's TV and TG are higher. This observation underscores the regulatory role of different vegetation types on local and regional climate dynamics.

D. Sensible/latent Heat Fluxed (HS and LE)

Fig. 8 presented the spatial distribution of sensible HS for July 2002, utilizing data from various sources: GLDAS; the CTL experiment; the difference between CTL and GLDAS; and the difference between the CN experiment and GLDAS, particularly emphasizing regions where notable improvements have been observed. Generally, an increase in vegetation cover is associated with a diminution in HS, whereas a decrease in vegetation cover tends to elevate HS levels. LAI clearly affect the distribution of energy balance between the land and atmosphere.

Simulations have identified a tendency towards the overestimation of HS in specific geographical locals, notably in northern and northeastern Eurasia, central South America, and northern and eastern North America. Importantly, the integration of DA-estimated LAI has proven effective in rectifying these overestimations, with pronounced corrections observed particularly in central South America and northern and eastern North America. This correctional capability of the DA-estimated LAI is indicative of the enhanced precision of the model in capturing land-atmosphere interaction phenomena.

Moreover, the DA-estimated LAI also facilitates the rectification of underestimation in HS. This correction is more pronounced in instances where the CN experiment and GLDAS exhibit a negative discrepancy, suggesting a considerable improvement in the model's ability to accurately represent HS can also be corrected, while the negative difference of CN and GLDAS is larger than positive difference

Fig. 9 presented the spatial distribution of latent heat flux (LE) for July 2002 utilizing data from various sources: GLDAS; the CTL experiment; the difference between CTL and GLDAS; and

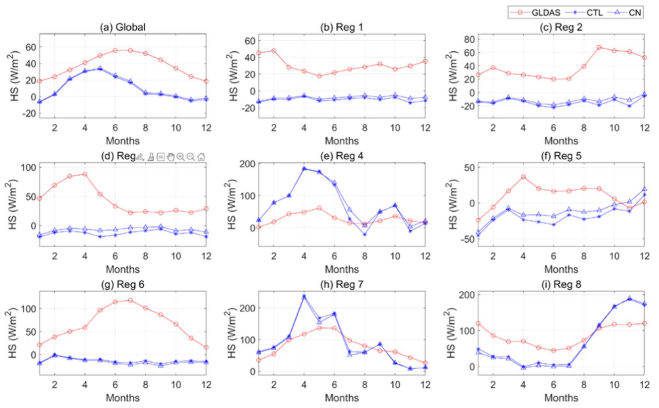

Fig. 10. Evolution of regional averaged HS over time at global and regional scales in 2002.

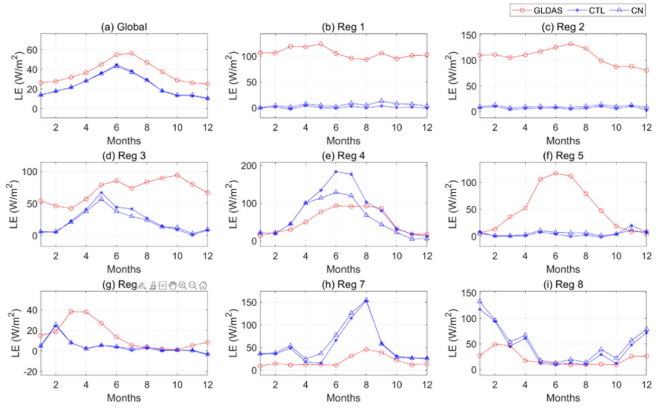

Fig. 11. Evolution of regional averaged LE over time at global and regional scales in 2002.

the difference between the CN experiment and GLDAS, particularly emphasizing regions where notable improvements have been observed. The regions characterized by sparse vegetation cover are associated with diminished LE values, contrary to the trends observed in HS.

Pronounced overestimations of LE were observed in certain regions, particularly in the northern and southern extremities of South America, across various parts of Africa, in Australia and within the northwestern and southern sectors of Eurasia. These regions, predominantly covered by lush vegetation including tropical rainforests, temperate forests, and boreal forests, appear to significantly influence the overestimated LE values, indicating a potential area for model refinement to more accurately represent the LE in these ecologically complex zones. In contrast, LE is consistently underestimated, most notably in central South America and within the latitude range of 50°N–60°N.

In eastern Asia, Southeast Asia, and eastern Australia, the DA-estimated LE exhibits a decreasing trend, a pattern that is concurrently observed with reduction in DA-estimated LAI. The decrease in LAI implies less foliage for transpiration, consequently leading to a reduction in LE. Conversely, an intriguing phenomenon is observed in Central Africa, Europe and South America, where LE increases despite a decrease in LAI. This inverse relationship indicates a more nuanced interplay, potentially influenced by the predominant vegetation types in these regions. The reduction in LAI suggests a decrease in transpiration rates due to less leaf surface area. However, this decrease in transpiration appears to be offset or even surpassed by an increase in evaporation from the land surface. It is possible that as LAI diminishes, more of the land surface is exposed, leading to higher rates of direct evaporation from the soil or water bodies, thereby elevating the overall LE.

Figs. 10 and 11 illustrate the temporal evolution of regional averaged HS and LE respectively, at both global and regional scales during the year 2002. Both the CTL and CN experiments are capable of accurately depicting the seasonal variations in HS and LE across regions 4, 5, 7, and 8, which are primarily situated in the mid-latitudes. However, in forested areas, both HS and LE tend to be underestimated. Despite this, the introduction of adjusted LAI values offers a measure of correction to this underestimation, enhancing the accuracy of these estimations to a certain degree.

On a global scale, assimilating LAI has not significantly improved the modeling capability of HS at the global level. The modeling capability of HS is slightly better than LE. At the regional scale, after assimilating LAI, HS over region 5 exhibits a 20% reduction in mean relative difference and a decrease of 6.1 W/m^2 in RMSD, consistent with the conclusions drawn for $T_{2\mathrm{m}}$. The most noticeable improvement in LE is observed in region 4, where mean relative difference and RMSD have decreased by 22% and 15.21 W/m^2, respectively.

E. Soil Variables

The land surface temperature exhibits an inverse relationship with LAI values, meaning that an increase in LAI leads to a decrease in surface soil temperature, and conversely, a decrease in LAI results in an increase in surface soil temperature. Both the CN and CTL experiments are adept at capturing the amplitude and seasonal patterns of soil temperature, as showed in Fig. 12(a)–(i).

Fig. 12(j)–(r) also presents the differences in soil temperature across four distinct layers, specifically at depths of 0–7 cm, 7–28 cm, 28–100 cm, and 100–255 cm. In regions characterized by dense forest cover, both surface and deeper soil layers exhibit a high sensitivity to variations in LAI. In contrast, in areas with sparse vegetation, such as grasslands and open scrublands predominantly found in the northern hemisphere, the deeper soil layer show a more pronounced response to changes in LAI compared to the surface soil layers. This observed pattern suggests that the land surface temperature at the top soil layer is not the most reactive to alterations in LAI. This phenomenon could potentially be linked to the depth of plant roots, as root depth may influence how soil temperatures respond to changes in LAI.

Fig. 13 illustrates the temporal evolution of the volumetric water content in both surface and subsurface soil at various depth: 0–7 cm; 7–28 cm, 28–100; and 100–255 cm. The data highlights distinct patterns of SM response to changes in LAI across different regions and depths.

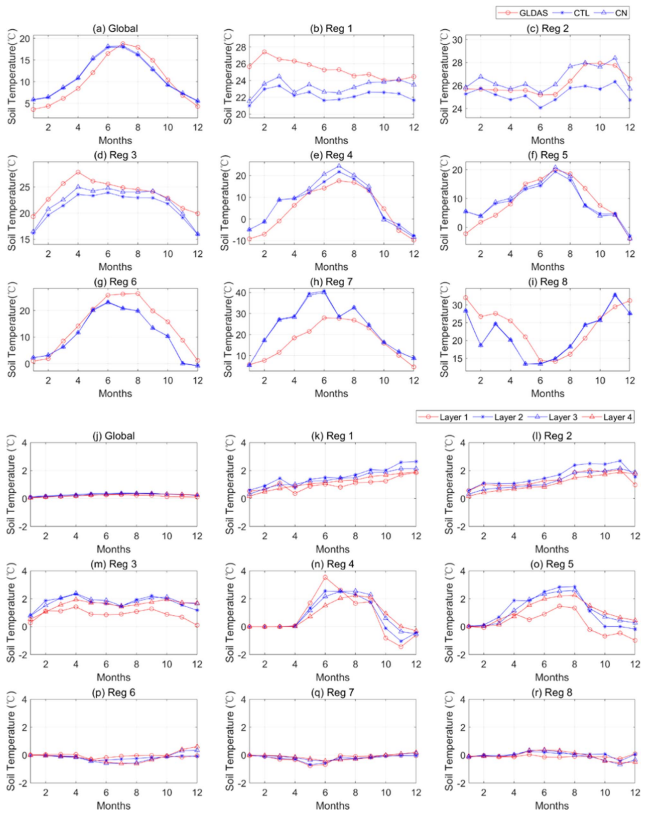

Fig. 12. (a)–(i) Evolution of regional averaged land surface temperature and (j)–(r) soil temperature at different depths over time at global and regional scales in 2002.

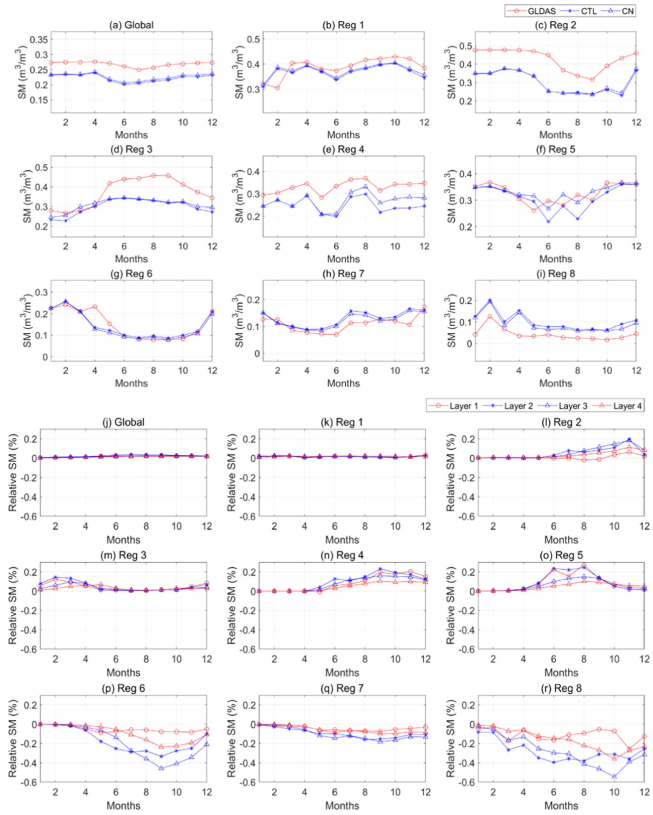

Fig. 13. (a)–(i) Evolution of regional averaged SM and (j)–(r) differences of SM at four depths over time at global and regional scales in 2002.

In forested areas (regions 1–3), the influence of LAI alterations on SM across various strata remains relatively subdued over time. This suggests a stable interaction between vegetation density and SM within these densely canopied ecosystems. Conversely, in regions characterized by sparser vegetation (regions 6–8), a pronounced responsiveness of SM is observed, particularly in the intermediate depth range of 28–100 cm. This implies that in environments with less vegetation cover, changes in LAI have a more marked impact on SM, especially at deeper soil levels, with the most significant fluctuations occurring in the 28–100 cm layer.

IV. Discussion

A. *Influence of Updating LAI, Leaf C, and Leaf N Simultaneously*

The experimental design of this article involves assimilating LAI and simultaneously updating Leaf C and Leaf N. From Fig. A3, it is evident that the global distribution patterns of Leaf C and Leaf N consistently align with LAI. Moreover, Leaf C levels are observed to be 1–2 orders of magnitude higher than Leaf N.

As discussed in [13] and [17], LAI value from the previous period does not directly influence the changes in LAI for the next period. As depicted in Fig. 1, during the variable updating process, DART incorporates model allocation rules pertinent to ecological processes. This integration signifies the imposition of constraints related to carbon and nitrogen dynamics. The discerned robust correlation existing between LAI and the botanical constituents, namely Leaf C and Leaf N, culminates in the refinement of ensemble initial conditions for subsequent temporal increments.

B. *Regional Disparities at the Scale of Different Regions*

Following the assimilation of LAI, conspicuous enhancements are foremost observed in LAI itself, followed by discernible improvements in radiative processes, with subsequent amelioration in temperature. The impact on heat flux, however, manifests a relatively moderate effect.

Regionally, a discernible heterogeneity in improvement effects becomes apparent, intricately linked to the prevailing vegetation types and climatic contexts. In low-latitude regions characterized by dense vegetation cover, exemplified in regions 1, 2, and 3, LAI assimilation yields superior enhancements in NLW. Conversely, in mid-latitude regions with dense vegetation cover, as exemplified in regions 4 and 5, where vegetation exhibits pronounced periodic variations, LAI assimilation engenders notable enhancements not only in LAI but also in temperature and heat flux.

Additionally, the regions wherein LAI improvements are most conspicuous do not necessarily align with those demonstrating the most pronounced enhancements in land surface variables. This discrepancy is attributed not only to vegetation types and climatic conditions but also to the direct interactions between the land and the atmosphere.

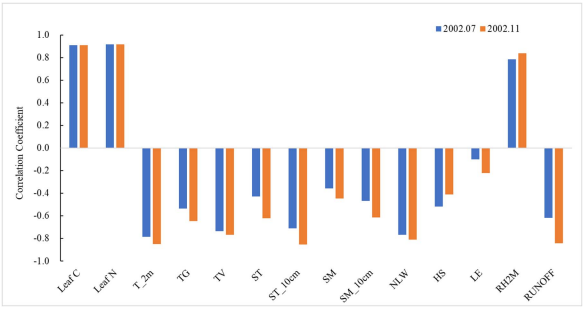

Fig. 14. Correlation coefficients between latitudinal average of LAI and latitudinal average of land surface variables in July and November 2002.

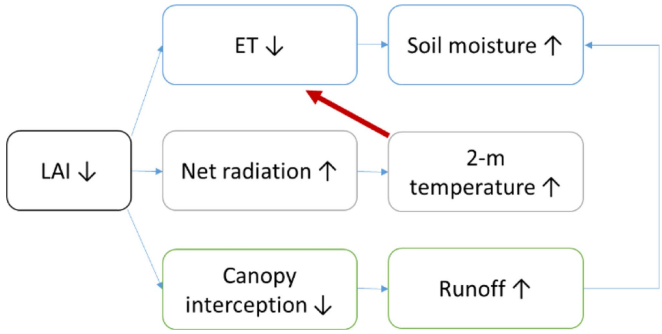

Fig. 15. Mechanism of LAI that affects surface state variables and land-atmosphere fluxes.

C. Interrelationships Between Different State Variables

Fig. 14 presents the correlation coefficients that quantify the relationship between the latitudinal distributions in the differences of various variables and the difference in LAI observed between the CN and CTL experiments during July and November of 2002. There is a marked positive correlation between the differences in Leaf C, Leaf N, and 2-m relative humidity (RH2M) with the difference in LAI. Specifically, the correlation coefficients for Leaf C, Leaf N, and RH2M in relation to LAI are 0.91 and 0.91, 0.92 and 0.92, and 0.79 and 0.84 for the months of July and November, respectively.

The variables that exhibit the most significant changes due to the latitudinal distribution shifts in LAI are the T2m and NLW. Their respective correlation coefficients with LAI differences are -0.79 and -0.77 for July, and -0.85 and -0.81 for November, indicating a strong inverse relationship. Furthermore, the changes in $T_{2\mathrm{m}}$ are more pronounces than those in TV and TG. The variations in soil temperature and SM at a depth of 10 cm are found to be greater than those at the surface.

Intriguingly, the response of all variables to changes in LAI during winter is more pronounces compared to summer, with the sole exception being the changes in HS. This suggests a heightened sensitivity of these variables to LAI variations in colder months.

D. Mechanism of LAI Influence Land Surface Variables

The mechanism through which LAI influences surface state variables and land-atmosphere fluxes are elucidated in Fig. 15. A decrease in LAI primarily leads to a reduction in canopy density, which in turn allows for greater penetration of solar radiation to the ground surface. This increase in solar radiation subsequently elevates longwave radiation, ultimately resulting in an increase in surface temperatures, including TSA, TG and TV.

Concurrently, the diminished canopy density also intensifies canopy interception, effectively increasing the volume of water droplets that reach the ground surface. Moreover, a reduction in LAI impacts ET by altering stomatal conductivity within the vegetation, leading to a decrease in soil water uptake by plant roots and a consequent increase in SM content.

However, it is important to note that changes in surface temperature also affect the evaporation process of vegetation, introducing a degree of uncertainty in modeling SM content. This aspect of SM content dynamics is beyond the scope of this article. Therefore, when analyzing the influence of vegetation on surface state, energy, and water balance, it is crucial to take into account various factors, including the type of vegetation, the geographic longitude and latitude of the study region, the height of the vegetation, the depth of the vegetation's roots, and other relevant factors.

V. Conclusion

In this article, assimilations (CN experiments) are compared with simulations (CTL experiments) to analyze the improvement of LAI on the surface state variables and land-atmosphere fluxes. GLASS LAI is assimilated into the CLM4CN model by utilizing DART, and Leaf C as well as Leaf N is updated every 8 days.

The regions at low latitude regions dominated by dense vegetation are the improved regions for simulating LAI, as well as NLW, $T_{2\mathrm{m}}$, HS, and LE. While the regions where LAI is not improved, model performance for simulating NLW, $T_{2\mathrm{m}}$, HS and LE is also not improved.

CN and CTL experiments can represent consistent characteristic of seasonal patterns for most variables, except for LE, which could be interpreted that LE is influenced by many factors including LAI. Assimilation of remotely sensed LAI can correct the underestimation for all variables, while the simulation ability of the model itself remains an important factor to consider.

The future direction of this research will consider assimilating more variables, for example, SM, PFTs, among other factors. Furthermore, extending the simulation/assimilation time is necessary.

Acknowledgment

The authors would like to thank the Jackson School of Geosciences, the University of Texas at Austin, and the Texas Advanced Computing Center. We are grateful to the High Performance Computing Center of Nanjing University for doing the numerical calculations in this article on its blade cluster system. We would like to thank the editor and the anonymous reviewers for exceptionally thoughtful reviews and suggestions that greatly improved this article.

REFERENCES

[1] P. J. Sellers, "Modeling the exchanges of energy, water, and carbon between continents and the atmosphere," *Science*, vol. 275, no. 5299, pp. 502–509, 1997.

[2] G. B. Bonan, D. Pollard, and S. L. Thompson, "Effects of boreal forest vegetation on global climate," *Nature*, vol. 359, no. 6397, pp. 716–718, 1992.

[3] J. M. Chen and T. A. Black, "Defining leaf area index for non-flat leaves," *Plant Cell Environ.*, vol. 15, no. 4, pp. 421–429, 1992.

[4] G. B. Bonan et al., "Improving canopy processes in the community land model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data," *J. Geophys. Res.-Biogeosci.*, vol. 116, 2011, Art. no. G02014.

[5] R. Myneni et al., "MODIS/Aqua Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V061," *Distrib. NASA EOSDIS Land Processes Distrib. Act. Arch. Center*, 2021. Accessed: Apr. 10, 2023. [Online]. Available: https://doi.org/10.5067/MODIS/MYD15A2H.061

[6] X. Zhao et al., "The global land surface satellite (GLASS) remote sensing data processing system and products," *Remote Sens.*, vol. 5, no. 5, pp. 2436–2450, 2013.

[7] F. Baret et al., "GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production," *Remote Sens. Environ.*, vol. 137, pp. 299–309, 2013.

[8] H. Fang et al., "An overview of global leaf area index (LAI): Methods, products, validation, and applications," *Rev. Geophys.*, vol. 57, no. 3, pp. 739–799, 2019.

[9] P. E. Thornton and N. E. Zimmermann, "An improved canopy integration scheme for a land surface model with prognostic canopy structure," *J. Climate*, vol. 20, no. 15, pp. 3902–3923, 2007.

[10] P. J. Lawrence and T. N. Chase, "Investigating the climate impacts of global land cover change in the community climate system model," *Int. J. Climatol.*, vol. 30, no. 13, pp. 2066–2087, 2010.

[11] Y. Xue, P. Sellers, J. Kinter, and J. Shukla, "A simplified biosphere model for global climate studies," *J. Climate*, vol. 4, no. 3, pp. 345–354, 1991.

[12] K. W. Oleson et al., "Technical description of version 4.0 of the community land model," 2010.

[13] X. L. Ling et al., "Assimilation of remotely sensed LAI Into CLM4CN using DART," *J. Adv. Model. Earth Syst.*, vol. 11, no. 8, pp. 2768–2786, 2019.

[14] B. Bonan et al., "An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the land data assimilation system LDAS-monde: Application over the Euro-Mediterranean region," *Hydrol. Earth Syst. Sci.*, vol. 24, no. 1, pp. 325–347, 2020.

[15] C. Albergel et al., "Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-monde assessment over the Euro-Mediterranean area," *Geoscientific Model Develop.*, vol. 10, no. 10, pp. 3889–3912, 2017.

[16] Y. Sawada, T. Koike, and J. P. Walker, "A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics," *J. Geophys. Res.-Atmos.*, vol. 120, no. 12, pp. 5910–5930, 2015.

[17] A. M. Fox et al., "Evaluation of a data assimilation system for land surface models using CLM4.5," *J. Adv. Model. Earth Syst.*, vol. 10, no. 10, pp. 2471–2494, 2018.

[18] A. L. Barbu et al., "Assimilation of soil wetness index and leaf area index into the ISBA-A-GS land surface model: Grassland case study," *Biogeosciences*, vol. 8, no. 7, pp. 1971–1986, 2011.

[19] A. Rahman et al., "The joint assimilation of remotely sensed leaf area index and surface soil moisture into a land surface model," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 20.

[20] H. S. Kang, Y. K. Xue, and G. J. Collatz, "Impact assessment of satellite-derived leaf area index datasets using a general circulation model," *J. Climate*, vol. 20, no. 6, pp. 993–1015, 2007.

[21] A. Anav et al., "Evaluation of land surface models in reproducing satellite derived leaf area index over the high-latitude northern hemisphere. Part II: Earth system models," *Remote Sens.*, vol. 5, no. 8, pp. 3637–3661, 2013.

[22] A. Maheu et al., "Is the annual maximum leaf area index an important driver of water fluxes simulated by a land surface model in temperate forests?," *Can. J. Forest Res.*, vol. 51, no. 4, pp. 595–603, 2021.

[23] M. Notaro et al., "Regional climate modeling of vegetation feedbacks on the Asian-Australian monsoon systems," *J. Climate*, vol. 30, no. 5, pp. 1553–1582, 2017.

[24] N. Montaldo, A. Gaspa, and R. Corona, "Multiscale assimilation of sentinel and landsat data for soil moisture and leaf area index predictions using an ensemble-Kalman-filter-based assimilation approach in a heterogeneous ecosystem," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 27.

[25] Q. Liu et al., "Assimilation of satellite reflectance data into a dynamical leaf model to infer seasonally varying leaf areas for climate and carbon models," *J. Geophys. Res.-Atmos.*, vol. 113, no. D19, 2008, Art. no. 025462.

[26] H. Dokoohaki et al., "Development of an open-source regional data assimilation system in PEcAn v. 1.7.2: Application to carbon cycle reanalysis across the contiguous US using SIPNET," *Geoscientific Model Develop.*, vol. 15, no. 8, pp. 3233–3252, 2022.

[27] S. N. Li et al., "Simulating carbon and water fluxes using a coupled process-based terrestrial biosphere model and joint assimilation of leaf area index and surface soil moisture," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 24, pp. 6311–6337, 2022.

[28] N. MacBean et al., "Quantifying and reducing uncertainty in global carbon cycle predictions: Lessons and perspectives from 15 years of data assimilation studies with the ORCHIDEE terrestrial biosphere model," *Glob. Biogeochemical Cycles*, vol. 36, no. 7, 2022, Art. no. 29.

[29] B. Raczka et al., "Improving CLM5.0 biomass and carbon exchange across the western United States using a data assimilation system," *J. Adv. Model. Earth Syst.*, vol. 13, no. 7, 2021, Art. no. 23.

[30] X. L. He et al., "Improving predictions of evapotranspiration by integrating multi-source observations and land surface model," *Agricultural Water Manage.*, vol. 272, 2022, Art. no. 20.

[31] O. M. Igder et al., "Multivariate assimilation of satellite-based leaf area index and ground-based river streamflow for hydrological modelling of irrigated watersheds using SWAT," *J. Hydrol.*, vol. 610, 2022, Art. no. 18.

[32] A. Rajib et al., "Watershed modeling with remotely sensed Big Data: MODIS leaf area index improves hydrology and water quality predictions," *Remote Sens.*, vol. 12, no. 13, 2020, Art. no. 17.

[33] X. X. Zhang et al., "The influence of assimilating leaf area index in a land surface model on global water fluxes and storages," *Hydrol. Earth Syst. Sci.*, vol. 24, no. 7, pp. 3775–3788, 2020.

[34] A. V. M. Ines et al., "Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction," *Remote Sens. Environ.*, vol. 138, pp. 149–164, 2013.

[35] D. M. Mocko et al., "Assimilation of vegetation conditions improves the representation of drought over agricultural areas," *J. Hydrometeorol.*, vol. 22, no. 5, pp. 1085–1098, 2021.

[36] H. Zare et al., "Combining crop modeling with remote sensing data using a particle filtering technique to produce real-time forecasts of winter wheat yields under uncertain boundary conditions," *Remote Sens.*, vol. 14, no. 6, 2022, Art. no. 26.

[37] H. Seo and Y. Kim, "Role of remotely sensed leaf area index assimilation in eco-hydrologic processes in different ecosystems over East Asia with community land model version 4.5-biogeochemistry," *J. Hydrol.*, vol. 594, 2021, Art. no. 16.

[38] P. R. Lin et al., "Snow data assimilation-constrained land initialization improves seasonal temperature prediction," *Geophys. Res. Lett.*, vol. 43, no. 21, pp. 11423–11432, 2016.

[39] K. E. Mitchell, "The multi-institution North American land data assimilation system (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system," *J. Geophys. Res.*, vol. 109, 2004, Art. no. D07S90.

[40] Y. L. Xia et al., "Comparative analysis of relationships between NLDAS-2 forcings and model outputs," *Hydrol. Processes*, vol. 26, no. 3, pp. 467–474, 2012.

[41] M. Pan et al., "Snow process modeling in the north American land data assimilation system (NLDAS): 2. Evaluation of model simulated snow water equivalent," *J. Geophys. Res.-Atmos.*, vol. 108, no. D22, 2003, Art. no. 024562.

[42] M. Rodell et al., "The global land data assimilation system," *Bull. Amer. Meteorol. Soc.*, vol. 85, no. 3, 2004, Art. no. 381.

[43] C. M. J. Jacobs et al., "Evaluation of European land data assimilation system (ELDAS) products using in situ observations," *Tellus Ser. a-Dyn. Meteorol. Oceanogr.*, vol. 60, no. 5, pp. 1023–1037, 2008.

[44] C. Huang and X. Li, "A review of land data assimilation system," *Remote Sens. Technol. Appl.*, vol. 19, no. 5, pp. 424–430, 2004.

[45] M. L. Carrera, S. Bélair, and B. Bilodeau, "The Canadian land data assimilation system (CaLDAS): Description and synthetic evaluation study," *J. Hydrometeorol.*, vol. 16, no. 3, pp. 1293–1314, 2015.

[46] E. Kalnay et al., "4-D-Var or ensemble Kalman filter?," *Tellus Ser. a-Dyn. Meteorol. Oceanogr.*, vol. 59, no. 5, pp. 758–773, 2007.

[47] P. Gauthier et al., "Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the meteorological service of Canada," *Monthly Weather Rev.*, vol. 135, no. 6, pp. 2339–2354, 2007.
[48] M. Eltahan and S. Alahmadi, "Numerical dust storm simulation using modified geographical domain and data assimilation: 3DVAR and 4DVAR (WRF-Chem/WRFDA)," *IEEE Access*, vol. 7, pp. 128980–128989, 2019.
[49] S. F. Kabir, G. A. Assumaning, and S. Y. Chang, "Efficiency of using 4DVar, 3DVar and EnKF data assimilation methods in groundwater contaminant transport modelling," *Eur. J. Environ. Civil Eng.*, vol. 23, no. 4, pp. 515–531, 2019.
[50] G. Tiwari and P. Kumar, "Predictive skill comparative assessment of WRF 4DVar and 3DVar data assimilation: An Indian Ocean tropical cyclone case study," *Atmos. Res.*, vol. 277, 2022, Art. no. 025462.
[51] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics," *J. Geophys. Res.-Oceans*, vol. 99, pp. 10143–10162, 1994.
[52] G. Evensen, *Data Assimilation, the Ensemble Kalman Filter*, Berlin, Germany: Springer, 2007, Art. no. 279.
[53] E. Kalnay, *Amospheric Modeling, Data Assimilation and Predictability*. Cambridge, U.K., MA, USA: Cambridge Univ. Press, 2003, Art. no. 512.
[54] J. L. Anderson, "An ensemble adjustment Kalman filter for data assimilation," *Monthly Weather Rev.*, vol. 129, no. 12, pp. 2884–2903, 2001.
[55] M. Zupanski, "Maximum likelihood ensemble filter: Theoretical aspects," *Monthly Weather Rev.*, vol. 133, no. 6, pp. 1710–1726, 2005.
[56] E. J. Fertig, J. Harlim, and B. R. Hunt, "A comparative study of 4DVAR and a 4D ensemble filter: Perfect model simulations with Lorenz-96," *Tellus Ser. a-Dyn. Meteorol. Oceanogr.*, vol. 59A, pp. 96–100, 2007.
[57] J. Harlim and B. R. Hunt, "Four-dimensional local ensemble transform Kalman filter: Numerical experiments with a global circulation model," *Tellus Ser. a-Dyn. Meteorol. Oceanogr.*, vol. 59, no. 5, pp. 731–748, 2007.
[58] J. Anderson et al., "The data assimilation research testbed a community facility," *Bull. Amer. Meteorol. Soc.*, vol. 90, no. 9, pp. 1283–1296, 2009.
[59] K. D. Raeder, J. L. Anderson, T. J. Hoar, N. S. Collins, H. Kershaw, and A. Chatterjee, "DART data assimilation with CESM models," in *Proc. CESM Workshop Univ. Corp. Atmos. Res. Nat. Center Atmospheric Res.*, 2015, pp. 1–6.
[60] K. Raeder et al., "DART/CAM: An Ensemble data assimilation system for CESM atmospheric models," *J. Climate*, vol. 25, no. 18, pp. 6304–6317, 2012.
[61] Y. Kwon et al., "Estimating snow water storage in North America using CLM4, DART, and snow radiance data assimilation," *J. Hydrometeorol.*, vol. 17, no. 11, pp. 2853–2874, 2016.
[62] Y. F. Zhang et al., "Assimilation of MODIS snow cover through the data assimilation research testbed and the community land model version 4," *J. Geophys. Res.-Atmos.*, vol. 119, no. 12, pp. 7091–7103, 2014.
[63] F. S. Castruccio et al., "An EnOI-based data assimilation system with DART for a high-resolution version of the CESM2 ocean component," *J. Adv. Model. Earth Syst.*, vol. 12, no. 11, 2020, Art. no. 045628.
[64] L. Zhao and Z. L. Yang, "Multi-sensor land data assimilation: Toward a robust global soil moisture and snow estimation," *Remote Sens. Environ.*, vol. 216, pp. 13–27, 2018.
[65] X. L. Ling et al., "Comparison of different sequential assimilation algorithms for satellite-derived leaf area index using the data assimilation research testbed (version Lanai)," *Geoscientific Model Develop.*, vol. 12, no. 7, pp. 3119–3133, 2019.
[66] T. T. Qian et al., "Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations," *J. Hydrometeorol.*, vol. 7, no. 5, pp. 953–975, 2006.
[67] M. J. Shi et al., "Spin-up processes in the Community Land Model version 4 with explicit carbon and nitrogen components," *Ecol. Model.*, vol. 263, pp. 308–325, 2013.
[68] A. Verger, F. Baret, and M. Weiss, "Near real-time vegetation monitoring at global scale," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3473–3481, Aug. 2014.
[69] F. Camacho et al., "GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products," *Remote Sens. Environ.*, vol. 137, pp. 310–329, 2013.
[70] Z. Y. Hu et al., "Temporal and spatial variations of soil moisture over Xinjiang based on multiple GLDAS datasets," *Front. Earth Sci.*, vol. 9, 2021, Art. no. 654848.
[71] X. Yan et al., "Jointly using the GLDAS 2.2 model and GRACE to study the severe Yangtze flooding of 2020," *J. Hydrol.*, vol. 610, 2022, Art. no. 127927.
[72] G. X. Zhang et al., "Drought monitoring and evaluation using ESA CCI and GLDAS-Noah soil moisture datasets across China," *Theor. Appl. Climatol.*, vol. 144, no. 3-4, pp. 1407–1418, 2021.
[73] M. Yin et al., "Climate impacts of parameterizing subgrid variation and partitioning of land surface heat fluxes to the atmosphere with the NCAR CESM1.2. Geosci," *Model Develop.*, vol. 16, pp. 135–156, 2023.

Xiaolu Ling born in Taixing City, Jiangsu Province, China on August 25, 1986. She received the Ph.D. degree in atmospheric sciences from Nanjing University, Nanjing, China, in 2016.

She is currently an Associate Professor with China University of Mining and Technology, Xuzhou, China. Her research primarily centers on the application of remote sensing data in understanding land-atmosphere interactions, and atmospheric environmental remote sensing.

Jian Gao born in Tianchang City, Anhui Province, China on July 1, 1998. He received the Bachelor of Engineering degree in surveying and mapping engineering from Liaoning Technical University, Fuxin, China, in 2021. He is currently working toward the master's degree in resources and environment (surveying and mapping engineering) with China University of Mining and Technology, Xuzhou, China.

His research interest includes on atmospheric environmental remote sensing.

Zeyu Tang born in Suzhou City, Jiangsu Province, China on April 3, 2000. He received the Bachelor of Engineering degree in remote sensing from Nanjing University of Information Science and Technology, Nanjing, China, in 2022. He is currently working toward the master's degree in photogrammetry and remote sensing with China University of Mining and Technology, Xuzhou, China. His primary research focus is atmospheric environmental remote sensing.

Wenhao Liu born in Harbin, Heilongjiang Province, China on December 14, 1998. He received the Bachelor of Science degree in geographic information science in 2021 from China University of Mining and Technology, Xuzhou, China, where he is currently working toward the doctorate degree in photogrammetry and remote sensing.

He actively participated in the International Geoscience and Remote Sensing Symposium in 2022 and 2023, a prestigious international conference on Earth Sciences and Remote Sensing organized by IEEE. He has contributed to the field with two conference papers published at these events.