ADVANCING EARTH AND SPACE SCIENCES

Reviews of Geophysics

INVITED ARTICLE
10.1029/2022RG000801

# Land Data Assimilation: Harmonizing Theory and Data in Land Surface Process Studies

**Xin Li**[1], **Feng Liu**[2], **Chunfeng Ma**[2], **Jinliang Hou**[2], **Donghai Zheng**[1], **Hanqing Ma**[2], **Yulong Bai**[3], **Xujun Han**[4], **Harry Vereecken**[5], **Kun Yang**[6], **Qingyun Duan**[7], and **Chunlin Huang**[2]

[1]National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China, [2]Key Laboratory of Remote Sensing of Gansu Province, Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China, [3]College of Physics and Electrical Engineering, Northwest Normal University, Lanzhou, China, [4]Chongqing Jinfo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing, China, [5]Forschungszentrum Jülich, Agrosphere (IBG 3), Jülich, Germany, [6]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China, [7]College of Hydrology and Water Resources, Hohai University, Nanjing, China

**Key Points:**

- Land data assimilation (LDA) advances scientific understanding and serves as an engineering tool for Earth system sciences
- LDA reflects the trend of harmonizing theory and data in the era of big data and artificial intelligence
- Future LDA research should expand the applications from pure geophysical systems to coupled natural and human systems

**Correspondence to:**

X. Li,
xinli@itpcas.ac.cn

**Citation:**

Li, X., Liu, F., Ma, C., Hou, J., Zheng, D., Ma, H., et al. (2024). Land data assimilation: Harmonizing theory and data in land surface process studies. *Reviews of Geophysics*, *62*, e2022RG000801. https://doi.org/10.1029/2022RG000801

Received 6 FEB 2023
Accepted 14 FEB 2024
Corrected 1 APR 2024

This article was corrected on 1 APR 2024. See the end of the full text for details.

**Author Contributions:**

**Conceptualization:** Xin Li, Feng Liu, Chunfeng Ma, Jinliang Hou, Donghai Zheng
**Investigation:** Feng Liu, Chunfeng Ma, Jinliang Hou, Donghai Zheng, Hanqing Ma
**Project administration:** Xin Li
**Visualization:** Feng Liu, Chunfeng Ma, Jinliang Hou, Donghai Zheng, Hanqing Ma
**Writing – original draft:** Xin Li, Feng Liu, Chunfeng Ma, Jinliang Hou, Donghai Zheng, Hanqing Ma, Yulong Bai
**Writing – review & editing:** Xin Li, Feng Liu, Chunfeng Ma, Jinliang Hou, Donghai Zheng, Hanqing Ma, Yulong Bai, Xujun Han, Harry Vereecken, Kun Yang, Qingyun Duan, Chunlin Huang



## Abstract

Data assimilation plays a dual role in advancing the “scientific” understanding and serving as an “engineering tool” for the Earth system sciences. Land data assimilation (LDA) has evolved into a distinct discipline within geophysics, facilitating the harmonization of theory and data and allowing land models and observations to complement and constrain each other. Over recent decades, substantial progress has been made in the theory, methodology, and application of LDA, necessitating a holistic and in-depth exploration of its full spectrum. Here, we present a thorough review elucidating the theoretical and methodological developments in LDA and its distinctive features. This encompasses breakthroughs in addressing strong nonlinearities in land surface processes, exploring the potential of machine learning approaches in data assimilation, quantifying uncertainties arising from multiscale spatial correlation, and simultaneously estimating model states and parameters. LDA has proven successful in enhancing the understanding and prediction of various land surface processes (including soil moisture, snow, evapotranspiration, streamflow, groundwater, irrigation and land surface temperature), particularly within the realms of water and energy cycles. This review outlines the development of global, regional, and catchment-scale LDA systems and software platforms, proposing grand challenges of generating land reanalysis and advancing coupled land–atmosphere DA. We lastly highlight the opportunities to expand the applications of LDA from pure geophysical systems to coupled natural and human systems by ingesting a deluge of Earth observation and social sensing data. The paper synthesizes current LDA knowledge and provides a steppingstone for its future development, particularly in promoting dual driven theory-data land processes studies.

## Plain Language Summary

Land Data Assimilation (LDA) integrates numerical models with observation data to enhance predictions of key variables related to land surface processes, including soil moisture, snow, evapotranspiration, and groundwater. Consequently, LDA effectively tackles two of the fundamental scientific challenges in the Earth system: how to enhance the usability of exponentially increasing land observations and how to improve the predictive capability and accuracy of land models. Significant advancements have been made in the theory, methods, and applications of LDA since the 21st century. Breakthroughs have been attained in characterizing uncertainties, estimating multiple variables and parameters simultaneously, and incorporating big data analytics. LDA systems at global, regional, and catchment-scales, along with widely used software platforms, have been developed. However, future challenges facing LDA comprise improving meteorological forcing data, creating long-term land reanalysis, devising operational applications, and broadening its range to cover the critical zone. In the era of big data, LDA will evolve further by assimilating big Earth data and incorporating machine learning to develop digital twins from pure geophysical systems to coupled natural and human systems.

# 1. Data Assimilation and Land Surface Research

## 1.1. Challenges in the Predictability and Observability of Land Surface Processes

Modeling and observation are two major methods for investigating the spatiotemporal evolution of the Earth system. A numerical model, as the formulation of theory, provides a dynamic framework for the continuous simulation of Earth system spatiotemporally, while observational data capture key elements of the Earth system with respect to the specific range of time and locations at multiple scales. Data assimilation (DA), also known as geophysical model data assimilation or model-data fusion in geophysics, aims to optimally combine dynamic modeling (i.e., theory) with observations (i.e., data) to improve the estimation of the whole Earth system and has become an important discipline in geoscience (Carrassi et al., 2018). DA was first introduced in atmospheric science (Daley, 1997; Derber & Rosati, 1989). The successful application has led to the gradual application of DA to other disciplinary fields of Earth system sciences (X. Li et al., 2020). DA is considered "a strategy for the Earth system sciences" (NRC, 1991) and has also been used to address two major scientific questions in the Earth system, that is, predictability and observability of land surface processes.

Land models describe biogeophysical/geochemical processes in the Earth system using physically based governing equations combined with parameterizations of local-scale processes, which determine the state and exchange of energy, water, and carbon at local, regional, or global scales (Golaz et al., 2019; Lawrence et al., 2019; Wiltshire et al., 2021). Although land models have evolved from simple bucket models (Manabe, 1969) to more sophisticated biogeophysical/geochemical models (Dai et al., 2003; Lawrence et al., 2019; Sellers et al., 1986), challenges remain in continually improving their predictability and accuracy, especially given the drastic heterogeneity of land surface processes and the associated scaling challenges (Li et al., 2022). For instance, many governing equations of physical processes (e.g., soil water flow) adopted in current land models have been developed based on microscale observations (e.g., soil column scale), and uncertainties may arise when applying these equations at the macroscale (e.g., watershed scale), mainly due to the strong spatiotemporal heterogeneities of model parameters (e.g., soil hydraulic properties), initial conditions (i.e., model states such as soil moisture), and boundary conditions (i.e., near-surface atmospheric states such as precipitation). In addition, some land surface processes remain unknown or are difficult to parameterize, which also affects the predictability and accuracy of land models. For example, typical processes of the cryosphere (e.g., glaciers and permafrost) and fully closed water cycles consisting of soil-aquifer interactions have not been well represented by current land models (Y. Fan et al., 2019; H. Lu et al., 2020). Therefore, major challenges remain in reducing predictive uncertainties and improving the simulation accuracy of current land models (K. J. Beven & Cloke, 2012; Fisher & Koven, 2020; MacBean et al., 2022).

Land observations provide measurements of key variables, such as soil moisture and land surface temperature that govern land surface processes. These observational data are taken at specific observation times and within their represented space. With the rapid development of different observation technologies, observational data of land surface processes are exponentially increasing but irregularly distributed spatiotemporally, which creates substantial challenges in enhancing the usability of land observations to improve the understanding of land surface processes. Since most observational data represent different spatiotemporal scales of a specific land surface parameter or variable, their usability at various spatial scales is hampered (Moody & Woodcock, 1995). For example, the representative space of in situ observations varies with different observation variables or parameters and different observation methods and times. There may be a significant representativeness error when transforming the in situ observations to other spatial units (e.g., model grid), which is closely related to the spatial heterogeneity of observation variables. Similarly, the transformation of remote sensing observations to a model grid may also introduce representativeness error. In addition, forward radiative transfer models are often needed to transform the land surface state variables into estimated remote sensing observations, and the scale representativeness of these models may also lead to representativeness errors. Therefore, it is a challenging task to reduce the representativeness error of land observations to enhance their usability (Dokoohaki et al., 2022).

Furthermore, a deluge of big Earth data, such as massive quantities of Earth observation data and hyperdimensional and unstructured nonmainstream data, provides both opportunities and challenges. Big Earth data significantly enrich observational information but also result in the "curse of dimensionality" (Nguyen et al., 2020). This situation exacerbates the challenges of improving the usability of these observational data, and thus, a combination of land models and land observations is urgently needed to overcome both the predictability and observability of land surface processes. Land data assimilation (LDA), which allows land models and

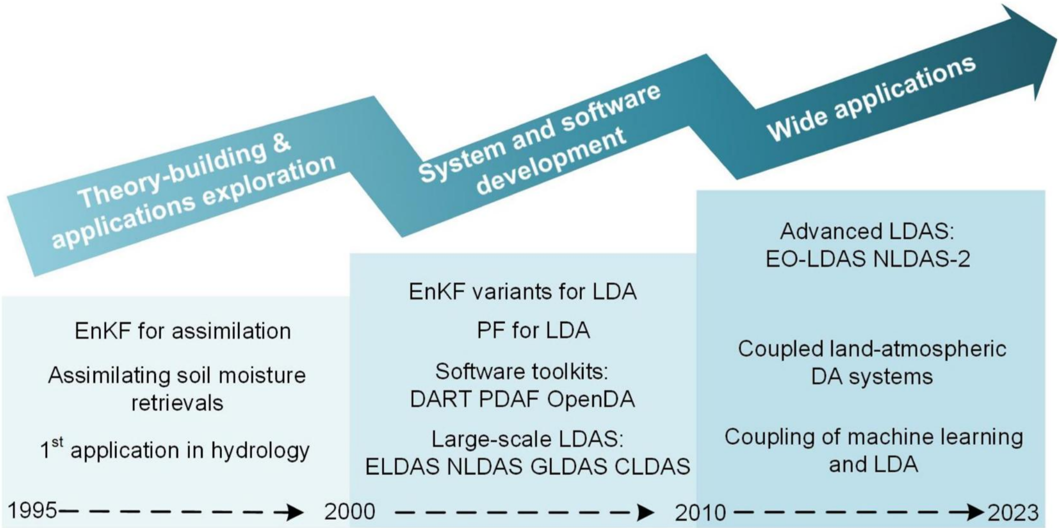


**Figure 1.** The developments of land data assimilation (LDA) during the recent decades. The first phase focused on theory building and application exploration. The second phase introduced a few pioneering practices in LDA systems (LDAS) and software. The third phase addressed the challenges of coupled land–atmosphere DA and integration of machine learning and LDA. (EnKF: Ensemble Kalman Filter, PF: Particle Filter, DART: Data Assimilation Research Testbed, PDAF: Parallel Data Assimilation Framework, OpenDA: Open Data Assimilation library, ELDAS: European LDAS, NLDAS: North American LDAS, GLDAS: Global LDAS, CLDAS: China LDAS, EO-LDAS: Earth Observation LDAS, NLDAS-2: NLDAS second phase).

observations to complement and constrain each other, has emerged as a new discipline in land surface research to address the fundamental challenges in the predictability and observability of land surface processes.

### 1.2. Land Data Assimilation: Model-Data Fusion to Generate the Optimized Estimation of Land Surface States

LDA was introduced into land surface research in the early 1990s (McLaughlin, 1995), which aims to produce estimates of land surface states, parameters, and fluxes that are as accurate as possible by using available observational data in land models weighted by model and observation errors. Within the framework of LDA, land models track the historical trajectory and predict the future change in land surface states and provide theoretical interpretations for land observations; land observations, after quality control, provide "true values" of land surface states in a specific space and at a specific time, which are ingested into model states via a feedback mechanism and used as references for adjusting the model trajectory (X. Li, 2014). The key objective of LDA is to integrate direct and indirect observations at different scales and from different sources within the framework of land models. As such, LDA is able to characterize and control uncertainties and thus provides optimized estimation of land surface states by maximally harmonizing modeling and observational data.

During the past three decades, LDA has seen rapid development and has become a mainstream methodology in land surface research, and important progress has been made in regard to its theory, methodology and application (Figure 1). From 1995 to 2000, LDA was introduced in hydrologic research, and LDA studies mainly focused on the assimilation of soil moisture retrievals based on the ensemble Kalman filter (EnKF) algorithm. Since 2000, more sophisticated methods, such as particle filters (PFs), have been introduced in LDA, and various large-scale land data assimilation systems (LDASs) (see Section 4) and LDA software (see Section 5) have been developed. After 2010, coupled land–atmosphere DA systems were developed, and the integration of machine learning and LDA has received increasing research interest (see Section 6), reflecting the trend of combing theory and data in the era of big data and artificial intelligence.

### 1.3. Essentials to Review Land Data Assimilation

Since significant progress has been witnessed in LDA in terms of theory, methodology and application, a holistic review is needed to summarize its milestones, point out knowledge gaps and provide potential directions for future research. To the best of our knowledge, to date, there have been few comprehensive review articles on LDA. Two early studies are either limited to the methods (e.g., linearization and uncertainties) and applications (e.g., precipitation) of hydrologic data assimilation (McLaughlin, 1995) or place more emphasis on the advantages of epistemic understanding in a general geophysical assimilation framework (Christakos, 2005). Other early studies focused on two major approaches, that is, introducing a series of LDAs for assimilating remote sensing data by using four-dimensional DA (Bach & Mauser, 2003) and reviewing hydrological DA with the extended Kalman filter (EKF) and EnKF (L. Sun et al., 2016). A recently published review focused on some specific aspect(s) of LDASs. For example, Y. Liu et al. (2012) conducted an in-depth review of operational hydrological DA by describing the theorem and mathematical work, uncertainty, utilization of data and common software. Montzka et al. (2012) reviewed assimilation strategies for different spatial scales and multisource observations. Xia et al. (2019) summarized the developments of regional, global, and project-based LDAS and the associated applications and validations. Additionally, Carrassi et al. (2018) provided a basic tutorial on LDA, and Baatz et al. (2021) reviewed the various reanalysis (estimations of historical observational data spanning to the present using an assimilation scheme) approaches in Earth system sciences, but with a specific focus on terrestrial ecosystem reanalysis.

The key novelties and value additions of this synthesis compared to previous reviews are described as follows. Our study aims at a comprehensive and in-depth review of the full spectrum of LDA, including distinguished characteristics (distinct from atmospheric and oceanographic DA) in methodological developments, typical applications in terrestrial water and energy budgets, developments of global and catchment-scale LDASs, and the development of software tools for LDA. In addition, we propose some grand challenges, including generating high-quality land surface forcing and long-time series land reanalysis, developing LDA operational system, and exploring coupled land-atmospheric DA and critical zone DA. We prospect new trends in assimilating nontraditional observational data, and combining big data and deep learning with LDA.

## 2. Theoretical and Methodological Innovations in Land Data Assimilation

LDA has distinguished itself as an innovative discipline from atmospheric and ocean DA due to the following features: (a) high nonlinearity combined with states, fluxes and parameters that are scale dependent and spatially correlated; (b) the non-Gaussian nature of model and observation errors; (c) the larger importance of forcing compared to that of initial values; and (d) the high sensitivity of model prediction to parameters. These issues require innovative theoretical and methodological development to reduce uncertainties from the nonlinearity and multiscale heterogeneity of land models and observations (Figure 2).

LDA theory and methods are anticipated to incorporate new approaches such as machine learning and to address uncertainty and its propagation. In addition, they are expected to establish general principles that can address the complexities of real-world scenarios, including nonlinear and non-Gaussian characteristics, and to develop robust and adaptive algorithms. These efforts will help LDA address the challenges arising from the analysis of big earth data, leading to the construction of real-time control systems for digital twins of land surface processes.

### 2.1. Continuous Versus Sequential Data Assimilation

DA methods are generally classified into three categories: continuous, sequential, and combined continuous-sequential methods. Continuous DA methods are predominated by variational methods, while sequential DA methods mainly consist of various Bayesian filter methods, such as the EnKF, PFs, and variants of these filters. The combined continuous-sequential DA methods integrate sequential methods and continuous methods. This section introduces the advantages and disadvantages of different DA methods, aiming to provide insight into a comprehensive understanding of the LDA methodological framework, thus guiding the application of these methods.

#### 2.1.1. Continuous Data Assimilation Methods

The 3D- and 4D-variational (4DVAR) DA (Lorenc et al., 2000; Rawlins et al., 2007) are the most frequently used continuous DA methods, but they require an adjoint model for minimizing the cost function. However, few efforts have been devoted to developing adjoint land models because the work is tedious and error prone (Margulis & Entekhabi, 2001; Reichle, McLaughlin, & Entekhabi, 2001; Raoult et al., 2016). Alternatively, heuristic optimization methods, such as simulated annealing (Pathmathevan et al., 2003) and genetic algorithms (Dumedah &

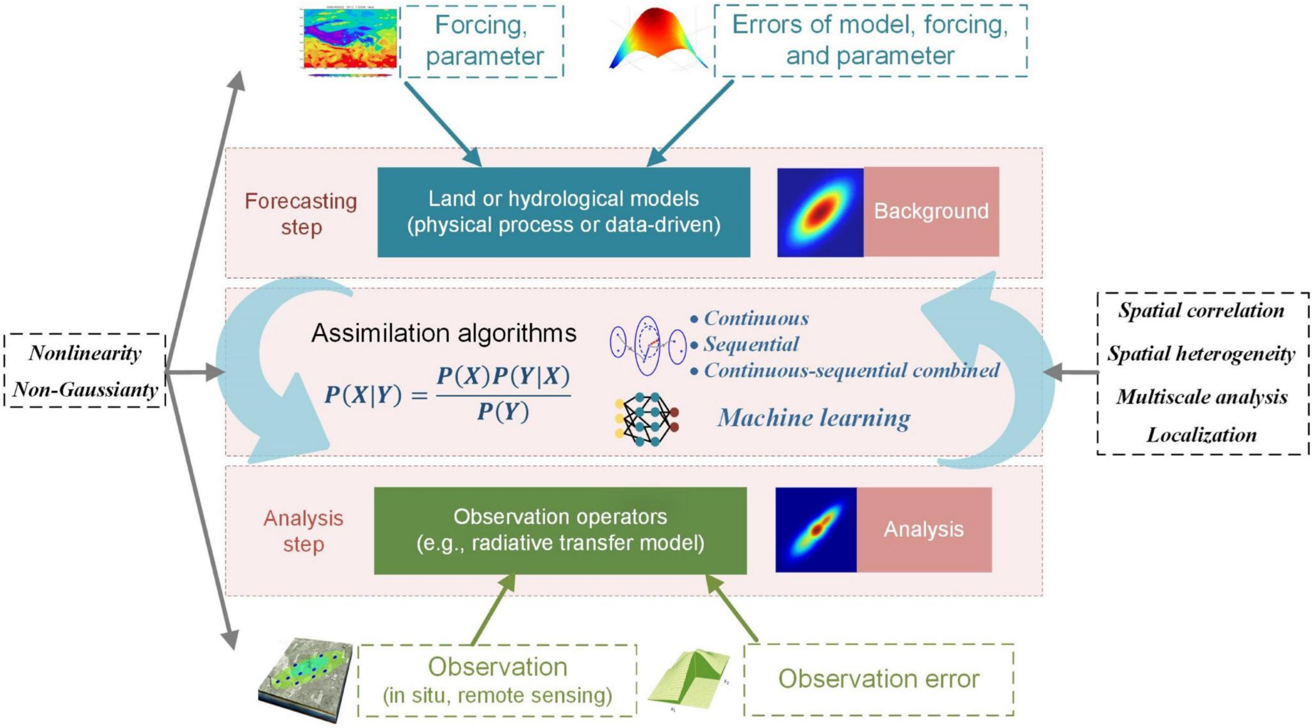


**Figure 2.** The generalized framework of Land Data Assimilation (LDA) consists of two primary steps: forecasting and analysis. During the forecasting step, land models are employed to generate predictions. In the analysis step, various observations are assimilated by mapping the model states to the observations in a specific spatial location and temporal range using the observation operators (e.g., radiative transfer models).

Coulibaly, 2013b), have been successfully applied in LDA. The heuristic methods outperform variational methods in finding global optima in the hilly structure of the cost function. However, they usually require tens of thousands of iterations to find the global optimum, hampering their operational use in LDA.

#### 2.1.2. Sequential Data Assimilation Methods

Linear and Gaussian assumptions are usually unrealistic for land models because high-dimensional and complicated land models involve full nonlinearity physics and non-Gaussian states (Apte et al., 2007). There are also many thresholds in land models, for example, the upper and lower bounds of soil moisture, meaning that the state variable is distributed in a truncated Gaussian state (Baguis & Roulin, 2017). Therefore, the EnKF is appealing for LDA because it can tackle the nonlinearity problem. Additionally, the EnKF avoids computing the tangent linear or adjoint of the model and observation operators, making it easier to implement (Evensen, 2003; Kalnay et al., 2007). Moreover, many variants of the EnKF, such as the ensemble transform Kalman filter (ETKF, (Bishop et al., 2001)), have been proposed to avoid observation perturbation. Spatiotemporally chaotic behavior is an important characteristic of the EnKF, which may lead to spurious correlations between distant locations in the background covariance matrix (Hunt et al., 2007). To this end, the local ensemble transform Kalman filter (LETKF, Hunt et al., 2007) is introduced to capture the space of forecast uncertainties via a localization scheme, that is, a cutoff radius is applied to eliminate spurious correlations. Other nonlinear and non-Gaussian KFs, such as the unscented KF (Julier & Uhlmann, 2004) and central difference KF (Van Der Merwe, 2004), have not yet been applied in LDA because some of these methods cannot handle high-dimensional problems.

To overcome the Gaussian assumption of EnKF-based methods, a more universal method, that is, PF, adopting Monte Carlo and importance sampling, has been successfully introduced in LDA to solve the non-Gaussian/nonlinear assumption (Moradkhani, Hsu, et al., 2005; Van Leeuwen, 2017), which demonstrated that PF is superior to EnKF in several synthetic tests (DeChant & Moradkhani, 2012; Kivman, 2003). However, the MCMC resampling method subjectively determines the resampling step and acceptance probability of the candidate points, hence artificially enforcing the convergence of the sampled parameter values to a too narrow distribution. To overcome the limitation of parameter convergence, a particle-Differential Evolution Adaptive Metropolis (particle-DREAM) method was developed (Vrugt, 2016; Vrugt et al., 2013; Vrugt & Ter Braak, 2011). The particle-DREAM applied a different definition of the Metropolis acceptance probability to guarantee that MCMC resampling leaves the target distribution invariant. In addition to synthetic studies, real case studies regarding the PF have witnessed success in LDA (Abbaszadeh et al., 2018, 2019, 2020; Gavahi et al., 2020; L. Xu et al., 2020). However, the time consumption, filter degeneracy, and PF dimensionality problems hamper the precise estimation of the posterior. Resampling and PF variants can partially solve this issue by mitigating weight degeneration for particles (Bocquet et al., 2010). Advanced PFs, for instance, equivalent-weights PF (Ades & van Leeuwen, 2013), ensemble transform PF (Reich, 2013), and particle batch smoother (Dong et al., 2015), were developed and used in LDA. Another direction focuses on the hybrid scheme of PF and EnKF (Frei & Künsch, 2013) to exploit their respective strengths. Additionally, the first attempt to use Markov chain Monte Carlo (MCMC) in conjunction with PF in hydrologic data assimilation (Moradkhani et al., 2012) effectively improves the resampling of particles and increases the number of effective particles (Abbaszadeh et al., 2018). Furthermore, PF and genetic algorithms have been coupled and applied to improve soil moisture DA (Gavahi et al., 2020; L. Xu et al., 2020) and land model performance (C. Zhang et al., 2021). Nevertheless, these new approaches should be adopted in more realistic nonlinear and non-Gaussian LDAs to validate their usefulness and efficiency.

Ensemble filters cannot handle asynchronous observations, probably leading to the loss of water and energy balance constraints due to the limited ensemble size (Kalnay et al., 2007). To address the asynchronous observations, the ensemble Kalman smoother (EnKS) (Evensen & van Leeuwen, 2000) and particle batch smoother (Dong et al., 2015) were developed for use in LDA (Dunne & Entekhabi, 2006). To overcome the limitation of ensemble size that causes the underestimation of forecast error variance, covariance inflation is usually adopted (Miyoshi, 2011), in which various adaptive inflation approaches are applied to obtain optimal inflation parameters, such as Bayesian estimation-based (J. Anderson et al., 2009) and statistically based methods (H. Li et al., 2009).

#### 2.1.3. Combined Continuous-Sequential Data Assimilation Methods

The development of hybrid approaches that combine the advantages of continuous and sequential DA methods is a promising direction. A typical scheme uses the ensemble-estimated forecast error covariances to replace (or be part of) the background error covariances in 4DVAR and resets the ensemble mean via 4DVAR analysis (Bannister, 2017; Buehner et al., 2010; M. Zhang & Zhang, 2012). This scheme outperforms in limited ensemble sizes and assimilation windows and generates smaller errors. However, as noted by X. Tian et al. (2018), the traditional ensemble four-dimensional variational (En4DVAR) DA still strongly depends on the adjoint model of the underlying forecast model and hence is more complicated and requires much more computational time. Thus, a nonlinear least square En4DVar (X. Tian et al., 2018) was proposed to overcome these issues. Additionally, a variational ensemble DA framework was developed for the Joint UK Land Environment Simulator (JULES) land surface model (Pinnington et al., 2020), and a coupled 4DVar-PF approach was proposed (Abbaszadeh et al., 2019). These novel combined continuous-sequential DA methods, being unified in the Bayesian theoretical framework, are expected to provide new insights into improvements in future LDAs. For instance, the combined method could simultaneously produce estimates of state variables and parameters and improve the estimation of the posterior distribution, hence simultaneously improving the estimation of the state variables and parameters as well as their uncertainties (Abbaszadeh et al., 2019).

### 2.2. Quantification and Reduction of Uncertainties

The quantification and reduction of uncertainties related to understanding non-Gaussian and nonlinear land surface processes, initial states, boundary conditions, model parameters, and observations are illustrated in Figure 3. Managing these uncertainties is inherently complex and intractable, and should consider their specific features, such as the forcing uncertainties caused by spatial-temporal averaging in hydrologic (Moradkhani & Sorooshian, 2008) and Hydrometeorological predictions (Moradkhani et al., 2019), and multiscale spatial correlations and heterogeneities (Vereecken et al., 2015, 2022).

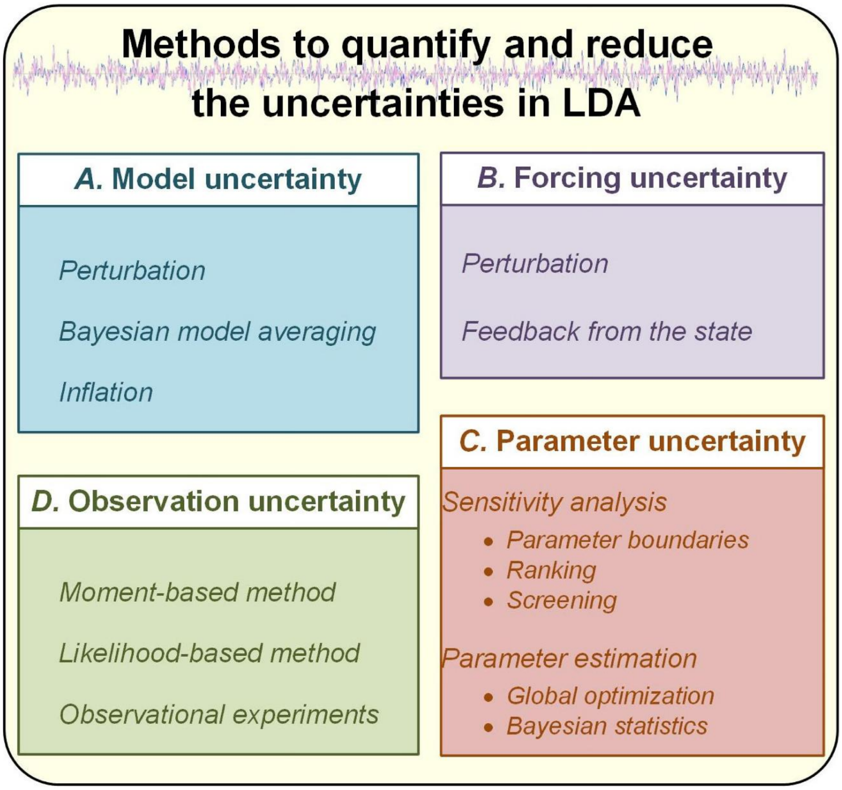


**Figure 3.** Methods to quantify and reduce the uncertainties in LDA: the uncertainties mainly stem from the model, forcing, parameter, and observation.

### 2.2.1. Model Uncertainty

Model uncertainty denotes the deviation of the state estimate of the model from the state truth (usually unknown), which partially arises from various sources, such as inadequate representations, omission of physical processes within the model structure (Abbaszadeh et al., 2019), and challenges in parameterizing structural error in land models (Pathiraja et al., 2018a, 2018b; Pathiraja, Moradkhani, et al., 2018). In LDA, model uncertainty is typically either ignored or simply treated by adding Gaussian white noise to the model states, for example, using the perturbation method (Figure 3a) (Crow & van den Berg, 2010). Another approach involves ensemble modeling (DeChant & Moradkhani, 2014). This approach entails running a set of independent ensemble members in a model or multiple models, and employing Bayesian model averaging to assess model structural uncertainty (Figure 3a) (Parrish et al., 2012).

An additional focus is to quantify posterior model uncertainty during land model simulations. This approach addresses the evolution of model error during the sequential update of model states and the error covariance matrix. In EnKF, filter collapse may occur as the sampling error is caused by the limited number of ensembles and usually underestimates the model error covariance. To address this underestimation, various inflation methods (Figure 3a) have been developed, such as multiplicative inflation (D. D. Anderson et al., 1999), additive inflation (Hamill & Whitaker, 2005), and the "relaxation-to-prior" method (H. Zhang et al., 2004). Such inflation approaches require ad hoc tuning to determine the optimal inflation factor. Therefore, some statistically based methods have been proposed to directly estimate lower-order model error statistics, such as those that are time-dependent and spatially correlated (Zheng, 2009). However, inflation methods assume unbiased model errors; therefore, bias correction techniques have been further applied to estimate and correct systematic errors (De Lannoy et al., 2007; Q. Zhang et al., 2019: Y. Zhang et al., 2019). Moreover, another stochastic parameterization scheme has been proposed to implicitly estimate model error statistics beyond unbiased Gaussian white noise, such as the crossover principle-based inflation (Bai & Li, 2011). Recently, an adaptive covariance inflation method was developed for model uncertainties in the EnKF (Raanes et al., 2019).

Capturing model uncertainty in the presence of input uncertainty in land surface modeling is challenging and involves controlling parameter or forcing uncertainty and addressing scale issues. One strategy for quantifying model error induced by input uncertainty is to conduct sensitivity analyses on parameter values or forcing data. A promising method is to use data-driven approaches, which possess advantages in handling limited prior knowledge of model uncertainty (Pathiraja et al., 2018a, 2018b; Pathiraja, Moradkhani, et al., 2018). To comprehend uncertainties stemming from spatial variability, a multiscale data assimilation (DA) framework is essential. Such a framework enables the assimilation of data sets with diverse spatiotemporal resolutions, facilitating the differentiation of model errors through the analysis of model outputs.

### 2.2.2. Forcing Uncertainty

The near-surface meteorological forcing (such as precipitation, radiation and air temperature) as the boundary condition of the model significantly affects the model state evolution of land models (Moradkhani et al., 2006). There are numerous methods to account for forcing data uncertainty. The most commonly used method is the forcing perturbation (Figure 3b). For example, a normally distributed random fluctuation with a mean of zero and a standard deviation equal to 50% of the nominal value was added to the initial forcing in the EnKF (Margulis et al., 2002). Later, Margulis et al. (2006) proposed a simple ensemble-based disaggregation scheme incorporating remotely sensed precipitation data in hydrological applications. In addition, conditional simulation methods have the potential to provide more reliable uncertainty estimates of model forcing (Clark & Slater, 2006).

An alternative method involves using feedback from the state variable to correct the forcing error (Figure 3b). For example, a DA approach that feedbacks the surface soil moisture via a soil water balance model to the rainfall magnitude to correct the rainfall accumulation error and analysis increments contains recent rainfall errors (Crow & Ryu, 2009). This approach has been further refined by considering the non-Gaussian distribution of the errors in rainfall by introducing the EnKF and PF (Crow et al., 2011). However, most of these works are empirical. Using robust/adaptive DA methods, such as the H-infinity filter, is a promising direction because it does not need prior information of model error and is insensitive to the forcing uncertainty (H. Lü et al., 2010; D. Wang & Cai, 2008).

### 2.2.3. Parameter Uncertainty

In contrast to atmospheric and oceanic DA, quantification of parameters uncertainties is of unique importance in LDA because overparameterization (C. Wang et al., 2016) and equifinality (K. Beven & Freer, 2001) are two major challenges in LDA. Estimation of model parameter uncertainty is usually preceded by sensitivity analysis, involving the establishment of parameter boundaries, ranking parameters, and subsequently screening the most sensitive ones (Figure 3c) (Gan et al., 2014; C. Wang et al., 2016). Parameter uncertainty estimation methods can be grouped into two categories: global optimization or Bayesian statistics. Global optimization converts parameter estimation for the land models into an optimization problem under the assumption of a unique optimal parameter set. Many attempts have been made with probabilistic parameter inference methods by mapping model uncertainty onto model parameters, such as SCE-UA algorithms (shuffled complex evolution method developed at the University of Arizona) (Duan et al., 1992). A typical LDA application entails the Shuffled Complex Evolution Metropolis (SCEM) algorithm in vadose zones and groundwater hydrology (Vrugt et al., 2003). The second type of method converts parameter estimation for the land models into a statistical inference problem that can be solved using Bayesian statistics, assuming parameters are random variables following a jointed probability distribution. In practice, a framework of sequential Monte Carlo with a genetic algorithm, differential evolution algorithm and Metropolis–Hasting algorithm was developed to evolve a population of particles to approximate the posterior parameter distributions (Zhu et al., 2018). Recent advances consider time-variant errors of parameters and then simultaneously estimate the model state and time-variant parameters, as detailed in Section 2.3. While comprehensive reviews on the quantification and reduction of parameter errors have been conducted by Y. Liu et al. (2012) and C. Wang et al. (2016), this paper provides a brief overview of parameter error estimation. Further developments are expected to produce more meaningful uncertainty estimation methods.

### 2.2.4. Observation Uncertainty

Observation uncertainty can be decomposed into instrument and representativeness errors (J. A. Waller et al., 2014). Instrument (or measurement) errors are associated with the corresponding measuring devices. Representativeness error, also called the forward model error, observation-operator error (Janjić et al., 2018), refers to the error generated either by mapping the model state from its model unit to the representative space of a certain observation or by mapping a model variable to the raw remote sensing observation (X. Li, 2014). LDA utilizes abundant raw remote sensing observations (Lewis et al., 2012) and directly assimilates these raw observations, such as reflectance, thermal emissivity, backscattering coefficient, and brightness temperature from passive microwave remote sensing. However, incorporating these raw observations necessitates the development of sophisticated observation operators. Typically, these observation operators are typically nonlinear and have remarkable representativeness errors. Methods to estimate representativeness error include moment-based methods (i.e., mean and variance methods), likelihood-based methods, and observational experiment-based estimation methods (Figure 3d).

Moment-based methods assume that the theoretical and empirical moments of the innovations (the deviations between forecasted and updated states or between the observations and states mapped in the observation space) are equal. Within the DA community, the Desroziers diagnostic is popular for examining various innovation statistics in the observation space (Desroziers et al., 2005). Most of the methods developed to date use this diagnostic. Kotsuki et al. (2017) estimated the real observation error covariance matrix, including horizontal observation error correlations, for densely observed data from satellites and radars. Guillet et al. (2019) proposed a method to represent spatially correlated observation errors in variational DA when the number of horizontally error-correlated observations is much larger. Another way to estimate error covariances is to exploit cross-correlations between lag innovations, namely, innovations between consecutive times. Empirical methods, such as the one proposed by Berry and Sauer (2013), have been established for nonlinear systems in DA. However, these methods have yet to be tested in the LDA system, and the idea of using lagged innovations seems to have great potential.

The statistics-based methods that employ maximized likelihood to determine optimal parameter values can be roughly divided into two categories. One approach uses a Bayesian framework, in which the elements of the covariance matrices are assumed to have a priori distributions with hyperparameters. The second involves maximizing a likelihood function using the iterative expectation–maximization algorithm. Pulido et al. (2018) used the expectation–maximization procedure to produce estimation matrices in Lorenz-96 systems. Recently, Cocucci et al. (2021) proposed an online adaptation of the expectation–maximization algorithm for the estimation of observation error at each time step. For LDA application, J. Tian et al. (2022) developed a dual-cycle data assimilation approach for soil moisture, in which the model and observation errors are adjusted through an optimization algorithm to optimize the likelihood function by innovation time series.

Finally, the observation errors are statistically a priori; therefore, they should be understood and quantified from observational experiments, particularly densely distributed, multiscale experiments (X. Li et al., 2013). For example, representativeness errors of solar radiation, surface fluxes, such as evapotranspiration and carbon dioxide flux, land surface temperature, and soil moisture, were investigated during a multiscale land surface process experiment (Pan & Wood, 2009). The representativeness error of in situ observations is very large, and the probability of using a single in situ observation to represent a model grid-scale ground truth is very low.

In general, quantifying and reducing the uncertainties in LDA is challenging when considering the interactions among the above four types of uncertainties. Many existing strategies are either empirical or statistical, often relying on prior information that cannot be obtained, such as randomly distributed observational data. This finding highlights the need for further study to better understand the underlying processes involved. Regarding the unique characteristics of LDA uncertainties, data-driven approaches and robust methods, such as the H-infinity filter, hold potential for developing knowledge-based techniques and robust, adaptive assimilation algorithms.

### 2.3. Simultaneous Estimation of Model States and Parameters

In land models, parameters could be more impactful than the initial field of state variables on the model performance. However, many parameter values are either unknown or only available at specific locations (i.e., limited spatial coverage) and are time-dependent. Traditional LDA only updates the model state while utilizing the default model parameter values, which inevitably leads to large uncertainties in the model predictions due to inaccurate model parameter values (Smith et al., 2013). Thus, accurate estimation of both state variables and model parameters is indispensable. Two distinct methods have been proposed to alleviate uncertainty in LDA by performing joint estimations of model states and parameters.

The first type of method is the state augmentation method, which regards the model parameters as state variables, whereby the original state vector is augmented with both the state variables and model parameters (Hendricks Franssen & Kinzelbach, 2008; Moradkhani, Sorooshian, et al., 2005). Thus, parameters are simultaneously updated with state variables in the LDA. The newly analyzed parameter values are then fed back into the model to affect the next forecast of state variables. Through this analysis-forecast evolution, the model gradually generates more accurate parameters and state estimates, and therefore, the method can effectively diminish the long-term systematic deviations and weaken the stochastic errors of the model forecast by estimating the time-varying characteristics of model parameters (X. J. Han et al., 2016; J. Qin et al., 2009). The state augmentation method explicitly accounts for the cross-covariance between the state and parameters, with the advantages of a simple principle, easy implementation, and high efficiency. However, this approach suffers from limitations arising from several aspects, such as challenges in accurately estimating parameters and states in cases with sparse observations and controlling the convergence of the model simulation and parameter estimation due to complex interactions between state and parameters in nonlinear dynamic systems (Moradkhani, Sorooshian, et al., 2005). In addition, treating parameters as state variables implies that they are time-varying, which may not be reasonable because some model parameters (e.g., soil texture) are essentially time-invariant (J. Tian et al., 2022).

The second type of method is the dual-pass/dual-cycle approach, in which system updating entails two passes or two cycles compared to the traditional DA approach. This method first estimates parameters by using DA or an optimization algorithm and then estimates state variables based on the optimized parameters. Previous studies have provided two different methods to accomplish this task: (a) both parameters and states are recursively updated by fitting or assimilating observations in each time step, and the parameters are treated as time-varying variables (Moradkhani, Sorooshian, et al., 2005). (b) the first pass or outer loop estimates time-invariant parameters by fitting observations within a long-term window, and the second pass or inner loop estimates the land state by assimilating observations in a short-term window (J. Tian et al., 2022; Vrugt et al., 2005; K. Yang et al., 2007, 2009, 2016; C. Zhang et al., 2021). The separated representation of state variables and parameters in the dual-pass/dual-cycle approach makes it possible to obtain more appropriate estimates for state variables and parameters. This method avoids the complex interactions between states and parameters. Compared to the state augmentation approach, the dual-pass/dual-cycle approach usually shows better robustness, but its heavy computational load hinders large-scale applications. A noticeable issue of the dual-pass approach is the overuse of observational information (each observation is used twice), which may cause an overcorrection problem. In addition, a fundamental assumption of the type of ensemble Kalman algorithms is that the state variables and observations are independent. However, the dual-pass/dual-cycle approach utilizes observations to alternately estimate parameters and states, which may undermine this hypothesis.

The simultaneous estimation of model states and parameters has offered unique opportunities for LDA developments. First, it can reduce the reliance on extensive data collection, which is particularly useful when data collection is expensive or time-consuming. Second, it can provide insights into the underlying processes that govern system behavior. By estimating model parameters together with model states, researchers can identify the most influential factors driving system dynamics. However, one of the main challenges is that the states and parameters are often entangled, which may lead to the problem of parameter and state identifiability in the context of simultaneous estimation of states and parameters. In this case, the constrained estimation of states and parameters can be considered, which may reduce the entanglement of states and parameters. Overall, the simultaneous estimation of model states and parameters offers a promising avenue for improving our understanding of complex systems and enhancing their control.

### 2.4. Multiscale Land Data Assimilation

Heterogeneity and related uncertainties are intrinsic properties of the land surface state variables, boundary conditions, and parameters (X. Li, 2014), but they are usually assumed to be deterministic processes at the microscale. In addition, many Earth observations from different sensors recorded at different temporal and spatial scales reveal that the representative spaces of models and observations have multiscale characteristics (X. Li et al., 2020). The multiscale characteristics of land modeling and observations require that the LDAs integrate model predictions and observations at different spatial and temporal scales. Multiscale LDA addresses the abovementioned issues and includes two types of methods.

10.1029/2022RG000801

First, the EnKF, PF, or variational methods are applied to simultaneously assimilate variables at different scales when multiscale measurements are available, such as the DA strategy of simultaneously estimating the multiscale leaf area index (LAI) from multiresolution satellite observations (Q. Zhang et al., 2019a). To solve the problem of scale mismatch, either an observation operator with spatial weight is constructed (X. J. Han et al., 2016), or scaling techniques, such as the interpolation method (C. S. Draper et al., 2012), statistical upscaling (Zhao et al., 2016), prior fusion or downscaling (Merlin et al., 2006), are performed before assimilation. The most common examples for multiscale LDA are the assimilation of coarse-resolution soil moisture contents (Lievens et al., 2016; Reichle, Entekhabi, & McLaughlin, 2001) or snow water equivalents (De Lannoy et al., 2012; Durand & Margulis, 2008) into finer spatial scale hydrologic models.

Second, multiscale filtering approaches are developed. An optimal multiscale Kalman filtering method has been proposed (Parada & Liang, 2004) to assimilate near-surface soil moisture into the three-layer variable infiltration capacity (VIC-3L) land surface model. A multiscale Kalman smoother (MKS) has also been developed (S. Wang et al., 2011) to fuse precipitation data from diverse sources with different spatial resolutions. The results show that the MKS-based framework can significantly restore the loss of spatial patterns and magnitude of precipitation associated with white noise and bias. These approaches also put forward the criteria for determining the optimal spatial scale when using the given data at different resolutions. A multiscale autoregressive framework, which processes signals that exhibit multiscale features, can provide a method for identifying the multiscale structure in signals and a filtering procedure. The multiscale autoregressive provides an efficient way to solve the optimal estimation problem for many high-dimensional dynamic systems. Therefore, a multiscale ensemble filtering system can efficiently address multiscale DA problems based on the multiscale autoregressive framework (M. Pan et al., 2009).

In short, multiscale LDA presents valuable insights into the complex interactions between land surface processes at different scales by integrating data from various sources. Although the theory of the multiscale DA method has been applied in LDA, the usability and stability of multiscale LDA algorithms are still not at operational standards, as data from different sources may have different uncertainties, and it is also difficult to manage the propagation of uncertainty. Future research should focus on the development of new methods and tools, such as more advanced scaling methods, improved DA algorithms that can effectively integrate data from multiple sources, and uncertainty quantification methods.

### 2.5. Spatial Correlation and Localization

Considering Tobler's first law of geography, spatial correlation is inherent in land surface processes. Spatial correlations in land surface processes are typically dominant at smaller scales (less than 1 km) compared to those in meteorological and oceanic processes (ranging from dozens of to hundreds of kilometers). In addition, land surface processes present more complex multiscale spatial correlations. Spatial correlation in LDA can play multiple roles, including predicting the no-value region using spatial interpolation (the correlation ranges and basic semivariogram models among multiple scales should be estimated in advance), characterizing the model errors and observation errors, and reducing the sampling error of the background error for the localization problem in ensemble-based DA (Buehner & Charron, 2007).

The mainstream approaches to consider spatial correlations in LDA use geostatistical techniques. Some methods can improve the assimilation efficiencies, for example, to compensate for the inaccuracies of macroscale rainfall-runoff simulations by assimilating streamflow observations (Pugliese et al., 2018) and to determine an adequate observation sequence to be assimilated based on semivariogram analysis (X. Han et al., 2012). Other works include developing new assimilated algorithms, such as introducing kriging into EnKF to simplify the time-consuming assimilation algorithm (Tolosana-Delgado et al., 2011). The spatial correlation of observations has been explored in LDA (X. Han et al., 2012), in which Gaussian, exponential and spherical spatial correlation functions from geostatistics were fitted and used in observation localization to update the uncovered region.

In LDA applications, when the number of ensemble members in the EnKF is smaller than the degree of freedom in the model operator, spurious spatial correlations and large sampling errors in the approximation of the background error covariance can be generated. Therefore, covariance localization (J. L. Anderson, 2012) and observation localization (Hunt et al., 2007) have been proposed to eliminate spurious spatial correlations in ensemble-based high-dimensional and multivariable LDA. Devegowda et al. (2010) illustrated that the 50-member ensemble in flow data assimilation causes spurious correlations, while introducing localization can

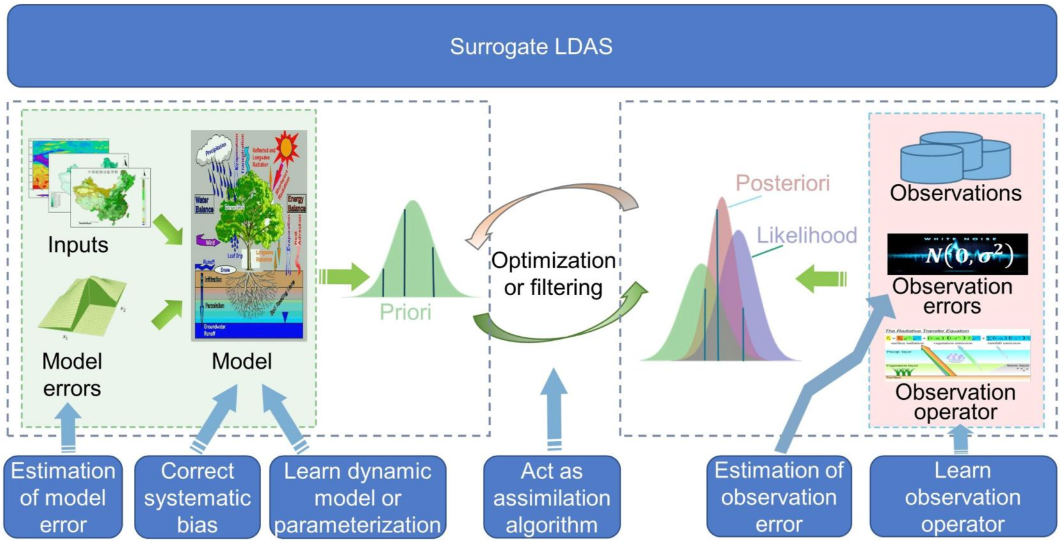


**Figure 4.** Integration of machine learning and LDA. Machine learning is used for model parameter optimization, model process/subprocess agent, assimilation system deviation correction, observation operator modeling, DA algorithm substitution, and LDAS surrogation.

correct this fault. The estimation with localization performs similarly to that from the 1000-member ensemble. The localization is generally implemented as a Schur product between the background (or observation) error covariance matrix and a spatial correlation-dependent function (or matrix). This localization function determines the observation assimilated during the analysis step. A typical function is GC (Gaspari & Cohn, 1999), but using different localization functions empirically with respect to observation types (such as temperature and u- and v-wind) might have limited improvement (L. L. Lei et al., 2015). The localization method encounters the following issues in LDA. (a) The localization function reduces the local analysis error at the cost of increasing the global analysis error. (b) The distance-dependent localization function needs to be refined for strongly heterogeneous and structured observations.

Current localization functions are statistical and empirical, and whether they vary with observation type (Y. Chen & Oliver, 2010) and spatial correlation characteristics in a specific study area has not yet been theoretically determined (L. L. Lei et al., 2015). Therefore, unifying the knowledge in geostatistics and the spatiotemporal representation of geophysical variables is a promising direction for improving studies on spatial correlation and localization.

### 2.6. Symbiotic Integration of Machine Learning and LDA

The data-driven machine learning (ML) method can thoroughly explore the spatiotemporal heterogeneity characteristics of land surface variables from historical data to optimize known parameters and processes or discover unknown parameters, state variables, and processes contained in the physical systems. In addition, it can learn the complex relationship between model or observation errors and related variables without a large amount of prior knowledge about various uncertainties in LDA systems. That is, ML leverages knowledge about the structure of data to enhance the accuracy and efficiency of physical process models. The integration of ML with LDA is a relatively new but quickly growing area of research and focuses on (Figure 4) (a) providing parametric time-varying boundaries or driving conditions in land surface models (Sawada, 2020; Q. Zhang et al., 2019a); (b) replacing the empirical or semiempirical process or subprocess and even the entire physical process model (Bao et al., 2020; Hou et al., 2021; Tang et al., 2021; Y. Wang et al., 2021); (c) correcting the system deviation of the physical model (King et al., 2020; Q. Zhang et al., 2019a); (d) developing the observation operator model (Kwon et al., 2019); (e) acting as a data assimilation algorithm (Boucher et al., 2020); and (f) surrogating the entire assimilation system to reduce computational complexity (P. J. Li et al., 2020). Moreover, DA can also be integrated into ML, although there is currently no relevant study available in LDA. However, there have been successful cases in the field of numerical weather prediction. For example, S. G. Penny et al. (2022) integrated

DA with RNNs to perform data-driven state estimation, where RNNs served as pretrained surrogate models for key components in the DA cycle in numerical weather prediction, and DA updated the hidden states of the RNN dynamics using observations of the target system. The study demonstrated that integrating DA and RNNs can generate reasonable representations of the system response to uncertainty in initial conditions.

Integrating DA into artificial intelligence (AI) yields new hybrid methods. DA provides an opportunity to address these pain points (e.g., gradient dependence learning process, offline training modes, low interpretability, and weak generalization ability) of AI, such as using DA to train AI models to construct gradient-free AI frameworks (B. Chen et al., 2022; C. Chen et al., 2022), embedding the kinetics of physical models through four-dimensional variational form loss function to obtain interpretable AI models (W. Wang et al., 2024), and developing online real-time AI models with high generalization ability by utilizing DA to optimize the parameters and states of learning models when new observations are available (Malartic et al., 2022). In short, established DA concepts and approaches are contributing to the development of innovative AI algorithms, which may provide some benefits even for moving toward artificial general intelligence (AGI).

It is encouraging that ML and LDA can be combined and complement each other (Buizza et al., 2022). The symbiotic integration of ML and LDA has two benefits. First, ML's ability to analyze a large amount of data in a shorter time and model the complex patterns in Earth system science may help LDA to (a) handle the dilemma that makes it difficult to obtain analytical solutions for complex partial differential equations in land/hydrological models (Tang et al., 2021; Y. Wang et al., 2021); (b) reduce the high computational complexity caused by a large number of iterative numerical calculations in solving physical equations (Bao et al., 2020; Hou et al., 2021; and (c) discover hidden knowledge and rules in the physical model to improve the simulation accuracy of land models (Yeung et al., 2022). Second, the constraints of physical knowledge contained in the physical model may help ML to (a) improve the spatiotemporal representation of training samples by enhancing the diversity of samples or features and strengthening the quality of training samples and (b) implicitly adhere to the relevant physical rules and relationships, which helps to improve the physical consistency, interpretability and extrapolation ability of ML models (Q. Z. He et al., 2020). In short, the integration of ML and LDA has the potential to provide a very appealing solution for a more accurate estimation of land surface variables, fluxes, and parameters.

Nevertheless, the symbiotic integration of LDA and ML still faces significant challenges. First, the observations in LDA are usually sparse, irregular, and locally distributed. That is, when integrating LDA and ML, few-shot learning or insufficient label sample (or even no label samples at all) learning must be addressed. In this case, how to use physical constraints to enhance the spatiotemporal representation of ML samples is a difficult problem. Second, it is also difficult to use the knowledge of geophysical laws, physical invariance, and physical relationships in the physical model to guide and optimize the learning process to obtain a physically consistent solution, that is, develop new physically guided ML models under the framework of LDA (Yeung et al., 2022). Third, the prediction of LDA presently relies mainly on geophysical theory-driven models. Under the new framework of coupling LDA and ML, it is necessary to consider how to alternately perform the assimilation process and ML modeling through the dual driven theory-data approach to enhance the forecasting ability of the system (Hou et al., 2021). This integration is expected to construct a real-time control layer for digital twins of land surface processes (X. Li et al., 2023).

## 3. Applications of Land Data Assimilation in Estimating Land Surface Variables

To date, LDA has witnessed a large amount of progress in the estimation of land surface variables by assimilating multisource and multiscale observations. This has distinguished LDA as an indispensably innovative solution for land surface science and inevitably provides insights into the accurate quantification of land processes in the future. This section comprehensively reviews the progress of assimilating key land surface variables (Figure 5), particularly, hydrological state variables and fluxes into different physical models to improve their predictability and accuracy. Table 1 summarizes the critical hydrology variables in LDA, including LDA methods, physical models, data products, as well as scientific significance. This summary aims to provide a skeleton of application of LDA in hydrology research. Through overviewing the milestone progress, the future insight and implication of LDA in hydrology research are expected to be identified.

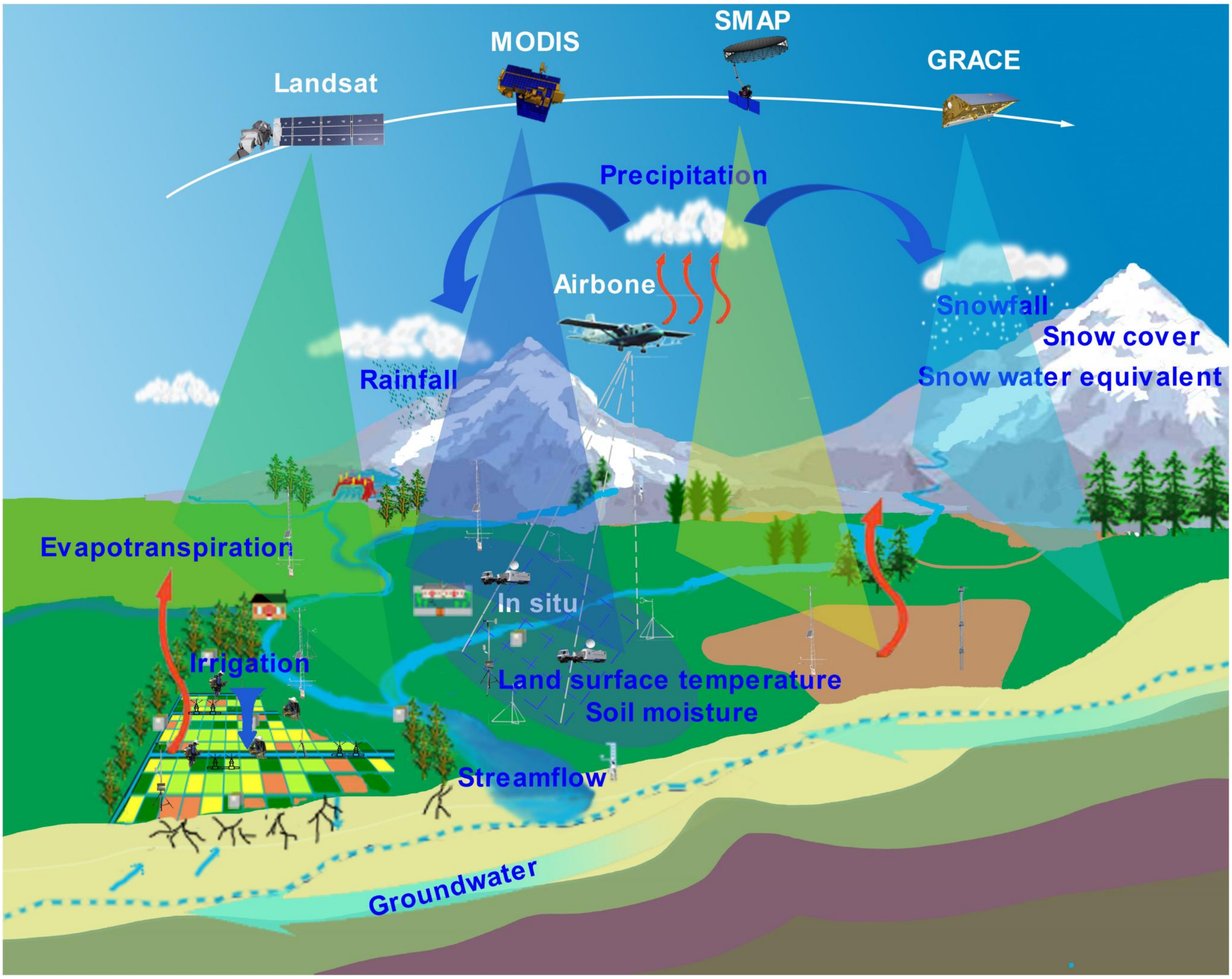


**Figure 5.** Assimilation of multisource and multiplatform observations (satellite, airborne, ground) into land models to improve the predictability of land surface variables (e.g., soil moisture, snow water equivalent/snow depth, groundwater, evapotranspiration, land surface temperature).

### 3.1. Soil Moisture

Soil moisture estimation has typically served as one of the most active applications of LDA. Vice-vera, LDA has become the main-stream approach to generate soil moisture products with spatiotemporal continuity and physical consistency. Although model-data fusion was used in estimating the soil moisture profile in the early 1980s (Jackson et al., 1981), advanced DA methods were introduced in this field in the mid- 1990s. For example, Entekhabi et al. (1994) demonstrated the feasibility of estimating soil moisture and temperature profiles by assimilating remote sensing observations and McLaughlin (1995) noted that LDA could not be fully realized until the availability of extensive sources of observations was developed. The work by Houser et al. (1998) was the first real case to assimilate airborne microwave radiometer observations into the TOPMODEL-based land–atmosphere transfer model based on a four-dimensional DA scheme. This pioneering work paved the way toward booming applications of DA in land surface studies.

Since then, soil moisture DA progressed along the pathway from single to multiple variables, from linear and Gaussian assumptions to nonlinear/non-Gaussian processes, and from 1D assimilation to consideration of spatial correlation and localization. The practice of soil moisture DA has demonstrated the applicability of sequence DA methods and demonstrated the importance of introducing spatial correlation. Given the nonlinearity of the Richards equation, which governs soil moisture flow, the EKF has been applied by developing the adjoints of the Richards equation (Entekhabi et al., 1994; J. P. Walker & Houser, 2001). EnKF is most commonly used because

10.1029/2022RG000801

**Table 1**
*Application of LDA to Key State and Flux Variables of Hydrologic Research*

| Variables | DA methods | Physical models | Widely used observation data sets | Scientific significance |
|---|---|---|---|---|
| Soil Moisture | EKF, EnKF, ETKF, LETKF | Land Surface Models (LSMs), such as the Community Land Model (CLM), Common Land Model (CoLM); hydrologic models, such as Variable Infiltration Capacity (VIC), Soil & Water Assessment Tool (SWAT) | Soil Moisture Active Passive (SMAP) Brightness Temperature (TB) and soil moisture product: (https://nsidc.org/data/smap) | Key state variables controlling partitioning of precipitation between infiltration and runoff |
| Snow-related Variables (SCF, SD/SWE) | Direct-insertion or optimized interpolation methods, Bayesian filtering methods such as EKF, EnKF and its variants, PF, Smoother-based DA schemes such as LETKF | LSMs, such as CoLM, Noah-MP; hydrologic models, such as VIC, SWAT | MODIS SCF and albedo Products (https://nsidc.org), AMSR-E TB (https://gportal.jaxa.jp/gpr/) | Improve the estimation accuracy of land surface water cycle elements |
| Evapotranspiration | EKF, EnKF, and PF | LSM and hydrological models, such as VIC, SWAT | Satellite-based ET estimations at catchment scale using the surface energy balance method such as SEBAL and SEBS | Key component of surface water and energy exchanges; its accurate estimation is important for improving numerical weather prediction, drought and irrigation monitoring |
| Streamflow-related Variables (Streamflow, Inundation map, Water depth/ Water stage) | PF, EKF, EnKF and its variants | Hydrological model, such as WRF-Hydro, LISFLOOD-FP | United States Geological Survey (USGS) gauges (https://earthexplorer.usgs.gov/), Sentinel-1 SAR data (https://search.asf.alaska.edu/) | Key for water resource management and flood prevention measures, such as reducing flood impacts |
| Groundwater-related Variables | EnKF, ETKF | LSMs, such as Community Atmosphere and Biosphere Land Exchange (CABLE); groundwater model, such as MIKE SHE | GRACE https://grace.jpl.nasa.gov/data/get-data/<br>SMOS https://earth.esa.int/eogateway/missions/smos/data<br>SMAP https://nsidc.org/data/smap | Key for improving the estimation of groundwater level, subsurface storage variations, rootzone SM |
| Irrigation | Ensemble smoother<br>Particle batch smoother | LSMs, such as Noah-MP; agro-hydrological models, such as soil, water, atmosphere, and plant (SWAP) | GRACE https://grace.jpl.nasa.gov/data/get-data/SMAP https://nsidc.org/data/smap | Key for maintaining sustainable development of agriculture |
| Land Surface Temperature | Variational DA approach, Ensemble-based filters such as EnKF and EnKS | SEB model, such as SEBS; LSMs, such as CLM | MODIS LST Product: https://ladsweb.modaps.eosdis.nasa.gov/search/ | Key for improving the estimation of soil temperature profiles and/or surface heat fluxes. |

of its strong ability to handle nonlinear problems without developing an adjoint (Dumedah & Walker, 2014; Huang, Li, Lu, & Gu, 2008; Huang et al., 2016; Margulis et al., 2002). Additionally, various localization techniques and ETKF/LETKF were used in soil moisture DA (Wu et al., 2016) to effectively cut off the spurious correlation. More recently, a more advanced particle filter DA was applied to soil moisture estimation and drought monitoring (L. Xu et al., 2020).

Quantifying and reducing uncertainty are some of the most important tasks of soil moisture DA (Moradkhani, Hsu, et al., 2005). In particular, the spatial representativeness error in soil moisture strongly hampers the advancement of soil moisture DA (X. Li, 2014). To address the challenge of quantifying spatial representativeness, a series of large-scale soil moisture experiments have been conducted since 2002 (J. M. Jacobs et al., 2004), such as the Australian Airborne Cal/val Experiments for SMOS (Peischl et al., 2012), SMAPEx (Ye et al., 2021), NAFE (Panciera et al., 2008), and HiWATER (X. Li et al., 2013). Pixel-scale and watershed-scale observations of soil moisture, by capturing heterogeneity, have enabled the quantification of unique representation errors at different scales. In addition, methods have been developed to quantify the scale-associated representativeness errors in soil moisture remote sensing forward modeling and inversion algorithms, which provide prior information for the observation error covariance of soil moisture in LDA.

In addition, LDA has also successfully been used in the downscaling of coarse-resolution soil moisture products, for example, SMOS and SMAP, to generate mid- to high-resolution soil moisture products to improve their usefulness in hydrological applications (Reichle, Entekhabi, & McLaughlin, 2001; Zupanski et al., 2011). Downscaling contains two approaches. The first approach assimilates downscaled remote sensing observations or retrievals (Dumedah et al., 2015; Sahoo et al., 2013), which is easy to implement in the DA process, and uncertainties caused by the scaling issue can be considered. The second approach assimilates coarse-resolution observations into the high-resolution output of land models. This approach avoids a priori observation partitioning and performs downscaling within the EnKF algorithm (Sahoo et al., 2013). However, an assumption of a certain spatial distribution may disturb the spatial coherence of the downscaled soil moisture. This shortcoming might be overcome by machine learning techniques (F. Liu et al., 2020).

Overall, the soil moisture DA has witnessed good progress in terms of LDA methods and downscaling coarser-resolution soil moisture products. In particular, the soil moisture DA method originated from a simple UKF to utilize the advance LETKF to cut off the spurious correlation. This milestone progress provides a tracible indicator for the future progress of soil moisture DA.

### 3.2. Snow-Related Variables

Snow cover is a fundamental component of both the radiation balance and the hydrological cycle of the Earth system. Assimilation of snow-related variables from in situ or remote sensing observations has made great progress in recent years. Variables such as snow cover fraction (SCF), snow cover area (SCA), snow depth (SD), snow water equivalent (SWE), snow albedo, microwave brightness temperature (TB), and terrestrial water storage (TWS) have been assimilated into land models or hydrological models using either simple methods, such as direct insertion, or more sophisticated assimilation schemes based on Bayesian inference to improve the estimation accuracy of land surface water cycle elements.

#### 3.2.1. Evolution of Snow Data Assimilation Methods

The snow DA method has gradually advanced. Simple direct-insertion or optimized interpolation methods correct the mismatches between the simulated snow state variables and snow observations according to physical rules (Roy et al., 2010). The inherent hypothesis of insertion methods is that observations are perfect, while model forecasts have biases, although the model prediction is sometimes more accurate than the observations in practical applications (C. Sun et al., 2004). However, interpolation operations usually violate the water balance within the physical process model (Magnusson et al., 2017). The drawbacks of the insertion DA method can be avoided by using more advanced Bayesian filtering methods (Largeron et al., 2020), such as EKF (C. Sun et al., 2004) and EnKF (Andreadis & Lettenmaier, 2006). The study of EnKF and insertion-based MODIS SCF assimilation showed that the EnKF method slightly improves snow simulation with little violation of the mass balance (Arsenault et al., 2013). Various variants of the EnKF, such as the ensemble square root filter (EnsRF) (Leisenring & Moradkhani, 2011), deterministic ensemble Kalman filter (DEnKF) (T. Xu et al., 2014), hybrid DEnKF-variational data assimilation method (DEnVar) (J. Xu & Shu, 2014) and ensemble adjustment Kalman filter

(EAKF) (Zhao et al., 2016), have also been used in snow DA, and these methods show better efficiency and performance than the EnKF with advantages such as not perturbing the observations. PF has been gaining popularity in snow DA, for example, comparative studies of the EnKF, EnSRF and PF methods have suggested that PF is superior to EnKF in improving the estimation accuracy of snow-related state variables, such as SD, SWE, and streamflow (Leisenring & Moradkhani, 2011). Recently, smoother-based DA schemes, such as LETKF (Oaida et al., 2019), the ensemble batch smoother (Margulis et al., 2015) and the particle batch smoother, which synchronously assimilate both prior and subsequent observations in a given time, have become more appropriate because they consider the seasonal temporal correlation between snow observations and state variables (Margulis et al., 2015).

#### 3.2.2. SCA/SCF Data Assimilation

From the perspective of data, an important branch of snow DA is SCA/SCF DA, owing to the satisfactory accuracy and higher spatial resolution of various optical remote sensing snow cover products. A key issue in SCA/SCF DA is to construct the nonlinear snow depletion curve (SDC), which relates the SCA to the model state (e.g., SD and SWE) as an observation operator. Various forms of SDCs have been proposed, and more consistent snow predictions are obtained by effectively constraining the state variables of the model using SCA (Andreadis & Lettenmaier, 2006; Niu & Yang, 2007; Zaitchik & Rodell, 2009). However, the existing SDCs still have difficulties expressing the relationship between SCA and model state, such as the weak correlation between SCA and model state caused by "saturation" and the lack of a consistent parameterized relationship in heterogeneous terrain environments. Improving the representation of the relationship between SCA and the model state remains a challenge.

#### 3.2.3. SD/SWE Data Assimilation

Assimilation of in situ and/or passive microwave remote sensing-retrieved SD/SWE is another important branch of snow DA, which has been shown to effectively increase the accuracy of snow water storage (i.e., SD, SWE, streamflow) (Kumar et al., 2015; Y. Q. Liu et al., 2013; Magnusson et al., 2014). However, in situ SD/SWE observations are usually sparse and do not match the model grid, and the SD/SWE retrieved by passive microwave remote sensing is trapped by the coarse resolution and large uncertainty of various inversion algorithms. The negative influence of erroneous SD/SWE retrievals can theoretically be avoided by directly assimilating radiance observations, and several studies have witnessed dramatic improvements in model predictability via the assimilation of microwave TBs (Che et al., 2014; Durand et al., 2009; Durand & Margulis, 2006; Huang et al., 2012). Moreover, snow data assimilation has another trend in the simultaneous assimilation of multivariate observation data, such as TB and in situ SD (Takala et al., 2011), SCF and snow albedo (J. Xu & Shu, 2014), SCF and SWE (De Lannoy et al., 2012), and SCF and GRACE TWS (Su et al., 2010). These studies have suggested that joint assimilation of combined multisource snow observations demonstrates a larger benefit by mitigating the limitations of individual sensors.

Overall, snow DA presents many opportunities for improving our understanding of snow systems. The focus and difficulties of future research lie in identifying the most informative sources of snow-related data and establishing new frameworks for snow data assimilation based on dual driven theory-data approaches, revealing multi-scale processes and identifying model deficiencies of snow physics, especially when ground observations of snow are scarce in high-mountain areas.

### 3.3. Evapotranspiration

Evapotranspiration (ET) is a key component of surface water and energy exchanges between land and atmosphere, and an accurate estimation of ET plays an important role in numerical weather prediction, drought and irrigation monitoring. Current research on ET-related assimilation is mainly focused on assimilating state variables, such as soil moisture (SM) and land surface temperature (LST), to indirectly improve ET estimation (Y. Lu et al., 2019; Xu et al., 2019), and only a few attempts have been made to assimilate satellite-based ET products directly (Deb et al., 2022). The latter mainly focuses on assimilating satellite-based ET estimations into a hydrological model, such as VIC and SWAT, at the catchment scale using mature DA methods such as EnKF and PF. The ET products are generally obtained by using the surface energy balance method, such as the Surface Energy Balance Algorithm for Land (SEBAL) and Surface Energy Balance System (SEBS) (Table 1).

A simple way to estimate ET is to update the model-simulated ET to the probabilistically optimal ET values by directly assimilating satellite ET products using mature DA methods, such as EKF, EnKF, and PF (M. Pan et al., 2008). The assimilation effect is limited to ET estimation instead of the whole model performance since ET is not a state variable. Alternatively, satellite-based ET products are treated as observations to calibrate the model parameters that are associated with other state (e.g., SM) or nonstate (e.g., streamflow) variables via the DA approach, which may not guarantee optimization of the hydrological cycles as a whole (Gelsinari et al., 2020; Xiong et al., 2019).

The bottleneck of assimilating ET is further resolved by establishing a relationship between ET and the state variable SM, enabling feedback from assimilating ET to update the model state variable and subsequent fluxes (e.g., sensible heat flux and streamflow) (Pipunic et al., 2008; Zou et al., 2017). One method is to construct an observation operator relating SM with ET that converts ET into a state variable, and then, DA is implemented to update the ET and SM (Pipunic et al., 2008). This method requires a long computational time and may cause a rank-deficient problem in calculating the covariance matrix. The other approach is to build a unit matrix operator to first update the modeled ET based on the DA method, and then, the state variable SM is calculated to adjust other hydrological variables (e.g., streamflow) (Yin et al., 2016; Zou et al., 2017). Although this method resolves the rank-deficient problem and is computationally effective, it demands an explicit time response relationship between ET and SM to realize the feedback of assimilation to the model.

Although the assimilation of ET shows the potential to improve other hydrological variables, the assimilation performance is largely affected by the quality of satellite-based ET products (Pipunic et al., 2008). In addition, it remains a challenge to construct an appropriate DA framework that enables the feedback of ET DA to update the model state variables and optimize the hydrologic cycles as a whole. Joint assimilation with the SM product may improve the assimilation results.

### 3.4. Streamflow

Accurate and reliable streamflow predictions are fundamental for water resource management and flood prevention measures, such as reducing flood impacts. Streamflow DA, which can improve the estimations of the states and parameters of a watershed, plays an important role in the modeling and prediction of rainfall-runoff processes.

Current research on streamflow DA is mainly focused on assimilating surface state variables, such as soil moisture, to indirectly improve streamflow prediction (Visweshwaran et al., 2021). Most researchers use streamflow to correct the initial conditions of watershed-scale hydrological models to improve streamflow prediction. For example, streamflow DA is used to update the states and parameters of watershed-scale models through the EKF, EnKF or PF (Aubert et al., 2003; Clark et al., 2008; Y. R. Fan et al., 2017; L. Sun et al., 2015; Vrugt et al., 2006). Variants of EnKF, for example, EnsRF and recursive EnKF, are also used to assimilate streamflow to update model states (H. Chen et al., 2013; McMillan et al., 2013). DA can also improve the initial conditions for streamflow prediction (DeChant & Moradkhani, 2011, 2014). Localization and inflation techniques can also be applied to improve the efficiency of streamflow DA (Gharamti et al., 2021). The variational data assimilation method and KF are rarely used because hydrological processes are highly nonlinear (Ercolani & Castelli, 2017; Mazrooei et al., 2021).

Flood inundation can be forecasted by assimilating water depth, discharge and flood extent maps. Traditionally, gauge station-observed discharge and water stage data have been assimilated into hydrological models to predict water depths and flood inundation (Cao et al., 2019; Jafarzadegan et al., 2021). Currently, an increasing number of river flood-plain elevation and inundation maps, which are derived from satellite synthetic aperture radar (SAR) images and altimetry data, have been assimilated into hydrological models to forecast flood inundation (Cooper et al., 2019; Dasgupta et al., 2021; Dubey et al., 2021; Matgen et al., 2010). Apart from satellite and gauge stations, observations can be derived from social media platforms; for example, water depth values gathered from volunteered geographic information (VGI) can be assimilated into a 2D hydraulic model for real time flood forecasting (Annis & Nardi, 2019). With real-time SAR-derived flood extents or stage observations, floods can be forecasted in a timely manner based on DA (Hostache et al., 2018; X. Y. Xu et al., 2017).

Additionally, the simultaneous updating of multiple states and parameters has attracted increasing attention in streamflow DA (Piazzi et al., 2021; Prakash & Mishra, 2022; F. Wang et al., 2022). However, spatiotemporal state updating in distributed hydrological models via streamflow DA remains a challenge due to the large dimensional disparity between the model state space and observation space. One solution is to use spatiotemporal variance–covariance from simulated climatology models as a surrogate for model error to spatially distribute constraints from gauge measurements (S. Tian et al., 2022).

In situ discharge data are critical in real-time DA because the temporal resolution of discharge measurements is superior, with hourly data versus the typically much longer revisit times of satellites. However, there have been very few attempts to directly assimilate in situ streamflow observations. For example, streamflow observations were assimilated to update upstream river water storage and improve downstream streamflow forecasting (S. F. Liu et al., 2012; Seo et al., 2021; Y. Zhang et al., 2017). In these applications, the densities of observations upstream dramatically affect the DA performance.

Assimilating remotely sensed observations of water levels and water-inundated areas is a promising direction. In particular, the Surface Water and Ocean Topography (SWOT) satellite, which was launched in 2022, provides high-resolution observations of the water body height and area and can be directly assimilated into hydrological models to improve the accuracy of streamflow estimation globally (D. Y. Li et al., 2020; Wongchuig-Correa et al., 2020). Additionally, another promising direction in streamflow DA is the combination of DA and ML (as introduced in Section 2.6) (Boucher et al., 2020; Quilty et al., 2022).

### 3.5. Groundwater-Related Variables

Assimilating remote sensing or other observations into a groundwater model is important for improving estimations of groundwater level (X. Li et al., 2020), subsurface storage variations (Khaki et al., 2018), and root zone soil moisture (H. Zhang et al., 2018). For example, soil moisture products from SMOS and SMAP are assimilated into land models to improve the estimation of groundwater (Tangdamrongsub et al., 2020). Soil moisture and GRACE data are assimilated simultaneously to estimate both surface water and groundwater (D. Zhang et al., 2016).

Methodologically, groundwater DA has special features, that is, the traditional EnKF cannot properly handle strongly skewed pressure distributions, which are caused by extremely negative pressure heads in the unsaturated zone during dry periods (Panzeri et al., 2014). A feasible solution is to simultaneously assimilate the soil moisture and observed ground water head into coupled stochastic moment equations with the EnKF or ETKF using localization (Panzeri et al., 2014; H. Zhang et al., 2018) as this process can effectively correct the skewness of the negative pressure heads. However, solute transport, when coupled with groundwater flow, requires the hydraulic conductivity and initial condition of solute sources to be simultaneously estimated (Tong et al., 2012). Therefore, a DA scheme that simultaneously updates the parameter and state for improving solute transport prediction is expected to be an effective solution.

### 3.6. Irrigation

To efficiently use and avoid unnecessary losses of fresh water to maintain sustainable development of agriculture under a changing climate system, the estimation of irrigation water use and optimal irrigation management are indispensable. Different from other hydrological state variables (e.g., soil moisture), irrigation can be regarded as a "flux" variable that is seldom assimilated or estimated directly. LDA has shown satisfactory performance in the estimation and management of irrigation by assimilating other variables (e.g., soil moisture) into land and hydrological models.

Given that irrigation is not a state variable in land and hydrological models, the general scheme involves improving the estimation of soil moisture (Abolafia-Rosenzweig et al., 2019), ET (Irmak & Kamble, 2009), and hydraulic parameters by assimilating remote sensing or in situ observations and then using the assimilated variables and parameters in optimal irrigation management via agro-hydrological models (Irmak & Kamble, 2009; L. Wang et al., 2009). This trend has evolved into using the above LDA scheme in real-time control of irrigation (X. J. Han et al., 2016) and assimilation of high-resolution remote sensing (~30 m) data (F. N. Lei et al., 2020) so that the assimilated information can be more useful in precision agriculture.

DA also has the potential to improve the estimation of irrigation magnitude. Synthetic experiments show that the smoothing DA algorithms, for example, the particle batch smoother, outperform the DA approach that assimilates a single point in a time window, thus improving the daily to seasonal irrigation magnitude by assimilating remote sensing soil moisture data (Abolafia-Rosenzweig et al., 2019). The combination of GRACE assimilation and irrigation process simulation in the Noah-MP model has demonstrated its effectiveness in improving the estimation of anthropogenic water withdrawals and irrigation over the USA High Plains Aquifer (Nie et al., 2019). However, DA for irrigation magnitude at a large scale has not been used for practical applications. The estimation of irrigation water use at continental, regional, and river basin scales is a promising research direction.

### 3.7. Land Surface Temperature

LST is a key state variable that determines the upward longwave radiation and influences the turbulent sensible and latent heat fluxes to the atmosphere as well as the ground heat flux into the soil column. Therefore, LST DA has attracted much research interest because assimilating LST can improve the estimation of soil temperature profiles and/or surface heat fluxes. The variational DA approach and ensemble-based filters are most often adopted to assimilate the LST into either the simple surface energy balance (SEB) equation or complex land surface models (LSMs), whereby the MODIS LST products are most widely used (Table 1).

One major development path entails a variational DA approach for assimilating LST using a simple SEB equation (e.g., force-restore equation or heat diffusion equation) or complex LSMs as the physical constraint. Numerous studies have assimilated LST measurements into the simple force-restore equation based on the variational DA framework for estimating surface heat fluxes (Sini et al., 2008). This approach has been further advanced by using the full heat diffusion equation as a constraint (Bateni & Entekhabi, 2012; T. Xu et al., 2014, 2019). A more sophisticated approach involves the use of LSM, such as the Common Land Model and ORCHIDEE LSM, as a physical constraint, in which the adjoint of the LSM needs to be developed (Benavides Pinjosovsky et al., 2017; X. L. He et al., 2020; Meng et al., 2009).

The other path for assimilating LST is based on ensemble-based filters, such as EnKF and EnKS. In one simple process, LST is assimilated into the SEB model with the heat diffusion equation using the EnKF or EnKS method to overcome the shortcomings of the variational DA approach (X. L. He et al., 2019). In comparison to the variational DA method, the ensemble-based algorithm has the following advantages: (a) it is easy to implement and does not need an adjoint model, and (b) it can provide both model estimates and their uncertainties. Accordingly, a more sophisticated approach has been developed to assimilate LST into complex LSMs considering the inherent coupling between water and energy using the EnKF or EnKS method for point- and regional-scale applications (Huang, Li, & Lu, 2008; Reichle et al., 2010). Simultaneous optimization of state variables (i.e., LST and SM) and model parameters that affect the surface heat fluxes can be determined based on the EnKF or EnKS method.

Although LST assimilation improves the estimation of the soil temperature profile and surface heat fluxes, two limitations have been demonstrated: (a) this process performs poorly under dense vegetation or wet soil conditions, and (b) the satellite-derived LST products are often contaminated by clouds that reduce data availability. A promising direction is the joint assimilation of LST and SM or brightness temperature measurements, which can overcome the above deficiencies (Y. Lu et al., 2017, 2019).

### 3.8. Multivariable Data Assimilation

Multivariable LDA refers to the simultaneous assimilation of at least two types of observations to update model state variables or parameters. The availability of simultaneous multisource observation pairs (even if these observation pairs are not captured at the same moment, it is critical that they are captured within a certain time window) is an important characteristic of multivariable LDA (Montzka et al., 2012). With the development of comprehensive space-air-ground Earth observation systems, multisource and multiscale Earth observation data are becoming increasingly abundant and have formed complementary advantages in temporal and spatial resolution.

However, the potential value of these massive observation data and their combinations has not been fully exploited in LDA compared to univariate LDA. To date, a large number of multivariable LDA schemes have been designed to consistently correct terrestrial system states and parameters by jointly assimilating multiple types of data, for example, joint assimilation of TB and ET (M. Pan et al., 2008), TB, LST and LAI (W. J. Chen et al., 2021), SM and LAI (Sabater et al., 2008), SM and ET (Gavahi et al., 2020), SM and LST (Huang et al., 2016), and SM and streamflow (Abbaszadeh et al., 2020), into LSM to improve water, energy and flux estimates (Barrett & Renzullo, 2009; W. J. Chen et al., 2021). SM, pressure head, and streamflow have been assimilated into hydrological models to improve subsurface states and river discharge estimates (Botto et al., 2018; D. Zhang et al., 2016). SM measurements from the SMOS and SMAP and TWS information from GRACE have been assimilated into LSM to improve regional SM and groundwater storage estimates (Tangdamrongsub et al., 2020). Joint assimilation of streamflow and actual ET observations has been performed to estimate the time-varying parameters of a hydrological model (Xiong et al., 2019). The combined assimilation of several snow-related variables (e.g., SCF, SD, SWE, LST, albedo, snow grain size, and snow density) has been conducted to improve snow cover modeling accuracy (Charrois et al., 2016; De Lannoy et al., 2012; Durand & Margulis, 2008; Piazzi et al., 2018). The most straightforward technique used in multivariable LDA is the augmented state-vector approach, which is commonly applied with EnKF and its variants, and PF.

Land surface process simulations could benefit from multivariate DA with respect to univariate DA. However, the performance of LDA results does not parallel improvements with the number of observed data types (Charrois et al., 2016). Undesirable worsening of the system performance has been observed for the combined assimilation of multiple variables, such as integrating additional observations with large uncertainties into hydrological models (Y. Zhang et al., 2017). Although multivariate LDA seems to be a relatively straightforward extension of univariate LDA, the complexity of multivariate LDA does not lie in an algorithmic nature but is mainly related to the following aspects: (a) it is difficult to assess which types of observation data are the most effective, and it is also increasingly difficult to determine the possible tradeoffs that might occur (Botto et al., 2018). Thus, specifying appropriate uncertainties for different types of measurement data is a major challenge for multivariable LDA. (b) Multivariable LDA often suggests a multiscale problem. How to coordinate the temporal and spatial scale discrepancies between different types of observations, as well as scale discrepancies between observations and model simulations, is also a major challenge. (c) Another incidental fact in multivariate DA is that the spurious correlation across different types of variables is typically more pronounced (Popov & Sandu, 2019; D. Zhang et al., 2016). Although some localization techniques alleviate spurious correlations, further efforts are needed to properly assign the degree of localization for each type of observation variable and account for the correlation changes spatiotemporally.

## 4. Development of Global, Regional, and Catchment-Scale Land Data Assimilation Systems

### 4.1. Global and Regional-Scale Land Data Assimilation Systems

The land data assimilation system (LDAS) is a system that exploits the methodological advances introduced in Section 2 to assimilate multisource observations into land models, providing optimized estimations of land surface states and fluxes. LDAS can be used in forecasting, nowcasting or reanalysis. Currently, most LDASs focus on producing long-term data sets, which manifest as the development of various global- and regional-scale LDASs mainly using the Fortran programming language (Baatz et al., 2021; Xia et al., 2019), including the North American LDAS (NLDAS) (Mitchell et al., 2004; Xia et al., 2012), Global LDAS (GLDAS) (Rodell et al., 2004), and European LDAS (ELDAS) (C. M. J. Jacobs et al., 2008; Van Den Hurk, 2002), as well as LDASs developed for Asia (LDAS developed by China Meteorological Administration (CMA)) (C. Shi et al., 2011), China (Chinese LDAS (CLDAS)) (X. Li et al., 2007), Canada (Coupled Land and Atmosphere Data Assimilation System (CaLDAS)) (Carrera et al., 2015), South America (SALDAS) (de Goncalves et al., 2006), South Korea (KLDAS) (Lim et al., 2012), and Africa (McNally et al., 2017). These LDASs are typically run offline (i.e., they are not coupled with an atmospheric model) using the best-available meteorological forcing and are different from reanalysis systems, such as ERA5, which are produced by weakly coupled land–atmosphere DA systems (Hersbach et al., 2020). Notably, most of the abovementioned LDASs only assimilate limited observations of land surface variables or just perform multimodel ensemble simulations. The basic information on global and regional-scale LDASs is summarized in Table 2; here, we introduce some representative LDASs.

#### 4.1.1. NLDAS

The NLDAS is a multi-institution effort dedicated to providing high-quality, spatially and temporally consistent land surface data sets (Mitchell et al., 2004; Xia et al., 2012). Currently, NLDAS has been upgraded from the first phase (NLDAS-1) to the second phase (NLDAS-2), which runs hourly from January 1979 with a spatial resolution of 1/8th-degree for North America. NLDAS-2 produces a high-quality meteorological forcing data set from

10.1029/2022RG000801

**Table 2**
*Overview of Global and Regional-Scale Land Data Assimilation Systems (LDASs)*

| LDAS | Spatial extent | Time period | Spatial resolution | Temporal resolution | Meteorological forcing | LSMs | Assimilation method |
|---|---|---|---|---|---|---|---|
| NLDAS | North America | 1979-present | 0.125-degree | hourly | Merged NARR meteorology analysis, satellite- and ground-based precipitation and radiation data | Noah, Mosaic, VIC, SAC | 4D Var, EnKF |
| GLDAS | Global | 1979-present; 1948–2010 | 0.25-degree; 1-degree | 3-hourly; monthly | Merged GDAS meteorology analysis, satellite- and ground-based precipitation and radiation data | Noah, Mosaic, VIC, CLM | 4D Var, EnKF |
| ELDAS | Europe | 2000 | 0.2-degree | 3-hourly | NWP outputs; Merged ERA-40 meteorology reanalysis, satellite- and ground-based precipitation and radiation data | ISBA, TERRA, TESSEL | 4D Var, OI |
| CMA LDAS | Asia | 2017-present | 0.0625 -degree | hourly | Merged gauge observations from 2400 national meteorological stations with FY2 satellite measurements, and ECMWF and GFS meteorology analysis | CLM, CoLM, Noah-MP | – |
| CLDAS | China | 1979-present | 0.1-degree | 3-hourly | Merged ground-based observations with several remote sensing and reanalysis data sets, such as China Meteorological Administration measurements, GLDAS forcing data set, MERRA pressure data set, TRMM satellite-observed precipitation, and Global Energy and Water Exchanges/Surface Radiation Budget radiation product | CoLM | EnKF |
| CaLDAS | North America | – | – | 6-hourly | Merged NWP outputs and satellite- and ground-based precipitation | ISBA | OI, EnKF |
| SALDAS | South America | 2000–2004 | 0.125-degree | 3-hourly | Merged South American meteorology reanalysis and satellite- and ground-based precipitation and radiation data | SSiB | – |
| KLDAS | East Asia | 2004–2008 | 10-km | hourly | Merged Korean GDAPS meteorology analysis and satellite- and ground-based precipitation and radiation data | Noah | 3D Var |

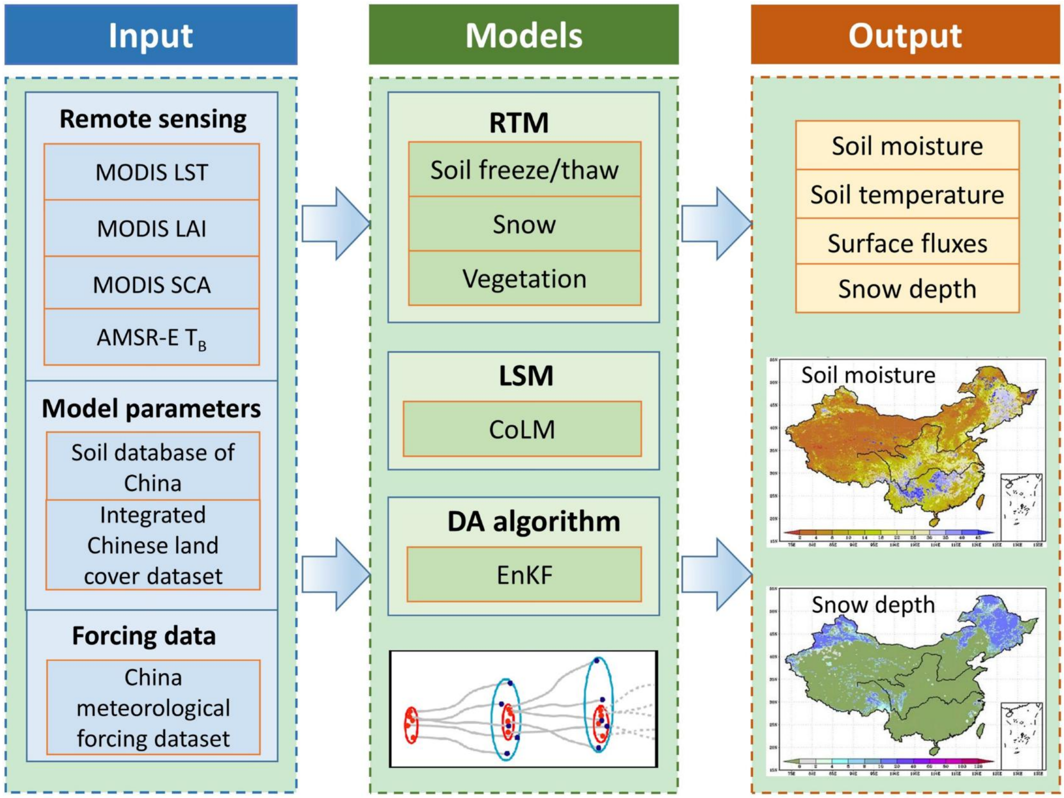


**Figure 6.** Framework of the Chinese land data assimilation system (CLDAS).

multisource ground- and satellite-based products/analysis and reanalysis data (Table 2). There are currently four LSMs, that is, Noah 2.8, Mosaic, VIC 4.0.3, and Sacramento Soil Moisture Accounting model, implemented by the NLDAS-2 to generate hourly model outputs. The NLDAS-2 project does not include the assimilation of satellite-based observations of land surface variables; however, efforts have been made to assimilate remote sensing retrievals into NLDAS-2 (Kumar et al., 2012, 2016).

### 4.1.2. GLDAS

The GLDAS led by NASA/GSFC also aims to use satellite- and ground-based observations with advanced LSMs and DA methods to produce optimal estimates of land surface states and fluxes worldwide (Rodell et al., 2004). Two meteorological forcing data sets are adopted by GLDAS: one forcing is produced by the combination of satellite- and ground-based products and reanalysis data available from 1979, and the other forcing is the Global Meteorological Forcing Data set produced by Princeton University available from 1948 to 2010 (Table 2). The two forcing data sets are used to drive four LSMs, that is, Noah, Mosaic, VIC and CLM2, to generate 3-hourly and monthly model outputs at spatial resolutions of 0.25° (Noah only) and 1° (all other LSMs). MODIS snow cover has been assimilated into GLDAS to constrain the estimation of snow water equivalent, and assimilation of other variables has also been implemented as part of a follow-up project (Peters-Lidard et al., 2011).

### 4.1.3. CLDAS

The CLDAS aims to produce high-resolution, spatiotemporally consistent land surface data sets for mainland China by assimilating various remote sensing observations (Figure 6) (Huang et al., 2016; X. Li et al., 2007). The CoLM (Dai et al., 2003) is used to simulate land surface processes, the EnKF algorithm is implemented as a DA method, and the radiative transfer models (RTM) of snow, vegetation, and thawed and frozen soils are used as observation operators. The data assimilated into CLDAS are mainly remote sensing observations of land surface variables, that is, passive microwave remote sensing data, including SSM/I, TMI and AMSR-E brightness temperature (TB) products, and optical remote sensing data, such as MODIS LST, LAI and SCA products. Inputs

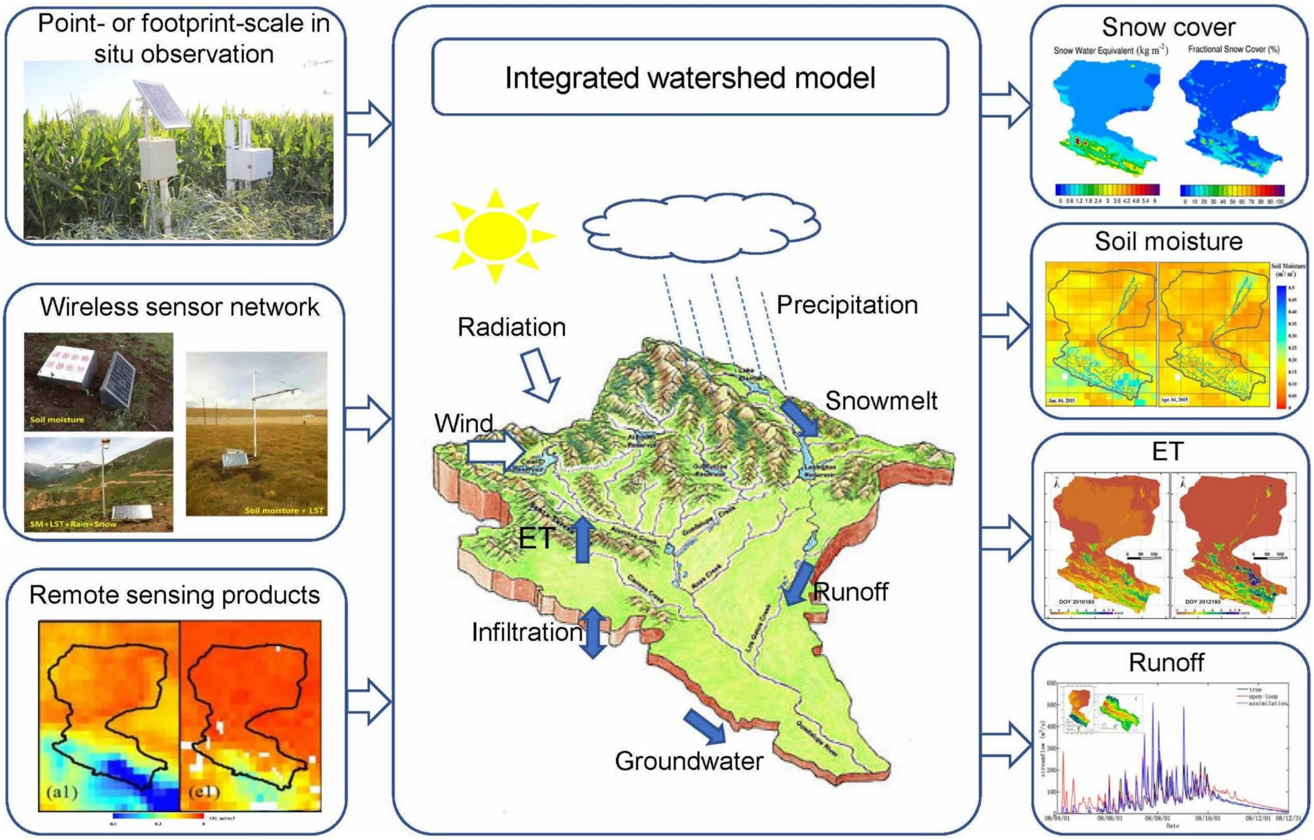


Figure 7. Conceptual scheme of catchment-scale LDAS.

include the China Meteorological Forcing Data set (CMFD) (J. He et al., 2020) and parameter sets for CoLM, such as soil texture derived from the Soil Database of China (Shangguan et al., 2012) and land cover taken from the Multisource Integrated Chinese Land Cover (MICLC) data set (Ran et al., 2012).

The long-term data sets of meteorological forcing, surface water and energy fluxes/variables produced by the abovementioned global and regional-scale LDASs have been widely adopted in various hydrological and meteorological applications (Xia et al., 2019). For instance, the GLDAS data set has been implemented to estimate groundwater depletion in India (Rodell et al., 2009), produce the ESA CCI soil moisture product (Dorigo et al., 2017), and quantify the global water cycle (Rodell et al., 2015) and freshwater availability (Rodell et al., 2018). The NLDAS data set has been widely used in drought (Xia et al., 2014) and flood (Alipour et al., 2020; David et al., 2013) monitoring, crop and water resource management (Dhungel et al., 2020), and numerical weather forecasting and climate prediction (Case et al., 2011; Saha et al., 2014) at the continental scale. Overall, the long-term data sets generated by those LDASs have greatly advanced our understanding of the global/regional-scale water cycle and energy budgets, especially in ungauged regions, which has provided substantial support to research scientists, operational forecasters, and decision makers.

### 4.2. Catchment-Scale Land Data Assimilation Systems

The catchment is a basic unit of land surface system and water resource management. A catchment (or river basin or watershed) encompasses more prominent complexity and heterogeneity due to many processes, including three-dimensional groundwater dynamics, which are not considered in global-scale models and must be presented in the catchment-scale model. Therefore, there is a strong need to extend the applicability of LDA from global and regional scales to the catchment scale (Figure 7), with a focus on improving the predictability of the hydrological cycle to support water resource management (Troch & Paniconi, 2003).

The most remarkable characteristic of catchment-scale LDA, which differs from large-scale LDA, is that a distributed hydrological model coupled with a groundwater model can be used, which can further integrate other important processes, such as dynamic vegetation growth models and socioeconomic models. Therefore, catchment-scale LDA may involve three-dimensional dynamics, which are rare in regional- and global-scale LDA. Additionally, the computational cost of catchment-scale LDA is generally higher, and the computationally tractable DA framework is also critical. Most importantly, high resolution (usually meters to tens of kilometers horizontally, 0.01∼1 m vertically) is a main advantage of catchment-scale LDA and has substantial practical value. For instance, (F. N. Lei et al., 2020) assimilated high-resolution thermal and radar remote sensing retrievals for soil moisture monitoring in a drip-irrigated vineyard.

To date, most catchment-scale LDAs have been conducted one dimensionally, either using one- or three-dimensional models. The focus of catchment-scale LDA has been on using simple conceptual models (Aubert et al., 2003; H. Chen et al., 2013; Pathiraja et al., 2016) and distributed hydrological models (Clark et al., 2008; Kim et al., 2021; Pauwels et al., 2001; C. Qin et al., 2008; Xie & Zhang, 2010) to assimilate surface soil moisture, streamflow, and snow data individually or jointly. A good example of catchment-scale LDA is the Soil & Water Assessment Tool (SWAT)-based hydrological DA system (SWAT-HDAS), which was developed to assimilate in situ or remotely sensed multisource observations, including SM, SWE and discharge, into a parallelized SWAT model for basin-scale hydrological predictions (Y. Zhang et al., 2017). In addition, three-dimensional models, for example, the process-based numerical model of coupled surface and subsurface flow (Hurkmans et al., 2006), and more complex distributed hydrological models, including lateral flows and surface water–groundwater interactions, for example, MIKE SHE (Boegh et al., 2004; Ridler et al., 2014; D. Zhang et al., 2016), have been used to assess the performance of DA algorithms. Combined streamflow and pressure head assimilations within the framework of the three-dimensional Richards equation are only used for test cases and not for real-world three-dimensional catchment data assimilations (Camporese et al., 2009). The parallel data assimilation framework (PDAF)-based catchment-scale LDAS, which fully couples the Consortium for Small-scale Modeling (COSMO) atmospheric model, the CLM, and a subsurface model (ParFlow), is a real three-dimensional catchment-scale LDAS (Kurtz et al., 2016; H. Zhang et al., 2018). This system offers the possibility to perform state and joint state–parameter estimations at the surface–subsurface part of the land system. Notably, catchment-scale LDAS that realizes three-dimensional data assimilation remains rare.

Advanced DA frameworks capable of assimilating multiscale observations and describing multiscale spatial heterogeneity should be developed in catchment-scale LDA. Multisource remote sensing observations, including SM, SD/SWE, SCA, ET, LAI, streamflow and surface freeze–thaw state, and some new types of data, for example, wireless sensor network data and COSMOS data, can be assimilated into integrated watershed models to generate high-resolution and consistent estimations of the main components of the terrestrial water cycle at the catchment scale (Farhadi et al., 2015; Gichamo & Tarboton, 2019; X. J. Han et al., 2014; C. Qin et al., 2008; Xie & Zhang, 2010; D. Zhang et al., 2016; Y. Zhang et al., 2017). Additionally, land processes can be assessed by catchment-scale LDAS. E. Han et al. (2012a) investigated how surface soil moisture data assimilation affects hydrologic simulations and how spatially varying inputs affect the potential capability of surface soil moisture assimilation at the watershed scale.

Methodologically, some advanced assimilation algorithms have been used for catchment-scale LDA, such as the integrated EnKF and PF, which can be developed to adapt nonlinear and non-Gaussian systems for catchment-scale hydrological data assimilation (Y. R. Fan et al., 2017). The evolutionary DA method was introduced to evolve ensembles of model states and parameter sets while adaptively estimating the model error for real-time forecasting operations (Dumedah & Coulibaly, 2013a). The recursive EnKF method has been developed to assimilate streamflow data in an operational flow forecasting system (McMillan et al., 2013). The nonstationary characteristics of the parameters were also considered, and the time-varying model structures have certain potential to reduce systematic error and improve the hydrological predictions with changing land surface conditions (Pathiraja, Moradkhani, et al., 2018; Pathiraja et al., 2016, 2018a, 2018b). The high-dimensional DA method can improve the performance of catchment-scale LDA. For instance, a three-dimensional EnKF has been used to assimilate AMSR-E coarse grid (25 km) soil moisture retrievals into the Noah land surface model for fine-scale (1 km) surface soil moisture estimation over the Little River Experimental Watershed, Georgia, USA (Sahoo et al., 2013). Some rescaling techniques, for example, linear regression and cumulative distribution function matching, were beneficial for removing the systematic deviations of catchment-scale LDA (Alvarez-Garreton et al., 2014; Loizu et al., 2018). The spatial patterns of observations can provide more useful information to update horizontal variables (e.g., lateral flow, streamflow) for catchment-scale LDA (Grayson et al., 2002; Pauwels et al., 2001). Multiscale DA has been an essential tool to assimilate various satellite-based and ground-based observations at the watershed scale (Montzka et al., 2012). In addition, high-performance computational technologies, such as PDAF and OpenDA (open data assimilation library) (Ridler et al., 2014), have been used to address the heavy computational requirements of catchment-scale LDA (Ridler et al., 2014; Y. Zhang et al., 2017; H. Zhang et al., 2018).

High spatiotemporal resolution land products are usually produced by catchment-scale LDAS and have been applied to drought and flood monitoring (Kim et al., 2021) and groundwater and agricultural irrigation water management (Kurtz et al., 2017), effectively improving the catchment-scale hydrological forecasting ability and water resource management level. For example, a GRACE terrestrial water storage product is assimilated into NASA's catchment land surface model (CLSM) at the global scale, generating groundwater storage time series that are useful for drought monitoring (B. L. Li et al., 2019). Agricultural water resources are properly managed before drought events are expended through soil moisture DA to detect and attribute drought-induced changes in hydrological behaviors at the catchment scale (Y. H. Liu et al., 2021). Flood extent representation is improved by assimilating synthetic aperture radar (SAR) images into hydrodynamic models with Telemac2D-EnKF (Nguyen et al., 2022). The opportunities presented by catchment-scale LDAS create an area of research with significant potential to inform decision-making processes related to water management and climate change adaptation.

## 5. Software and High-Performance Computing

A typical trend of DA software is providing flexible and "plug and play" tools for multidisciplinary applications for land surfaces. Data Assimilation Research Testbed (DART) (J. Anderson et al., 2009) is the first open-source DA tool and has been widely used in a large range of fields, including atmosphere, thermosphere and ionosphere, as well as land and hydrology studies. OpenDA is a more flexible framework that provides the open interface standard and model library for DA development. Both DART and OpenDA have been used in LDA experiments, such as snowpack estimates (Y.-F.Zhang et al., 2014) or system development (Poterjoy, 2016). The land information system (LIS) was designed specifically for LDA by integrating numerous land surface models and satellite and ground-based observations (Kumar et al., 2006). Applications based on the LIS include GLDAS (Rodell et al., 2004), the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) (McNally et al., 2017), and the National Climate Assessment LDAS (Kumar et al., 2019).

DA software was customized for multidisciplinary LDA applications. Most of them utilize specific land surface models. For example, Daspy (Han et al., 2015) and SWAT-HDAS (Y. Zhang et al., 2017) were designed for the Community Land Model (CLM) and SWAT, respectively. TerrSysMP-PDAF (Kurtz et al., 2016) provides the DA framework for the land surface–subsurface-atmosphere. ComDA (F. Liu et al., 2020) integrates various assimilation algorithms, multiple land models, and observation operators for multidisciplinary DA research. The most popular language among them is Fortran, while OpenDA, Daspy, and ComDA are written in Java, python, and C++, respectively. Other software integrate uncertainty analysis, parameter estimation, and sensitivity analysis toolkits for land models (UQ-PyL) (W. Wang et al., 2016) to address the problem that land models are highly nonlinear and sensitive to parameters.

High-performance computing is increasingly appealing to LDA to address the enormous amount of computing time and ensembles needed. Many software programs, such as Daspy, LIS, ComDA, and SWAT-HDAS, support parallel computing. The PDAF (Nerger & Hiller, 2013) specifically highlights its parallel performance. These software's parallel strategy is usually algorithm-based, such as parallel EnKF (Houtekamer et al., 2014) and LETKF. These software have supported the development of high-performance LDA frameworks, including GLDAS and TerrSysMP-PDAF (Kurtz et al., 2016). With the advent of GPU computing, the Compute Unified Device Architecture (CUDA) was introduced (De Luca et al., 2021) for LDA. An experiment (Quinn & Abarbanel, 2011) shows that a parallel speedup factor of approximately 300 has been achieved when using GPU computing. In the big Earth data era, data assimilation is needed to integrate ultrahigh-resolution numerical models and observations (Miyoshi et al., 2016) and enable the implementation of many ensemble members (more than 10 thousand, Miyoshi et al., 2014). High-performance computing is urgently needed to reduce the computational cost in these big data assimilation systems. Currently, the challenges of high-performance computing for LDA include the redesign of parallel computation schemes, construction of parallel architecture

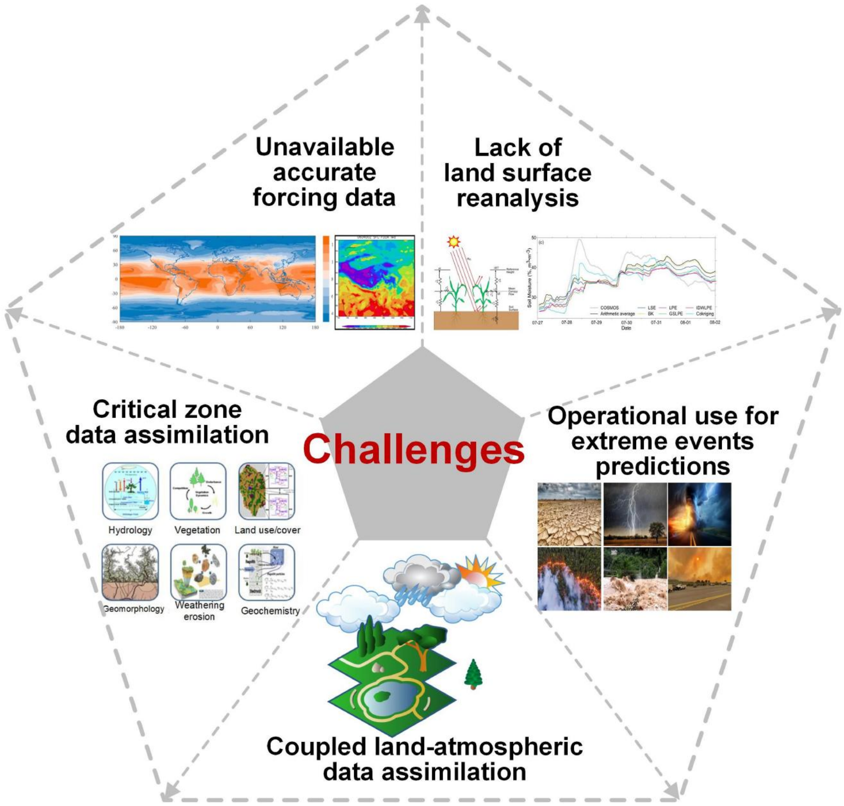


**Figure 8.** The challenges of LDA in forcing data, reanalysis, operational application, coupled land–atmosphere DA, and critical zone DA.

supercomputer systems, new software infrastructure focusing on flexible workflows (Bauer et al., 2021), and cloud computing with a more scalable and adaptable architecture for LDAS.

The future directions of LDA software include developing the scalability of the algorithms to integrate data-driven approaches and designing a better strategy to assimilate nontraditional scientific big data sets, such as those from social media. Additionally, by offering user-friendly interfaces, visualization and analysis tools, LDA software will be more accessible to the community.

## 6. Challenges and Opportunities

### 6.1. Challenges

LDA faces grand challenges, including increasing the reliability and resolution of near-surface atmospheric forcing, developing long time series land reanalysis by truly assimilating numerous land observations, using LDA in operational applications, developing coupled land-atmospheric DA systems, and expanding LDA from the land surface to the subsurface, that is, the critical zone (Figure 8).

#### 6.1.1. High-Quality Forcing for Land Data Assimilation Systems

Forcing is one of the most important components in LDAS that greatly impacts the ability of land models to produce reliable and robust predictions (Cosgrove et al., 2003). However, generating high-quality and high-resolution forcing remains challenging due to deficiencies in quality control, bias correction, and spatiotemporal downscaling. Most existing forcing data, particularly precipitation and solar radiation, have a spatial resolution of 25 km or are even coarser (Rodell et al., 2004; Sheffield et al., 2006), hindering their applications at regional and catchment scales. Furthermore, quality control and bias correction require systematic refinement; for instance, the GLDAS forcing has serious discontinuities and larger errors in precipitation and temperature and failed to capture the spatiotemporal variability in some regions (F. Yang et al., 2017). Similarly, the NLDAS retrospective forcing (1996–2002) was reported to have large biases in precipitation and solar radiation forcings (Luo et al., 2003; Xia et al., 2019). Thus, a potential direction of future research should emphasize the development of long time series, high spatiotemporal resolution, and high-quality forcing data (X. L. He et al., 2020). In particular, the development of high-quality forcing data for complex terrain and areas with extremely sparse observations is the main challenge (Jiang et al., 2022).

#### 6.1.2. New Land Reanalysis Based on the Land Data Assimilation System

Atmospheric reanalysis is available globally, however, land surface reanalysis is lagging behind. Currently, either the so-called land surface reanalysis, for example, ERA-interim/Land (Balsamo et al., 2015), ERA5-Land (Muñoz-Sabater et al., 2021), and MERRA-2 land (Reichle et al., 2017), or outputs from most of the LDASs (see Section 4), have not widely used advanced DA methods or operationally assimilated massive volumes of observational data. Therefore, the development of new land reanalysis by truly assimilating numerous land observations remains challenging.

The difficulties in producing new land reanalysis mainly originate from the following factors: the bias in the LDAS and the lack of long-term high-quality land observations to validate the LADS. The propagation of biased information, for example, errors associated with land models and observation data, through LDAS may lead to problems in closing the water and energy budgets that cause geophysical inconsistencies in land reanalysis (Lahoz & Menard, 2010). Land reanalysis is also highly sensitive to the land observations used to validate the LDAS, while currently, the quality control and consistency checks of land observations are not as adequate as those of atmospheric observations. Taking one of the most widely used land surface variables, for example, soil moisture, as an example, the International Soil Moisture Network (Dorigo et al., 2021) has effectively improved the scarcity of in situ observations. However, its quality control level still lags behind that of atmospheric observations, mainly because the data come from different organizations, observation instruments, and observation depths. The error correction methods are different among different organizations and even in different observation areas within a single organization. There is no unified standard specification. In addition, remote sensing products derived from various platforms also present some unique challenges, requiring cross intercalibration, reprocessing and bias correction (Bosilovich et al., 2013). The changing technical specifications of observation systems lead to difficulties in accurately quantifying the uncertainty characteristics of satellite data to be assimilated (Baatz et al., 2021).

#### 6.1.3. Land Data Assimilation for Operational Use

One purpose of LDAS is to support hydrological and ecological operational applications and management. It is now possible to operationally use the DA technique in decision making and environmental management (Doherty & Moore, 2020), flood prediction and mitigation (Emerton et al., 2016), drought monitoring and prediction (Sawada & Koike, 2016), and operational irrigation scheduling (D. Z. Li et al., 2018).

The challenges in operational use are to enhance the stability and reliability of LDA and improve the predictions of extreme events, such as floods and droughts. Extreme events rarely occur and seldom repeat in both real-world and physical models (Trenberth et al., 2015). Therefore, a large ensemble is needed to capture extreme events, however, its application in operational LDA is hampered by the vast computation. Additionally, the observations of extreme events are sporadic, characterized by skewness and heteroscedasticity in their probabilistic distribution. However, quality control, bias correction, and error quantification of observational data have not been performed operationally in LDA, hindering the operational application of LDA in flood and drought prediction (Y. Liu et al., 2012).

#### 6.1.4. Coupled Land-Atmospheric Data Assimilation

The development of strongly coupled land–atmosphere DA systems is still at the early stage due to challenges in handling the mismatch of spatial-temporal scales between land and atmospheric processes and computational limitations. There are ongoing efforts to couple the LDAS with the atmospheric DA system to improve its coupled land–atmosphere forecast skills (C. Draper & Reichle, 2019; Shahabadi et al., 2019). Efforts have mainly been focused on optimizing the initial soil moisture state by assimilating the observations of air temperature and relative humidity at the 2-m height or/and satellite observations of surface soil moisture. Rather, a coupled land and atmosphere data assimilation system developed by Rasmy et al. (2012) directly assimilates both low- and high-frequency passive microwave brightness temperatures to improve representations of initial land surface and atmospheric conditions, which is more feasible for near real-time applications. Recently, de Rosnay et al. (2022) summarized the coupled DA activities at the European Centre for Medium-Range Weather Forecasts (ECMWF) and presented developments in land–atmosphere forward operator coupling that show the potential for coupled radiative transfer modeling at low-frequency passive microwaves. In the above cases, the LDAS is generally weakly coupled into the atmospheric DA system, in which the land and atmospheric assimilations are performed independently, and their relative influences are added back to the coupled model system in the subsequent forecast cycle. P. Shi et al. (2022) demonstrated the contributions of weakly coupled DA in improving the interannual predictability of summer climate over Europe. Weakly coupled land–atmosphere DA systems have been implemented operationally at some NWP centers (Shahabadi et al., 2019) to produce reanalysis data, such as ERA5 (Hersbach et al., 2020); however, the assimilation of observations in one system (e.g., land DA systems) cannot directly influence the analysis of the other system (e.g., atmospheric DA systems). Although the implementation of strongly coupled land–atmosphere DA systems can address the above deficiencies, it is a relatively new area of research that presents significant challenges, such as more costly computation and difficulty in integrating land and atmospheric processes at different spatial-temporal scales (S. Penny & Hamill, 2017; Shahabadi et al., 2019).

#### 6.1.5. Critical Zone Data Assimilation

The critical zone is composed of rock, soil, water, air, living organisms and their interactions in the permeable near-surface area of the Earth (Bogena et al., 2018). The anthropogenic, climatic and tectonic interconnections from the land surface to the subsurface are inherent in critical zones, and therefore, a full understanding of critical zones requires the integration of multiple disciplines, including hydrology, geophysics, ecology, biogeochemistry, and geochemistry.

Critical zone DA remains challenging because (a) coupled critical zone models considering surface–subsurface multiscale processes, especially those coupled with reactive transport and isotopic fractionation, are rarely available (Li et al., 2022). (b) In addition to horizontal heterogeneity, vertical heterogeneity introduces more complex error propagation for multivariable state updates. (c) Assimilation of new observations, including isotopes, ground penetrating radar data, P-band SAR, 3D microelectrical resistivity tomography (Vanella et al., 2016), sensor networks (Bogena et al., 2018), and the Internet of Things, requires the development of new observation operators that are suitable for critical zone observatories.

### 6.2. Opportunities

The flourishing of various types of data, including remote sensing, sensor networks, the Internet of Things, and social sensing, provide new opportunities for LDA. Additionally, LDA theory may become more generalized and unified by means of advanced methodologies, such as stochastic calculus and data-driven approaches, particularly deep learning, deep reinforcement learning, causal inference, and other big data analytics (X. Li et al., 2023). With these advanced techniques, LDA applications can be extended from natural systems to coupled natural and human systems (CNHS).

#### 6.2.1. Deluge of Data

Traditionally, Earth system observations are scarce, especially in remote regions where monitoring devices are sparsely distributed or nonexistent. However, this situation has dramatically changed due to the rapid progress in sensing techniques, information and communication technology, especially wireless sensor networks and the Internet of Things (Hart & Martinez, 2015). Additionally, new-generation of remote sensing satellites overperform their predecessors with unprecedented capacity in observing terrestrial systems. For example, CubeSat simultaneously improves the spatial and temporal resolution by scanning the Earth daily with hundreds of small satellites (Poghosyan & Golkar, 2017), which overcomes the largest challenges in traditional satellite remote sensing of high spatial resolution and lower temporal resolution, providing global daily meter-order images. For airborne platforms, the introduction of unmanned aerial vehicles (UAVs) expands the applicability of assimilating high-resolution remote sensing data at local and catchment scales.

Therefore, we are witnessing an unprecedented opportunity for ubiquitous sensing in Earth system. These remarkable changes require new data analytics and artificial intelligence to extract useful information for LDA.

They also require a substantial change in the LDA approach that can address new challenges resulting from the data deluge.

#### 6.2.2. Assimilation of Social Data

LDA is evolving beyond pure geophysical systems to encompass CNHS (Carrassi et al., 2018). Thus, LDA should consider the assimilation of social sensing data (X. P. Liu et al., 2017) into CNHS models, particularly agent-based models (Ward et al., 2016). This process is methodologically difficult because conventional assimilation methodology is appropriate only for quantifiable and structured natural system observational data but not for categorical variables or unstructured/soft data (e.g., natural language) in the human system. Therefore, extending DA approaches to meet the new requirements of assimilating categorical and nonstructural data in CNHS is a new research direction. Bayesian inference, capable of handling categorical distributions, provides new opportunities to overcome the challenge of assimilating discrete categorical variables. For example, conjugate prior features of the Dirichlet distribution are used to assimilate categorical data of land use into the cellular automata model to update multiple discrete state variables (Hu et al., 2022). Additionally, advancements in the combination of data assimilation and agent-based models have paved the way for assimilating big data from social media, social surveys, and surveillance cameras (Ward et al., 2016). Overall, although in its infancy, the assimilation of social data provides new insights to increase the understanding of the human component of the Earth system in the Anthropocene.

#### 6.2.3. Maturation of LDA Theory and Methods

The development of a more mature theory and methods for LDA has borrowed from theoretical and methodological advancements from stochastic differential equations and ergodic theory, which are mathematical foundations of Bayesian filters, including PFs and EnKF. The maturation of LDA theory necessitates the establishment of the spatial scale as the independent variable, better representation of the multiscale heterogeneities, and characterization of the scale-dependent representativeness errors. The formulation of scale representation and scale transformations in DA (Janjić et al., 2018) by means of stochastic calculus (F. Liu & Li, 2017) and ergodic theory have provided solid opportunities to form a more mature theory and methodology in LDA.

#### 6.2.4. Deep Integration With Big Data Analytics

In the era of big data, one of the most promising directions for LDA is harmonizing transformative theory with the exponentially growing observation data of land surface processes. Moreover, the symbiotic integration of advanced big data analytics, such as deep reinforcement learning, casual inference, and surrogate modeling, has propelled the fast and accurate simulation of land surface processes. In addition, DA facilitates the learning of new processes and relationships in the data. For example, the latent data assimilation (Cheng et al., 2022), combined with encoder-decoder networks that map high-dimensional neural state variables to low-dimensional dynamical state variables, can identify state variables from high-dimensional observational data, without prior knowledge of the underlying physics (B. Chen et al., 2022; C. Chen et al., 2022). With these technological advancements, LDA can serve as an engine of digital twin of Earth.

## 7. Summary

LDA integrates land models and land surface observations to produce state variables, parameters, and fluxes of the Earth system that are as accurate as possible while considering the uncertainties of both land models and observations. During past decades, LDA has undergone rapid development in theories, methods and applications and is widely accepted in the land surface research community. This review provides a synthesis of LDA advances, covering the full spectrum of LDA, including distinctive features in its theoretical and methodological developments, typical applications in terrestrial water and energy cycles, and developments of global- and catchment-scale LDASs and software platforms. We also point out the grand challenges and opportunities for future LDA studies.

LDA has developed its unique characteristics because land surface processes are highly nonlinear, and states, fluxes and parameters are scale dependent and spatially correlated. In addition, observations are typically non-Gaussian, and meteorological forcings are more important than the initial field of model states. These characteristics of land surface research demand new theoretical and methodological developments in LDA, such as the implementation of nonlinear and non-Gaussian DA approaches, accurate quantification and reduction in model and observation uncertainties, simultaneous estimation of model state and parameters, consideration of multiscale spatial correlation, and integration of ML and LDA.

Landmark studies assimilating the major states and fluxes in land surface processes, including soil moisture, snow, LST, evapotranspiration, groundwater and streamflow, are extracted from the volume of related works. Special attention is given to multivariable LDA since multisource and multiscale Earth observations are becoming increasingly abundant. The popularity of LDA can be partially attributed to the iconic developments and applications of various LDASs for different scales (i.e., global, regional, and catchment scales) and several common LDA software platforms for customized research in various fields, which are also introduced in detail.

In regard to the existing challenges and potential development trends for LDA, we conclude that the critical challenges focus on improving the physical understanding of land surface processes, including the development of high-quality meteorological forcing and new land reanalysis data, implementation of LDA for operational use, and interdisciplinary explorations of advancing coupled land–atmosphere DA and critical zone DA. The opportunities for LDA focus on advancing its theory and methodology, including addressing the data deluge in both Earth observation systems and social sensing, deep integrations with advanced big data analytics, and strengthening from stochastic calculus and ergodic theory. LDA is developing toward maturation.

As LDA theoretical foundations and its applications continue to mature, LDA not only advances the “scientific” understanding of land surface processes but also serves as an “engineering tool” in Earth system science. The former aspect underscores the role of LDA in enhancing the predictability and observability of high-dimensional nonlinear phenomena in land surface process research. By synergistically combining theory and data, LDA employs geophysical theories to express the resolved scientific insights, and uses observational data to substitute for unresolved processes. Pioneering studies have been conducted, for example, the combination of global optimization and DA to analyze the time series of the state and output innovations, enabling the accurate description of the nondriven slow part of the hydrograph and improvement of model structure (Vrugt et al., 2005). However, the assimilation of soil moisture did not improve the prediction of evapotranspiration of European forests (Strebel et al., 2023), indicating that this is mostly likely due to inappropriate description of evapotranspiration fluxes, and highlighting the need of improving model description of soil hydrological processes in land surface models. Therefore, further LDA investigations remain urgent needed to find more “scientific” evidence in practice. On the other hand, as an engineering tool, LDA has become a primary instrument for forecasting and reanalysis in land surface processes. It has bolstered the development of operational forecasting systems for various applications, such as NWP and hydrological forecasting. Moreover, LDA, being a part of the Earth system data assimilation toolbox, aids in engineering digital twins of Earth. Overall, this review, through synthesis of the current knowledge of LDA, could provide a steppingstone for its future development, particularly in the era of big data and artificial intelligence.

## Glossary

| | |
|---|---|
| AMSR-E | advanced microwave scanning radiometer for Earth observation system |
| CaLDAS | coupled land and atmosphere data assimilation system |
| CCI | climate change initiative |
| CLDAS | Chinese land data assimilation system |
| CLM | community land model |
| CMA | China Meteorological Administration |
| CMFD | China meteorological forcing data set |
| CoLM | common land model |
| ComDA | a common land data assimilation platform |
| COSMOS | cosmic-ray soil moisture observing system |
| CubeSat | a class of miniaturized satellites based around a form factor consisting of 10 cm cubes |

| | |
|---|---|
| DA | data assimilation |
| DART | data assimilation research testbed |
| ECMWF | European Centre for Medium-range Weather Forecasts |
| EKF | extended Kalman filter |
| ELDAS | European land data assimilation system |
| En4DVar | ensemble four-dimensional variational data assimilation |
| EnKF | ensemble Kalman filter |
| EnKS | ensemble Kalman smoother |
| EnSRF | ensemble square root filter |
| ERA5 | ECMWF reanalysis version5 |
| ESA | European Space Agency |
| ET | evapotranspiration |
| ETKF | ensemble transform Kalman filter |
| GLDAS | global land data assimilation system |
| GRACE | gravity recovery and climate experiment |
| HiWATER | Heihe watershed allied telemetry experimental research |
| JULES | joint UK land environment simulator |
| KF | Kalman filter |
| KLDAS | South Korea land data assimilation system |
| LAI | leaf area index |
| LDA | land data assimilation |
| LDAS | land data assimilation system |
| LETKF | local ensemble transform Kalman filter |
| LIS | land information system |
| LSM | land surface model |
| LST | land surface temperature |
| MCMC | Markov Chain Monte Carlo |
| MICLC | multisource integrated Chinese land cover |
| ML | machine learning |
| MKS | multiscale Kalman smoother |
| MODIS | moderate resolution imaging spectroradiometer |
| NAFE | Australian National Airborne Field Experiment |
| NLDAS | North American land data assimilation system |
| NLS-En4DVar | nonlinear least square En4DVar |
| OpenDA | open data assimilation library |
| PDAF | parallel data assimilation framework |

| | |
|---|---|
| PF | particle filter |
| REnKF | recursive EnKF |
| RTM | radiative transfer models |
| SALDAS | South American land data assimilation |
| SCA | snow cover area |
| SCF | snow cover fraction |
| SD | snow depth |
| SDC | snow depletion curve |
| SM | soil moisture |
| SMAP | soil moisture active passive satellite |
| SMAPEx | soil moisture active passive experiment |
| SMOS | soil moisture and ocean salinity satellite |
| SSM/I | special sensor microwave/imager |
| SWAT | soil & water assessment tool |
| SWAT-HDAS | SWAT-based hydrological data assimilation system |
| SWE | snow water equivalent |
| TB | brightness temperature |
| TMI | tropical rainfall measuring mission's microwave imager |
| TWS | terrestrial water storage |
| VIC | variable infiltration capacity model |

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

The authors declare that no new data were used or created in producing this manuscript. Figures 1–4, and 8 were created using Microsoft Visio Professional 2013. Figure 5 was created using Adobe Photoshop 2020. The subfigure of observation in Figure 2 is reproduced from Li et al. (2013). Parts of the subfigures used to create Figure 7 are available at National Tibetan Plateau Data Center (Jia, 2015) (https://doi.org/10.3972/heihe.114.2013.db), Zhang et al. (2017), Journal of Advances in Modeling Earth Systems (https://doi.org/10.1002/2017MS001144), X. Pan et al. (2017), Remote Sensing (https://doi.org/10.3390/rs9080774), and Wang et al. (2021), IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (https://doi.org/10.1109/JSTARS.2021.3108432). Parts of the subfigures used to create Figure 8 are available at Earth's Future (https://doi.org/10.1029/2022EF002966) and Zhu et al. (2014), Geoscientific Model Development (https://doi.org/10.5194/gmd-7-1467-2014).

**Acknowledgments**
We thank Professor Paolo D'Odorico and the other editors of Reviews of Geophysics for their help and valuable comments on our manuscript. We also thank Professor Jasper A. Vrugt and another anonymous reviewer for their insightful comments and suggestions. This work was supported by the Basic Science Center for Tibetan Plateau Earth System project of the National Natural Science Foundation of China (Grant 41988101), and the National Natural Science Foundation of China (Grants 42130113 and 42271442).

## References

Abbaszadeh, P., Gavahi, K., & Moradkhani, H. (2020). Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. *Advances in Water Resources*, *145*, 103721. https://doi.org/10.1016/j.advwatres.2020.103721

Abbaszadeh, P., Moradkhani, H., & Daescu, D. N. (2019). The quest for model uncertainty quantification: A hybrid ensemble and variational data assimilation framework. *Water Resources Research*, *55*(3), 2407–2431. https://doi.org/10.1029/2018WR023629

Abbaszadeh, P., Moradkhani, H., & Yan, H. X. (2018). Enhancing hydrologic data assimilation by evolutionary particle filter and Markov Chain Monte Carlo. *Advances in Water Resources*, *111*, 192–204. https://doi.org/10.1016/j.advwatres.2017.11.011

Abolafia-Rosenzweig, R., Livneh, B., Small, E. E., & Kumar, S. V. (2019). Soil moisture data assimilation to estimate irrigation water use. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3670–3690. https://doi.org/10.1029/2019MS001797

Ades, M., & van Leeuwen, P. J. (2013). An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, *139*(672), 820–840. https://doi.org/10.1002/qj.1995

Alipour, A., Ahmadalipour, A., & Moradkhani, H. (2020). Assessing flash flood hazard and damages in the southeast United States. *Journal of Flood Risk Management*, *13*(2), e12605. https://doi.org/10.1111/jfr3.12605

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment. *Journal of Hydrology*, *519*, 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041

Anderson, D. D., Anderson, D. F., & Park, J. (1999). GCD-sets in integral domains. *Houston Journal of Mathematics*, *25*(1), 15–34.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. https://doi.org/10.1175/2009BAMS2618.1

Anderson, J. L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, *140*(7), 2359–2371. https://doi.org/10.1175/mwr-d-11-00013.1

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*(6), 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004

Annis, A., & Nardi, F. (2019). Integrating VGI and 2D hydraulic models into a data assimilation framework for real time flood forecasting and mapping. *Geo-spatial Information Science*, *22*(4), 223–236. https://doi.org/10.1080/10095020.2019.1626135

Apte, A., Hairer, M., Stuart, A. M., & Voss, J. (2007). Sampling the posterior: An approach to non-Gaussian data assimilation. *Physica D: Nonlinear Phenomena*, *230*(1), 50–64. https://doi.org/10.1016/j.physd.2006.06.009

Arsenault, K. R., Houser, P. R., De Lannoy, G. J. M., & Dirmeyer, P. A. (2013). Impacts of snow cover fraction data assimilation on modeled energy and moisture budgets. *Journal of Geophysical Research: Atmospheres*, *118*(14), 7489–7504. https://doi.org/10.1002/jgrd.50542

Aubert, D., Loumagne, C., & Oudin, L. (2003). Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. *Journal of Hydrology*, *280*(1), 145–161. https://doi.org/10.1016/S0022-1694(03)00229-4

Baatz, R., Hendricks Franssen, H. J., Euskirchen, E., Sihi, D., Dietze, M., Ciavatta, S., et al. (2021). Reanalysis in Earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics*, *59*(3), e2020RG000715. https://doi.org/10.1029/2020RG000715

Bach, H., & Mauser, W. (2003). Methods and examples for remote sensing data assimilation in land surface process modeling. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(7), 1629–1637. https://doi.org/10.1109/TGRS.2003.813270

Baguis, P., & Roulin, E. (2017). Soil moisture data assimilation in a hydrological model: A case study in Belgium using large-scale satellite data. *Remote Sensing*, *9*(8), 820. https://doi.org/10.3390/rs9080820

Bai, Y., & Li, X. (2011). Evolutionary algorithm-based error parameterization methods for data assimilation. *Monthly Weather Review*, *139*(8), 2668–2685. https://doi.org/10.1175/2011MWR3641.1

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., et al. (2015). ERA-interim/land: A global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389–407. https://doi.org/10.5194/hess-19-389-2015

Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 607–633. https://doi.org/10.1002/qj.2982

Bao, J., Li, L. P., & Redoloza, F. (2020). Coupling ensemble smoother and deep learning with generative adversarial networks to deal with non-Gaussianity in flow and transport data assimilation. *Journal of Hydrology*, *590*, 125443. https://doi.org/10.1016/j.jhydrol.2020.125443

Barrett, D. J., & Renzullo, L. J. (2009). On the efficacy of combining thermal and microwave satellite data as observational constraints for root-zone soil moisture estimation. *Journal of Hydrometeorology*, *10*(5), 1109–1127. https://doi.org/10.1175/2009jhm1043.1

Bateni, S. M., & Entekhabi, D. (2012). Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters. *Water Resources Research*, *48*(8). https://doi.org/10.1029/2011WR011542

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., & Wedi, N. P. (2021). The digital revolution of Earth-system science. *Nature Computational Science*, *1*(2), 104–113. https://doi.org/10.1038/s43588-021-00023-0

Benavides Pinjosovsky, H. S., Thiria, S., Ottlé, C., Brajard, J., Badran, F., & Maugis, P. (2017). Variational assimilation of land surface temperature within the ORCHIDEE land surface model version 1.2.6. *Geoscientific Model Development*, *10*(1), 85–104. https://doi.org/10.5194/gmd-10-85-2017

Berry, T., & Sauer, T. (2013). Adaptive ensemble Kalman filtering of non-linear systems. *Tellus A: Dynamic Meteorology and Oceanography*, *65*(1), 20331. https://doi.org/10.3402/tellusa.v65i0.20331

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1–4), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Beven, K. J., & Cloke, H. L. (2012). Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood, et al. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011WR010982

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, *129*(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129%3c0420:ASWTET%3e2.0.CO;2

Bocquet, M., Pires, C. A., & Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, *138*(8), 2997–3023. https://doi.org/10.1175/2010mwr3164.1

Boegh, E., Thorsen, M., Butts, M. B., Hansen, S., Christiansen, J. S., Abrahamsen, P., et al. (2004). Incorporating remote sensing data in physically based distributed agro-hydrological modelling. *Journal of Hydrology*, *287*(1), 279–299. https://doi.org/10.1016/j.jhydrol.2003.10.018

Bogena, H. R., White, T., Bour, O., Li, X., & Jensen, K. H. (2018). Toward better understanding of terrestrial processes through long-term hydrological observatories. *Vadose Zone Journal*, *17*(1), 180194–180210. https://doi.org/10.2136/vzj2018.10.0194

Bosilovich, M. G., Chaudhuri, A. H., & Rixen, M. (2013). Earth system reanalysis: Progress, challenges, and opportunities. *Bulletin of the American Meteorological Society*, *94*(8), ES110–ES113. https://doi.org/10.1175/BAMS-D-12-00191.1

Botto, A., Belluco, E., & Camporese, M. (2018). Multi-source data assimilation for physically based hydrological modeling of an experimental hillslope. *Hydrology and Earth System Sciences*, *22*(8), 4251–4266. https://doi.org/10.5194/hess-22-4251-2018

Boucher, M.-A., Quilty, J., & Adamowski, J. (2020). Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons. *Water Resources Research*, *56*(6), e2019WR026226. https://doi.org/10.1029/2019WR026226

Buehner, M., & Charron, M. (2007). Spectral and spatial localization of background-error correlations for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *133*(624), 615–630. https://doi.org/10.1002/qj.50

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review*, *138*(5), 1567–1586. https://doi.org/10.1175/2009MWR3158.1

Buizza, C., Casas, C. Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., et al. (2022). Data learning: Integrating data assimilation and machine learning. *Journal of Computational Science*, *58*, 101525. https://doi.org/10.1016/j.jocs.2021.101525

Camporese, M., Paniconi, C., Putti, M., & Salandin, P. (2009). Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2008WR007031

Cao, Z. G., Ma, R. H., Duan, H. T., & Xue, K. (2019). Effects of broad bandwidth on the remote sensing of inland waters: Implications for high spatial resolution satellite data applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *153*, 110–122. https://doi.org/10.1016/j.isprsjprs.2019.05.001

Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), e535. https://doi.org/10.1002/wcc.535

Carrera, M. L., Bélair, S., & Bilodeau, B. (2015). The Canadian land data assimilation system (CaLDAS): Description and synthetic evaluation study. *Journal of Hydrometeorology*, *16*(3), 1293–1314. https://doi.org/10.1175/jhm-d-14-0089.1

Case, J. L., Kumar, S. V., Srikishen, J., & Jedlovec, G. J. (2011). Improving numerical weather predictions of summertime precipitation over the southeastern United States through a high-resolution initialization of the surface state. *Weather and Forecasting*, *26*(6), 785–807. https://doi.org/10.1175/2011WAF2222455.1

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., & Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. *The Cryosphere*, *10*(3), 1021–1038. https://doi.org/10.5194/tc-10-1021-2016

Che, T., Li, X., Jin, R., & Huang, C. L. (2014). Assimilating passive microwave remote sensing data into a land surface model to improve the estimation of snow depth. *Remote Sensing of Environment*, *143*, 54–63. https://doi.org/10.1016/j.rse.2013.12.009

Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., & Lipson, H. (2022). Automated discovery of fundamental variables hidden in experimental data. *Nature Computational Science*, *2*(7), 433–442. https://doi.org/10.1038/s43588-022-00281-6

Chen, C., Dou, Y., Chen, J., & Xue, Y. (2022). A novel neural network training framework with data assimilation. *The Journal of Supercomputing*, *78*(17), 19020–19045. https://doi.org/10.1007/s11227-022-04629-7

Chen, H., Yang, D. W., Hong, Y., Gourley, J. J., & Zhang, Y. (2013). Hydrological data assimilation with the ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Advances in Water Resources*, *59*, 209–220. https://doi.org/10.1016/j.advwatres.2013.06.010

Chen, W. J., Huang, C. L., Yang, Z.-L., & Zhang, Y. (2021). Retrieving accurate soil moisture over the Tibetan Plateau using multisource remote sensing data assimilation with simultaneous state and parameter estimations. *Journal of Hydrometeorology*, *22*(10), 2751–2766. https://doi.org/10.1175/JHM-D-20-0298.1

Chen, Y., & Oliver, D. S. (2010). Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Computational Geosciences*, *14*(4), 579–601. https://doi.org/10.1007/s10596-009-9174-6

Cheng, S., Chen, J., Anastasiou, C., Angeli, P., Matar, O. K., Guo, Y. K., et al. (2022). Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. *Journal of Scientific Computing*, *94*(1), 11. https://doi.org/10.1007/s10915-022-02059-4

Christakos, G. (2005). Methodological developments in geophysical assimilation modeling. *Reviews of Geophysics*, *43*(2), RG2001. https://doi.org/10.1029/2004RG000163

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X. G., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Clark, M. P., & Slater, A. G. (2006). Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, *7*(1), 3–22. https://doi.org/10.1175/JHM474.1

Cocucci, T. J., Pulido, M., Lucini, M., & Tandeo, P. (2021). Model error covariance estimation in Particle and ensemble Kalman filters using an online expectation–maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*, *147*(734), 526–543. https://doi.org/10.1002/qj.3931

Cooper, M., Brown, M. E., Azzarri, C., & Meinzen-Dick, R. (2019). Hunger, nutrition, and precipitation: Evidence from Ghana and Bangladesh. *Population and Environment*, *41*(2), 151–208. https://doi.org/10.1007/s11111-019-00323-8

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., et al. (2003). Real-time and retrospective forcing in the North American land data assimilation system (NLDAS) project. *Journal of Geophysical Research*, *108*(D22). https://doi.org/10.1029/2002JD003118

Crow, W. T., & Ryu, D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, *13*(1), 1–16. https://doi.org/10.5194/hess-13-1-2009

Crow, W. T., & van den Berg, M. J. (2010). An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2010WR009402

Crow, W. T., van den Berg, M. J., Huffman, G. J., & Pellarin, T. (2011). Correcting rainfall using satellite-based surface soil moisture retrievals: The soil moisture analysis rainfall tool (SMART). *Water Resources Research*, *47*(8). https://doi.org/10.1029/2011WR010576

Dai, Y. J., Zeng, X. B., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The common land model. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1024. https://doi.org/10.1175/BAMS-84-8-1013

Daley, R. (1997). Atmospheric data assimilation (Special issue: Data assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, *75*(1B), 319–329. https://doi.org/10.2151/jmsj1965.75.1B_319

Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R. N., & Walker, J. P. (2021). A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, *57*(2), e2020WR027859. https://doi.org/10.1029/2020WR027859

David, C. H., Yang, Z.-L., & Hong, S. (2013). Regional-scale river flow modeling using off-the-shelf runoff products, thousands of mapped rivers and hundreds of stream flow gauges. *Environmental Modelling & Software*, *42*, 116–132. https://doi.org/10.1016/j.envsoft.2012.12.011

Deb, P., Abbaszadeh, P., & Moradkhani, H. (2022). An ensemble data assimilation approach to improve farm-scale actual evapotranspiration estimation. *Agricultural and Forest Meteorology*, *321*, 108982. https://doi.org/10.1016/j.agrformet.2022.108982

DeChant, C. M., & Moradkhani, H. (2011). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrology and Earth System Sciences*, *15*(11), 3399–3410. https://doi.org/10.5194/hess-15-3399-2011

DeChant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, *48*(4). https://doi.org/10.1029/2011WR011011

DeChant, C. M., & Moradkhani, H. (2014). Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *Journal of Hydrology*, *519*, 2967–2977. https://doi.org/10.1016/j.jhydrol.2014.05.045

de Goncalves, L. G. G., Shuttleworth, W. J., Burke, E. J., Houser, P., Toll, D. L., Rodell, M., & Arsenault, K. (2006). Toward a South America land data assimilation system: Aspects of land surface model spin-up using the simplified simple biosphere. *Journal of Geophysical Research*, *111*(D17). https://doi.org/10.1029/2005JD006297

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., & Verhoest, N. E. C. (2007). State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency. *Water Resources Research*, *43*(6). https://doi.org/10.1029/2006WR005100

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., & Pauwels, V. R. N. (2012). Multiscale assimilation of advanced microwave scanning radiometer–EOS snow water equivalent and moderate resolution imaging spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011WR010588

De Luca, P., Galletti, A., Giunta, G., & Marcellino, L. (2021). Recursive filter based GPU algorithms in a data assimilation scenario. *Journal of Computational Science*, *53*, 101339. https://doi.org/10.1016/j.jocs.2021.101339

Derber, J., & Rosati, A. (1989). A global oceanic data assimilation system. *Journal of Physical Oceanography*, *19*(9), 1333–1347. https://doi.org/10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al. (2022). Coupled data assimilation at ECMWF: Current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, *148*(747), 2672–2702. https://doi.org/10.1002/qj.4330

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3385–3396. https://doi.org/10.1256/qj.05.108

Devegowda, D., Arroyo-Negrete, E., & Datta-Gupta, A. (2010). Flow relevant covariance localization during dynamic data assimilation using EnKF. *Advances in Water Resources*, *33*(2), 129–145. https://doi.org/10.1016/j.advwatres.2009.10.001

Dhungel, R., Aiken, R., Lin, X. M., Kenyon, S., Colaizzi, P. D., Luhman, R., et al. (2020). Restricted water allocations: Landscape-scale energy balance simulations and adjustments in agricultural water applications. *Agricultural Water Management*, *227*, 105854. https://doi.org/10.1016/j.agwat.2019.105854

Doherty, J., & Moore, C. (2020). Decision support modeling: Data assimilation, uncertainty quantification, and strategic abstraction. *Ground Water*, *58*(3), 327–337. https://doi.org/10.1111/gwat.12969

Dokoohaki, H., Morrison, B. D., Raiho, A., Serbin, S. P., Zarada, K., Dramko, L., & Dietze, M. (2022). 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 Development*, *15*(8), 3233–3252. https://doi.org/10.5194/gmd-15-3233-2022

Dong, J. Z., Steele-Dunne, S. C., Judge, J., & van de Giesen, N. (2015). A particle batch smoother for soil moisture estimation using soil temperature observations. *Advances in Water Resources*, *83*, 111–122. https://doi.org/10.1016/j.advwatres.2015.05.017

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving Earth system science for over a decade. *Hydrology and Earth System Sciences*, *25*(11), 5749–5804. https://doi.org/10.5194/hess-25-5749-2021

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Draper, C., & Reichle, R. H. (2019). Assimilation of satellite soil moisture for improved atmospheric reanalyses. *Monthly Weather Review*, *147*(6), 2163–2188. https://doi.org/10.1175/mwr-d-18-0393.1

Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters*, *39*(4). https://doi.org/10.1029/2011GL050655

Duan, Q. Y., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985

Dubey, A. K., Kumar, P., Chembolu, V., Dutta, S., Singh, R. P., & Rajawat, A. S. (2021). Flood modeling of a large transboundary river using WRF-Hydro and microwave remote sensing. *Journal of Hydrology*, *598*, 126391. https://doi.org/10.1016/j.jhydrol.2021.126391

Dumedah, G., & Coulibaly, P. (2013a). Evolutionary assimilation of streamflow in distributed hydrologic modeling using in-situ soil moisture data. *Advances in Water Resources*, *53*, 231–241. https://doi.org/10.1016/j.advwatres.2012.07.012

Dumedah, G., & Coulibaly, P. (2013b). Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. *Advances in Water Resources*, *60*, 47–63. https://doi.org/10.1016/j.advwatres.2013.07.007

Dumedah, G., & Walker, J. P. (2014). Intercomparison of the JULES and CABLE land surface models through assimilation of remotely sensed soil moisture in southeast Australia. *Advances in Water Resources*, *74*, 231–244. https://doi.org/10.1016/j.advwatres.2014.09.011

Dumedah, G., Walker, J. P., & Merlin, O. (2015). Root-zone soil moisture estimation from assimilation of downscaled soil moisture and ocean salinity data. *Advances in Water Resources*, *84*, 14–22. https://doi.org/10.1016/j.advwatres.2015.07.021

Dunne, S., & Entekhabi, D. (2006). Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment. *Water Resources Research*, *42*(1). https://doi.org/10.1029/2005WR004334

Durand, M., Kim, E. J., & Margulis, S. A. (2009). Radiance assimilation shows promise for snowpack characterization. *Geophysical Research Letters*, *36*(2). https://doi.org/10.1029/2008GL035214

Durand, M., & Margulis, S. A. (2006). Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. *Journal of Hydrometeorology*, *7*(3), 443–457. https://doi.org/10.1175/JHM502.1

Durand, M., & Margulis, S. A. (2008). Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization. *Journal of Geophysical Research*, *113*(D2), D02105. https://doi.org/10.1029/2007JD008662

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., et al. (2016). Continental and global scale flood forecasting systems. *WIREs Water*, *3*(3), 391–418. https://doi.org/10.1002/wat2.1137

Entekhabi, D., Nakamura, H., & Njoku, E. G. (1994). Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote Sensing*, *32*(2), 438–448. https://doi.org/10.1109/36.295058

Ercolani, G., & Castelli, F. (2017). Variational assimilation of streamflow data in distributed flood forecasting. *Water Resources Research*, *53*(1), 158–183. https://doi.org/10.1002/2016WR019208

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G., & van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, *128*(6), 1852–1867. https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al. (2019). Hillslope hydrology in global change research and earth system modeling. *Water Resources Research*, *55*(2), 1737–1772. https://doi.org/10.1029/2018WR023903

Fan, Y. R., Huang, G. H., Baetz, B. W., Li, Y. P., Huang, K., Chen, X., & Gao, M. (2017). Development of integrated approaches for hydrological data assimilation through combination of ensemble Kalman filter and particle filter methods. *Journal of Hydrology*, *550*, 412–426. https://doi.org/10.1016/j.jhydrol.2017.05.010

Farhadi, L., Reichle, R. H., De Lannoy, G. J. M., & Kimball, J. S. (2015). Assimilation of freeze–thaw observations into the NASA catchment land surface model. *Journal of Hydrometeorology*, *16*(2), 730–743. https://doi.org/10.1175/jhm-d-14-0065.1

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

Frei, M., & Künsch, H. R. (2013). Bridging the ensemble Kalman and particle filters. *Biometrika*, *100*(4), 781–800. https://doi.org/10.1093/biomet/ast020

Gan, Y. J., Duan, Q. Y., Gong, W., Tong, C., Sun, Y. W., Chu, W., et al. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, *51*, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gavahi, K., Abbaszadeh, P., Moradkhani, H., Zhan, X. W., & Hain, C. (2020). Multivariate assimilation of remotely sensed soil moisture and evapotranspiration for drought monitoring. *Journal of Hydrometeorology*, *21*(10), 2293–2308. https://doi.org/10.1175/JHM-D-20-0057.1

Gelsinari, S., Doble, R., Daly, E., & Pauwels, V. R. N. (2020). Feasibility of improving groundwater modeling by assimilating evapotranspiration rates. *Water Resources Research*, *56*(2), e2019WR025983. https://doi.org/10.1029/2019WR025983

Gharamti, M. E., McCreight, J. L., Noh, S. J., Hoar, T. J., RafieeiNasab, A., & Johnson, B. K. (2021). Ensemble streamflow data assimilation using WRF-Hydro and DART: Novel localization and inflation techniques applied to hurricane Florence flooding. *Hydrology and Earth System Sciences*, *25*(9), 5315–5336. https://doi.org/10.5194/hess-25-5315-2021

Gichamo, T. Z., & Tarboton, D. G. (2019). Ensemble streamflow forecasting using an energy balance snowmelt model coupled to a distributed hydrologic model with assimilation of snow and streamflow observations. *Water Resources Research*, *55*(12), 10813–10838. https://doi.org/10.1029/2019WR025472

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. https://doi.org/10.1029/2018MS001603

Grayson, R. B., Blöschl, G., Western, A. W., & McMahon, T. A. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources*, *25*(8), 1313–1334. https://doi.org/10.1016/S0309-1708(02)00060-X

Guillet, O., Weaver, A. T., Vasseur, X., Michel, Y., Gratton, S., & Gürol, S. (2019). Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. *Quarterly Journal of the Royal Meteorological Society*, *145*(722), 1947–1967. https://doi.org/10.1002/qj.3537

Hamill, T. M., & Whitaker, J. S. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Monthly Weather Review*, *133*(11), 3132–3147. https://doi.org/10.1175/mwr3020.1

Han, E., Merwade, V., & Heathman, G. C. (2012). Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *Journal of Hydrology*, *416*, 98–117. https://doi.org/10.1016/j.jhydrol.2011.11.039

Han, X., Li, X., He, G., Kumbhar, P., Montzka, C., Kollet, S., et al. (2015). *DasPy 1.0: The open source multivariate land data assimilation framework in combination with the community land model 4.5* (pp. 7395–7444). Geoscientific Model Development Discussion. https://doi.org/10.5194/gmdd-8-7395-2015

Han, X., Li, X., Hendricks Franssen, H. J., Vereecken, H., & Montzka, C. (2012). Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. *Hydrology and Earth System Sciences*, *16*(5), 1349–1363. https://doi.org/10.5194/hess-16-1349-2012

Han, X. J., Franssen, H.-J. H., Montzka, C., & Vereecken, H. (2014). Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. *Water Resources Research*, *50*(7), 6081–6105. https://doi.org/10.1002/2013WR014586

Han, X. J., Hendricks Franssen, H.-J., Jiménez Bello, M. Á., Rosolem, R., Bogena, H., Alzamora, F. M., et al. (2016). Simultaneous soil moisture and properties estimation for a drip irrigated field by assimilating cosmic-ray neutron intensity. *Journal of Hydrology*, *539*, 611–624. https://doi.org/10.1016/j.jhydrol.2016.05.050

Hart, J. K., & Martinez, K. (2015). Toward an environmental internet of things. *Earth and Space Science*, *2*(5), 194–200. https://doi.org/10.1002/2014EA000044

He, J., Yang, K., Tang, W. J., Lu, H., Qin, J., Chen, Y., & Li, X. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, *7*(1), 25. https://doi.org/10.1038/s41597-020-0369-y

He, Q. Z., Barajas-Solano, D., Tartakovsky, G., & Tartakovsky, A. M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, *141*, 103610. https://doi.org/10.1016/j.advwatres.2020.103610

He, X. L., Xu, T. R., Bateni, S. M., Ek, M., Liu, S. M., & Chen, F. (2020). Mapping regional evapotranspiration in cloudy skies via variational assimilation of all-weather land surface temperature observations. *Journal of Hydrology*, *585*, 124790. https://doi.org/10.1016/j.jhydrol.2020.124790

He, X. L., Xu, T. R., Bateni, S. M., Neale, C. M. U., Liu, S. M., Auligne, T., et al. (2019). Mapping regional turbulent heat fluxes via assimilation of MODIS land surface temperature data into an ensemble Kalman smoother framework. *Earth and Space Science*, *6*(12), 2423–2442. https://doi.org/10.1029/2019EA000705

Hendricks Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, *44*(9). https://doi.org/10.1029/2007WR006505

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hostache, R., Chini, M., Giustarini, L., Neal, J., Kavetski, D., Wood, M., et al. (2018). Near-real-time assimilation of SAR-derived flood maps for improving flood forecasts. *Water Resources Research*, *54*(8), 5516–5535. https://doi.org/10.1029/2017WR022205

Hou, J. L., Huang, C. L., Chen, W. J., & Zhang, Y. (2021). Improving snow estimates through assimilation of MODIS fractional snow cover data using machine learning algorithms and the Common Land Model. *Water Resources Research*, *57*(7), e2020WR029010. https://doi.org/10.1029/2020WR029010

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, *34*(12), 3405–3420. https://doi.org/10.1029/1998WR900001

Houtekamer, P. L., He, B., & Mitchell, H. L. (2014). Parallel implementation of an ensemble Kalman filter. *Monthly Weather Review*, *142*(3), 1163–1182. https://doi.org/10.1175/mwr-d-13-00011.1

Hu, X. L., Liu, F., Qi, Y., Zhang, J. L., & Li, X. (2022). A novel strategy to assimilate category variables in land-use models based on Dirichlet distribution. *Environmental Modelling & Software*, *149*, 105324. https://doi.org/10.1016/j.envsoft.2022.105324

Huang, C. L., Chen, W. J., Li, Y., Shen, H. F., & Li, X. (2016). Assimilating multi-source data into land surface model to simultaneously improve estimations of soil moisture, soil temperature, and surface turbulent fluxes in irrigated fields. *Agricultural and Forest Meteorology*, *230–231*, 142–156. https://doi.org/10.1016/j.agrformet.2016.03.013

Huang, C. L., Li, X., & Lu, L. (2008). Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sensing of Environment*, *112*(4), 1320–1336. https://doi.org/10.1016/j.rse.2007.03.028

Huang, C. L., Li, X., Lu, L., & Gu, J. (2008). Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sensing of Environment*, *112*(3), 888–900. https://doi.org/10.1016/j.rse.2007.06.026

Huang, C. L., Margulis, S. A., Durand, M. T., & Musselman, K. N. (2012). Assessment of snow grain-size model and stratigraphy representation impacts on snow radiance assimilation: Forward modeling evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(11), 4551–4564. https://doi.org/10.1109/TGRS.2012.2192480

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Hurkmans, R., Paniconi, C., & Troch, P. A. (2006). Numerical assessment of a dynamical relaxation data assimilation scheme for a catchment hydrological model. *Hydrological Processes*, *20*(3), 549–563. https://doi.org/10.1002/hyp.5921

Irmak, A., & Kamble, B. (2009). Evapotranspiration data assimilation with genetic algorithms and SWAP model for on-demand irrigation. *Irrigation Science*, *28*(1), 101–112. https://doi.org/10.1007/s00271-009-0193-9

Jackson, T. J., Schugge, T. J., Nicks, A. D., Coleman, G. A., & Engman, E. T. (1981). Soil moisture updating and microwave remote sensing for hydrological simulation. *Hydrological Sciences Bulletin*, *26*(3), 305–319. https://doi.org/10.1080/02626668109490889

Jacobs, C. M. J., Moors, E. J., Ter Maat, H. W., Teuling, A. J., Balsamo, G., Bergaoui, K., et al. (2008). Evaluation of European land data assimilation system (ELDAS) products using in situ observations. *Tellus A: Dynamic Meteorology and Oceanography*, *60*(5), 1023–1037. https://doi.org/10.1111/j.1600-0870.2008.00351.x

Jacobs, J. M., Mohanty, B. P., Hsu, E.-C., & Miller, D. (2004). SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sensing of Environment*, *92*(4), 436–446. https://doi.org/10.1016/j.rse.2004.02.017

Jafarzadegan, K., Abbaszadeh, p., & Moradkhani, H. (2021). Sequential data assimilation for real-time probabilistic flood inundation mapping. *Hydrology and Earth System Sciences*, *25*(9), 4995–5011. https://doi.org/10.5194/hess-25-4995-2021

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., et al. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. https://doi.org/10.1002/qj.3130

Jia, L. (2015). The evapotranspiration data in the Heihe River basin (2009–2011) [Dataset]. *National Tibetan Plateau / Third Pole Environment Data Center*. https://doi.org/10.3972/heihe.114.2013.db

Jiang, Y. Z., Yang, K., Qi, Y. C., Zhou, X., He, J., Lu, H., et al. (2022). TPHiPr: A long-term (1979–2020) high-accuracy precipitation dataset (1/30°, daily) for the third pole region based on high-resolution atmospheric modeling and dense observations. *Earth System Science Data*, *15*(2), 1–32. https://doi.org/10.5194/essd-15-621-2023

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, *92*(3), 401–422. https://doi.org/10.1109/JPROC.2003.823141

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabrera-Poy, J. (2007). 4D-Var or ensemble Kalman filter? *Tellus A: Dynamic Meteorology and Oceanography*, *59*(5), 758–773. https://doi.org/10.1111/j.1600-0870.2007.00261.x

Khaki, M., Forootan, E., Kuhn, M., Awange, J., van Dijk, A. I. J. M., Schumacher, M., & Sharifi, M. (2018). Determining water storage depletion within Iran by assimilating GRACE data into the W3RA hydrological model. *Advances in Water Resources*, *114*, 1–18. https://doi.org/10.1016/j.advwatres.2018.02.008

Kim, S., Shen, H. J., Noh, S., Seo, D.-J., Welles, E., Pelgrim, E., et al. (2021). High-resolution modeling and prediction of urban floods using WRF-Hydro and data assimilation. *Journal of Hydrology*, *598*, 126236. https://doi.org/10.1016/j.jhydrol.2021.126236

King, F., Erler, A. R., Frey, S. K., & Fletcher, C. G. (2020). Application of machine learning techniques for regional bias correction of snow water equivalent estimates in Ontario, Canada. *Hydrology and Earth System Sciences*, *24*(10), 4887–4902. https://doi.org/10.5194/hess-24-4887-2020

Kivman, G. A. (2003). Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, *10*(3), 253–259. https://doi.org/10.5194/npg-10-253-2003

Kotsuki, S., Miyoshi, T., Terasaki, K., Lien, G.-Y., & Kalnay, E. (2017). Assimilating the global satellite mapping of precipitation data with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM). *Journal of Geophysical Research: Atmospheres*, *122*(2), 631–650. https://doi.org/10.1002/2016JD025355

Kumar, S. V., Jasinski, M., Mocko, D. M., Rodell, M., Borak, J., Li, B. L., et al. (2019). NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. *Journal of Hydrometeorology*, *20*(8), 1571–1593. https://doi.org/10.1175/jhm-d-17-0125.1

Kumar, S. V., Peters-Lidard, C. D., Arsenault, K. R., Getirana, A., Mocko, D., & Liu, Y. Q. (2015). Quantifying the added value of snow cover area observations in passive microwave snow depth data assimilation. *Journal of Hydrometeorology*, *16*(4), 1736–1741. https://doi.org/10.1175/jhm-d-15-0021.1

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, *21*(10), 1402–1415. https://doi.org/10.1016/j.envsoft.2005.07.004

Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., & Santanello, J. A. (2012). A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resources Research*, *48*(3). https://doi.org/10.1029/2010wr010261

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B. L., et al. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *Journal of Hydrometeorology*, *17*(7), 1951–1972. https://doi.org/10.1175/jhm-d-15-0157.1

Kurtz, W., He, G. W., Kollet, S. J., Maxwell, R. M., Vereecken, H., & Hendricks Franssen, H.-J. (2016). TerrSysMP–PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geoscientific Model Development*, *9*(4), 1341–1360. https://doi.org/10.5194/gmd-9-1341-2016

Kurtz, W., Lapin, A., Schilling, O. S., Tang, Q., Schiller, E., Braun, T., et al. (2017). Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. *Environmental Modelling & Software*, *93*, 418–435. https://doi.org/10.1016/j.envsoft.2017.03.011

Kwon, Y., Forman, B. A., Ahmad, J. A., Kumar, S. V., & Yoon, Y. (2019). Exploring the utility of machine learning-based passive microwave brightness temperature data assimilation over terrestrial snow in high mountain Asia. *Remote Sensing*, *11*(19), 2265. https://doi.org/10.3390/rs11192265

Lahoz, B. K. W., & Menard, R. (2010). *Data assimilation*. Springer -Verlag Berlin Heidelberg.

Largeron, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., et al. (2020). Toward snow cover estimation in mountainous areas using modern data assimilation methods: A review. *Frontiers in Earth Science*, *8*, 325. https://doi.org/10.3389/feart.2020.00325

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

Lei, F. N., Crow, W. T., Kustas, W. P., Dong, J. Z., Yang, Y., Knipper, K. R., et al. (2020). Data assimilation of high-resolution thermal and radar remote sensing retrievals for soil moisture monitoring in a drip-irrigated vineyard. *Remote Sensing of Environment*, *239*, 111622. https://doi.org/10.1016/j.rse.2019.111622

Lei, L. L., Anderson, J. L., & Romine, G. S. (2015). Empirical localization functions for ensemble Kalman filter data assimilation in regions with and without precipitation. *Monthly Weather Review*, *143*(9), 3664–3679. https://doi.org/10.1175/mwr-d-14-00415.1

Leisenring, M., & Moradkhani, H. (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, *25*(2), 253–270. https://doi.org/10.1007/s00477-010-0445-5

Lewis, P., Gómez-Dans, J., Kaminski, T., Settle, J., Quaife, T., Gobron, N., et al. (2012). An earth observation land data assimilation system (EO-LDAS). *Remote Sensing of Environment*, *120*, 219–235. https://doi.org/10.1016/j.rse.2011.12.027

Li, B. L., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*(9), 7564–7586. https://doi.org/10.1029/2018WR024618

Li, D. Y., Andreadis, K. M., Margulis, S. A., & Lettenmaier, D. P. (2020). A data assimilation framework for generating space-time continuous daily SWOT river discharge data products. *Water Resources Research*, *56*(6). https://doi.org/10.1029/2019WR026999

Li, D. Z., Hendricks Franssen, H.-J., Han, X. J., Jiménez-Bello, M. A., Martínez Alzamora, F., & Vereecken, H. (2018). Evaluation of an operational real-time irrigation scheduling scheme for drip irrigated citrus fields in Picassent, Spain. *Agricultural Water Management*, *208*, 465–477. https://doi.org/10.1016/j.agwat.2018.06.022

Li, H., Kalnay, E., & Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, *135*(639), 523–533. https://doi.org/10.1002/qj.371

Li, L. C., Bisht, G., & Leung, L. R. (2022). Spatial heterogeneity effects on land surface modeling of water and energy partitioning. *Geoscientific Model Development*, *15*(14), 5489–5510. https://doi.org/10.5194/gmd-15-5489-2022

Li, P. J., Zha, Y. Y., Shi, L. S., Tso, C.-H. M., Zhang, Y. G., & Zeng, W. Z. (2020). Comparison of the use of a physical-based model with data assimilation and machine learning methods for simulating soil water dynamics. *Journal of Hydrology*, *584*, 124692. https://doi.org/10.1016/j.jhydrol.2020.124692

Li, X. (2014). Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems. *Science China Earth Sciences*, *57*(1), 80–87. https://doi.org/10.1007/s11430-013-4728-9

Li, X., Cheng, G. D., Liu, S. M., Xiao, Q., Ma, M. G., Jin, R., et al. (2013). Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, *94*(8), 1145–1160. https://doi.org/10.1175/bams-d-12-00154.1

Li, X., Feng, M., Ran, Y. H., Su, Y., Liu, F., Huang, C. L., et al. (2023). Big data in earth system science and progress towards a digital twin. *Nature Reviews Earth & Environment*, *4*(5), 319–332. https://doi.org/10.1038/s43017-023-00409-w

Li, X., Huang, C. L., Che, T., Jin, R., Wang, S. G., Wang, J. M., et al. (2007). Development of a Chinese land data assimilation system: Its progress and prospects. *Progress in Natural Science*, *17*(8), 881–892. https://doi.org/10.1080/10002007088537487

Li, X., Liu, F., & Fang, M. (2020). Harmonizing models and observations: Data assimilation in earth system science. *Science China Earth Sciences*, *63*(8), 1059–1068. https://doi.org/10.1007/s11430-019-9620-x

Lievens, H., De Lannoy, G. J. M., Al Bitar, A., Drusch, M., Dumedah, G., Hendricks Franssen, H.-J., et al. (2016). Assimilation of SMOS soil moisture and brightness temperature products into a land surface model. *Remote Sensing of Environment*, *180*, 292–304. https://doi.org/10.1016/j.rse.2015.10.033

Lim, Y.-J., Byun, K.-Y., Lee, T.-Y., Kwon, H., Hong, J., & Kim, J. (2012). A land data assimilation system using the MODIS-derived land data and its application to numerical weather prediction in East Asia. *Asia-Pacific Journal of Atmospheric Sciences*, *48*(1), 83–95. https://doi.org/10.1007/s13143-012-0008-4

Liu, F., & Li, X. (2017). Formulation of scale transformation in a stochastic data assimilation framework. *Nonlinear Processes in Geophysics*, *24*(2), 279–291. https://doi.org/10.5194/npg-24-279-2017

Liu, F., Wang, L. X., Li, X., & Huang, C. L. (2020). ComDA: A common software for nonlinear and non-Gaussian land data assimilation. *Environmental Modelling & Software*, *127*, 104638. https://doi.org/10.1016/j.envsoft.2020.104638

Liu, S. F., Shao, Y. P., Yang, C. G., Lin, Z. H., & Li, M. (2012). Improved regional hydrologic modelling by assimilation of streamflow data into a regional hydrologic model. *Environmental Modelling & Software*, *31*, 141–149. https://doi.org/10.1016/j.envsoft.2011.12.005

Liu, X. P., He, J. L., Yao, Y., Zhang, J. B., Liang, H. L., Wang, H., & Hong, Y. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science*, *31*(8), 1675–1696. https://doi.org/10.1080/13658816.2017.1324976

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., et al. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, *16*(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Liu, Y. H., Liu, P., Zhang, L., Zhang, X. J., Zhang, Y. F., & Cheng, L. (2021). Detecting and attributing drought-induced changes in catchment hydrological behaviours in a southeastern Australia catchment using a data assimilation method. *Hydrological Processes*, *35*(7), e14289. https://doi.org/10.1002/hyp.14289

Liu, Y. Q., Peters-Lidard, C. D., Kumar, S., Foster, J. L., Shaw, M., Tian, Y. D., & Fall, G. M. (2013). Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. *Advances in Water Resources*, *54*, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., & Casalí, J. (2018). On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Advances in Water Resources*, *111*, 86–104. https://doi.org/10.1016/j.advwatres.2017.10.034

Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., et al. (2000). The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *126*(570), 2991–3012. https://doi.org/10.1002/qj.49712657002

Lü, H., Li, X., Yu, Z., Horton, R., Zhu, Y., Hao, Z., & Xiang, L. (2010). Using a $H_\infty$ filter assimilation procedure to estimate root zone soil water content. *Hydrological Processes*, *24*(25), 3648–3660. https://doi.org/10.1002/hyp.7778

Lu, H., Zheng, D., Yang, K., & Yang, F. (2020). Last-decade progress in understanding and modeling the land surface processes on the Tibetan Plateau. *Hydrology and Earth System Sciences*, *24*(12), 5745–5758. https://doi.org/10.5194/hess-24-5745-2020

Lu, Y., Dong, J., & Steele-Dunne, S. C. (2019). Impact of soil moisture data resolution on soil moisture and surface heat flux estimates through data assimilation: A case study in the southern Great Plains. *Journal of Hydrometeorology*, *20*(4), 715–730. https://doi.org/10.1175/jhm-d-18-0234.1

Lu, Y., Steele-Dunne, S. C., Farhadi, L., & van de Giesen, N. (2017). Mapping surface heat fluxes by assimilating SMAP soil moisture and GOES land surface temperature data. *Water Resources Research*, *53*(12), 10858–10877. https://doi.org/10.1002/2017WR021415

Luo, L., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., et al. (2003). Validation of the North American land data assimilation system (NLDAS) retrospective forcing over the southern Great Plains. *Journal of Geophysical Research*, *108*(D22), 8843. https://doi.org/10.1029/2002JD003246

MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E., Kuppel, S., et al. (2022). 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. *Global Biogeochemical Cycles*, *36*(7), e2021GB007177. https://doi.org/10.1029/2021GB007177

Magnusson, J., Gustafsson, D., Hüsler, F., & Jonas, T. (2014). Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water Resources Research*, *50*(10), 7816–7835. https://doi.org/10.1002/2014WR015302

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., & Jonas, T. (2017). Improving physically based snow simulations by assimilating snow depths using the particle filter. *Water Resources Research*, *53*(2), 1125–1143. https://doi.org/10.1002/2016WR019092

Malartic, Q., Farchi, A., & Bocquet, M. (2022). State, global, and local parameter estimation using local ensemble Kalman filters: Applications to online machine learning of chaotic dynamics. *Quarterly Journal of the Royal Meteorological Society*, *148*(746), 2167–2193. https://doi.org/10.1002/qj.4297

Manabe, S. (1969). Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface. *Monthly Weather Review*, *97*(11), 739–774. https://doi.org/10.1175/1520-0493(1969)09760;0739:catoc62;2.3.co;2

Margulis, S. A., & Entekhabi, D. (2001). A coupled land surface–boundary layer model and its adjoint. *Journal of Hydrometeorology*, *2*(3), 274–296. https://doi.org/10.1175/1525-7541(2001)002<0274:Aclsbl>2.0.Co;2

Margulis, S. A., Girotto, M., Cortés, G., & Durand, M. (2015). A particle batch smoother approach to snow water equivalent estimation. *Journal of Hydrometeorology*, *16*(4), 1752–1772. https://doi.org/10.1175/jhm-d-14-0177.1

Margulis, S. A., McLaughlin, D., Entekhabi, D., & Dunne, S. (2002). Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resources Research*, *38*(12), 35-31–35-18. https://doi.org/10.1029/2001WR001114

Margulis, S. A., Wood, E. F., & Troch, P. A. (2006). The terrestrial water cycle: Modeling and data assimilation across catchment scales. *Journal of Hydrometeorology*, *7*(3), 309–311. https://doi.org/10.1175/jhm999.1

Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza, D., et al. (2010). Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: Proof of concept. *Hydrology and Earth System Sciences*, *14*(9), 1773–1785. https://doi.org/10.5194/hess-14-1773-2010

Mazrooei, A., Sankarasubramanian, A., & Wood, A. W. (2021). Potential in improving monthly streamflow forecasting through variational assimilation of observed streamflow. *Journal of Hydrology*, *600*, 126559. https://doi.org/10.1016/j.jhydrol.2021.126559

McLaughlin, D. (1995). Recent developments in hydrologic data assimilation. *Reviews of Geophysics*, *33*(S2), 977–984. https://doi.org/10.1029/95RG00740

McMillan, H. K., Hreinsson, E. Ö., Clark, M. P., Singh, S. K., Zammit, C., & Uddstrom, M. J. (2013). Operational hydrological data assimilation with the recursive ensemble Kalman filter. *Hydrology and Earth System Sciences*, *17*(1), 21–38. https://doi.org/10.5194/hess-17-21-2013

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., et al. (2017). A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data*, *4*(1), 170012. https://doi.org/10.1038/sdata.2017.12

Meng, C. L., Li, Z.-L., Zhan, X., Shi, J. C., & Liu, C. Y. (2009). Land surface temperature data assimilation and its impact on evapotranspiration estimates from the Common Land Model. *Water Resources Research*, *45*(2). https://doi.org/10.1029/2008WR006971

Merlin, O., Chehbouni, A., Boulet, G., & Kerr, Y. (2006). Assimilation of disaggregated microwave soil moisture into a hydrologic model using coarse-scale meteorological data. *Journal of Hydrometeorology*, *7*(6), 1308–1322. https://doi.org/10.1175/jhm552.1

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, *109*(D7). https://doi.org/10.1029/2003JD003823

Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, *139*(5), 1519–1535. https://doi.org/10.1175/2010mwr3570.1

Miyoshi, T., Kondo, K., & Imamura, T. (2014). The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophysical Research Letters*, *41*(14), 5264–5271. https://doi.org/10.1002/2014GL060863

Miyoshi, T., Kunii, M., Ruiz, J., Lien, G.-Y., Satoh, S., Ushio, T., et al. (2016). "Big data assimilation" revolutionizing severe weather prediction. *Bulletin of the American Meteorological Society*, *97*(8), 1347–1354. https://doi.org/10.1175/BAMS-D-15-00144.1

Montzka, C., Pauwels, V. R. N., Franssen, H.-J. H., Han, X., & Vereecken, H. (2012). Multivariate and multiscale data assimilation in terrestrial systems: A review. *Sensors*, *12*(12), 16291–16333. https://doi.org/10.3390/s121216291

Moody, A., & Woodcock, C. E. (1995). The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landscape Ecology*, *10*(6), 363–379. https://doi.org/10.1007/BF00130213

Moradkhani, H., DeChant, C. M., & Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, *48*(12), W12520. https://doi.org/10.1029/2012WR012144

Moradkhani, H., Hsu, K., Hong, Y., & Sorooshian, S. (2006). Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting. *Geophysical Research Letters*, *33*(12), L12401. https://doi.org/10.1029/2006GL026855

Moradkhani, H., Hsu, K.-L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*(5). https://doi.org/10.1029/2004WR003604

Moradkhani, H., Nearing, G., Abbaszadeh, P., & Pathiraja, S. (2019). Fundamentals of data assimilation and theoretical advances. In *Handbook of hydrometeorological ensemble forecasting*. Springer. https://doi.org/10.1007/978-3-642-40457-3_30-1

Moradkhani, H., & Sorooshian, S. (2008). General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In *Hydrological modelling and the water cycle: Coupling the atmospheric and hydrological models* (pp. 1–24). https://doi.org/10.1007/978-3-540-77843-1_1

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

National Research Council (NRC): Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research. (1991). *Four dimensional model assimilation of data: A strategy for the Earth system sciences*. National Academies Pres.

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, *55*, 110–118. https://doi.org/10.1016/j.cageo.2012.03.026

Nguyen, T. H., Ricci, S., Fatras, C., Piacentini, A., Delmotte, A., Lavergne, E., & Kettig, P. (2022). Improvement of flood extent representation with remote sensing data and data assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–22. https://doi.org/10.1109/TGRS.2022.3147429

Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Transactions on Cybernetics*, *50*(9), 3826–3839. https://doi.org/10.1109/tcyb.2020.2977374

Nie, W., Zaitchik, B. F., Rodell, M., Kumar, S. V., Arsenault, K. R., Li, B., & Getirana, A. (2019). Assimilating GRACE into a land surface model in the presence of an irrigation-induced groundwater trend. *Water Resources Research*, *55*(12), 11274–11294. https://doi.org/10.1029/2019WR025363

Niu, G.-Y., & Yang, Z.-L. (2007). An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *Journal of Geophysical Research*, *112*(D21), D21101. https://doi.org/10.1029/2007JD008674

Oaida, C. M., Reager, J. T., Andreadis, K. M., David, C. H., Levoe, S. R., Painter, T. H., et al. (2019). A high-resolution data assimilation framework for snow water equivalent estimation across the western United States and validation with the airborne snow observatory. *Journal of Hydrometeorology*, *20*(3), 357–378. https://doi.org/10.1175/jhm-d-18-0009.1

Pan, M., & Wood, E. F. (2009). A multiscale ensemble filtering system for hydrologic data assimilation. Part II: Application to land surface modeling with satellite rainfall forcing. *Journal of Hydrometeorology*, *10*(6), 1493–1506. https://doi.org/10.1175/2009JHM1155.1

Pan, M., Wood, E. F., McLaughlin, D. B., Entekhabi, D., & Luo, L. (2009). A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment. *Journal of Hydrometeorology*, *10*(3), 794–806. https://doi.org/10.1175/2009jhm1088.1

Pan, M., Wood, E. F., Wójcik, R., & McCabe, M. F. (2008). Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sensing of Environment*, *112*(4), 1282–1294. https://doi.org/10.1016/j.rse.2007.02.039

Pan, X., Li, X., Cheng, G., Chen, R., & Hsu, K. (2017). Impact analysis of climate change on snow over a complex mountainous region using weather research and forecast model (WRF) simulation and moderate resolution imaging spectroradiometer data (MODIS)-terra fractional snow cover products. *Remote Sensing*, *9*(8), 774. https://doi.org/10.3390/rs9080774

Panciera, R., Walker, J. P., Kalma, J. D., Kim, E. J., Hacker, J. M., Merlin, O., et al. (2008). The NAFE'05/CoSMOS data set: Toward SMOS soil moisture retrieval, downscaling, and assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(3), 736–745. https://doi.org/10.1109/TGRS.2007.915403

Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. P. (2014). Comparison of Ensemble Kalman Filter groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo simulation. *Advances in Water Resources*, *66*, 8–18. https://doi.org/10.1016/j.advwatres.2014.01.007

Parada, L. M., & Liang, X. (2004). Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models. *Journal of Geophysical Research*, *109*(D24), D24109. https://doi.org/10.1029/2004JD004745

Parrish, M. A., Moradkhani, H., & DeChant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*, *48*(3). https://doi.org/10.1029/2011WR011116

Pathiraja, S., Anghileri, D., Burlando, P., Sharma, A., Marshall, L., & Moradkhani, H. (2018a). Time-varying parameter models for catchments with land use change: The importance of model structure. *Hydrology and Earth System Sciences*, *22*(5), 2903–2919. https://doi.org/10.5194/hess-22-2903-2018

Pathiraja, S., Anghileri, D., Burlando, P., Sharma, A., Marshall, L., & Moradkhani, H. (2018b). Insights on the impact of systematic model errors on data assimilation performance in changing catchments. *Advances in Water Resources*, *113*, 202–222. https://doi.org/10.1016/j.advwatres.2017.12.006

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Advances in Water Resources*, *94*, 103–119. https://doi.org/10.1016/j.advwatres.2016.04.021

Pathiraja, S., Moradkhani, H., Marshall, L., Sharma, A., & Geenens, G. (2018). Data-driven model uncertainty estimation in data assimilation. *Water Resources Research*, *54*(2), 1252–1280. https://doi.org/10.1002/2018WR022627

Pathmathevan, M., Koike, T., Li, X., & Fujii, H. (2003). A simplified land data assimilation scheme and its application to soil moisture experiments in 2002 (SMEX02). *Water Resources Research*, *39*(12). https://doi.org/10.1029/2003WR002124

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., & De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, *251*(1), 88–102. https://doi.org/10.1016/S0022-1694(01)00440-1

Peischl, S., Walker, J., Rüdiger, C., Ye, N., Kerr, Y., Kim, E., et al. (2012). The AACES field experiments: SMOS calibration and validation across the Murrumbidgee river catchment. *Hydrology and Earth System Sciences*, *16*(6), 1697–1708. https://doi.org/10.5194/hess-16-1697-2012

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

Penny, S. G., Smith, T. A., Chen, T.-C., Platt, J. A., Lin, H.-Y., Goodliff, M., & Abarbanel, H. D. I. (2022). Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002843. https://doi.org/10.1029/2021MS002843

Peters-Lidard, C. D., Kumar, S. V., Mocko, D. M., & Tian, Y. (2011). Estimating evapotranspiration with land data assimilation systems. *Hydrological Processes*, *25*(26), 3979–3992. https://doi.org/10.1002/hyp.8387

Piazzi, G., Thirel, G., Campo, L., & Gabellani, S. (2018). A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment. *The Cryosphere*, *12*(7), 2287–2306. https://doi.org/10.5194/tc-12-2287-2018

Piazzi, G., Thirel, G., Perrin, C., & Delaigue, O. (2021). Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. *Water Resources Research*, *57*(4). https://doi.org/10.1029/2020WR028390

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., & Scoby, D. (2020). The land variational ensemble data assimilation framework: LAVENDAR v1.0.0. *Geoscientific Model Development*, *13*(1), 55–69. https://doi.org/10.5194/gmd-13-55-2020

Pipunic, R. C., Walker, J. P., & Western, A. (2008). Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sensing of Environment*, *112*(4), 1295–1305. https://doi.org/10.1016/j.rse.2007.02.038

Poghosyan, A., & Golkar, A. (2017). CubeSat evolution: Analyzing CubeSat capabilities for conducting science missions. *Progress in Aerospace Sciences*, *88*, 59–83. https://doi.org/10.1016/j.paerosci.2016.11.002

Popov, A. A., & Sandu, A. (2019). A Bayesian approach to multivariate adaptive localization in ensemble-based data assimilation with time-dependent extensions. *Nonlinear Processes in Geophysics*, *26*(2), 109–122. https://doi.org/10.5194/npg-26-109-2019

Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, *144*(1), 59–76. https://doi.org/10.1175/mwr-d-15-0163.1

Prakash, V., & Mishra, V. (2022). Soil moisture and streamflow data assimilation for streamflow prediction in the Narmada river basin. *Journal of Hydrometeorology*, *24*(8), 1377–1392. https://doi.org/10.1175/JHM-D-21-0139.1

Pugliese, A., Persiano, S., Bagli, S., Mazzoli, P., Parajka, J., Arheimer, B., et al. (2018). A geostatistical data-assimilation technique for enhancing macro-scale rainfall–runoff simulations. *Hydrology and Earth System Sciences*, *22*(9), 4633–4648. https://doi.org/10.5194/hess-22-4633-2018

Pulido, M., Tandeo, P., Bocquet, M., Carrassi, A., & Lucini, M. (2018). Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods. *Tellus A: Dynamic Meteorology and Oceanography*, *70*(1), 1–17. https://doi.org/10.1080/16000870.2018.1442099

Qin, C., Jia, Y., Su, Z., Zhou, Z., Qiu, Y., & Suhui, S. (2008). Integrating remote sensing information into a distributed hydrological model for improving water budget predictions in large-scale basins through data assimilation. *Sensors*, *8*(7), 4441–4465. https://doi.org/10.3390/s8074441

Qin, J., Liang, S., Yang, K., Kaihotsu, I., Liu, R., & Koike, T. (2009). Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal. *Journal of Geophysical Research*, *114*(D15), D15103. https://doi.org/10.1029/2008JD011358

Quilty, J. M., Sikorska-Senoner, A. E., & Hah, D. (2022). A stochastic conceptual-data-driven approach for improved hydrological simulations. *Environmental Modelling & Software*, *149*, 105326. https://doi.org/10.1016/j.envsoft.2022.105326

Quinn, J. C., & Abarbanel, H. D. (2011). Data assimilation using a GPU accelerated path integral Monte Carlo approach. *Journal of Computational Physics*, *230*(22), 8168–8178. https://doi.org/10.1016/j.jcp.2011.07.015

Raanes, P. N., Bocquet, M., & Carrassi, A. (2019). Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 53–75. https://doi.org/10.1002/qj.3386

Ran, Y. H., Li, X., Lu, L., & Li, Z. Y. (2012). Large-scale land cover mapping with the integration of multi-source information based on the Dempster–Shafer theory. *International Journal of Geographical Information Science*, *26*(1), 169–191. https://doi.org/10.1080/13658816.2011.577745

Raoult, N. M., Jupp, T. E., Cox, P. M., & Luke, C. M. (2016). Land-surface parameter optimisation using data assimilation techniques: The adJULES system V1.0. *Geoscientific Model Development*, *9*(8), 2833–2852. https://doi.org/10.5194/gmd-9-2833-2016

Rasmy, M., Koike, T., Kuria, D., Mirza, C. R., Li, X., & Yang, K. (2012). Development of the coupled atmosphere and land data assimilation system (caldas) and its application over the Tibetan Plateau. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(11), 4227–4242. https://doi.org/10.1109/TGRS.2012.2190517

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., et al. (2007). The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *133*(623), 347–362. https://doi.org/10.1002/qj.32

Reich, S. (2013). A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, *35*(4), A2013–A2024. https://doi.org/10.1137/130907367

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., & De Lannoy, G. J. M. (2017). Assessment of merra-2 land surface hydrology estimates. *Journal of Climate*, *30*(8), 2937–2960. https://doi.org/10.1175/jcli-d-16-0720.1

Reichle, R. H., Entekhabi, D., & McLaughlin, D. B. (2001). Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resources Research*, *37*(9), 2353–2364. https://doi.org/10.1029/2001WR000475

Reichle, R. H., Kumar, S. V., Mahanama, S. P. P., Koster, R. D., & Liu, Q. (2010). Assimilation of satellite-derived skin temperature observations into land surface models. *Journal of Hydrometeorology*, *11*(5), 1103–1122. https://doi.org/10.1175/2010jhm1262.1

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2001). Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1708–1718. https://doi.org/10.1109/36.942549

Ridler, M. E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A., & Madsen, H. (2014). Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI). *Environmental Modelling & Software*, *57*, 76–89. https://doi.org/10.1016/j.envsoft.2014.02.008

Rodell, M., Beaudoing, H. K., L'ecuyer, T., Olson, W. S., Famiglietti, J. S., Houser, P. R., et al. (2015). The observed state of the water cycle in the early twenty-first century. *Journal of Climate*, *28*(21), 8289–8318. https://doi.org/10.1175/JCLI-D-14-00555.1

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). 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

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999–1002. https://doi.org/10.1038/nature08238

Roy, A., Royer, A., & Turcotte, R. (2010). Improvement of springtime streamflow simulations in a boreal environment by incorporating snow-covered area derived from remote sensing data. *Journal of Hydrology*, *390*(1), 35–44. https://doi.org/10.1016/j.jhydrol.2010.06.027

Sabater, J. M., Rüdiger, C., Calvet, J.-C., Fritz, N., Jarlan, L., & Kerr, Y. (2008). Joint assimilation of surface soil moisture and LAI observations into a land surface model. *Agricultural and Forest Meteorology*, *148*(8), 1362–1373. https://doi.org/10.1016/j.agrformet.2008.04.003

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/jcli-d-12-00823.1

Sahoo, A. K., De Lannoy, G. J. M., Reichle, R. H., & Houser, P. R. (2013). Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. *Advances in Water Resources*, *52*, 19–33. https://doi.org/10.1016/j.advwatres.2012.08.007

Sawada, Y. (2020). Machine learning accelerates parameter optimization and uncertainty assessment of a land surface model. *Journal of Geophysical Research: Atmospheres*, *125*(20), e2020JD032688. https://doi.org/10.1029/2020JD032688

Sawada, Y., & Koike, T. (2016). Towards ecohydrological drought monitoring and prediction using a land data assimilation system: A case study on the Horn of Africa drought (2010–2011). *Journal of Geophysical Research: Atmospheres*, *121*(14), 8229–8242. https://doi.org/10.1002/2015JD024705

Sellers, P., Mintz, Y., Sud, Y. e. a., & Dalcher, A. (1986). A simple biosphere model (SiB) for use within general circulation models. *Journal of the Atmospheric Sciences*, *43*(6), 505–531. https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2

Seo, B.-C., Krajewski, W., & Quintero, F. (2021). Multi-scale hydrologic evaluation of the national water model streamflow data assimilation. *Journal of the American Water Resources Association*, *57*(6), 875–884. https://doi.org/10.1111/1752-1688.12955

Shahabadi, M. B., Bélair, S., Bilodeau, B., Carrera, M. L., & Garand, L. (2019). Impact of weak coupling between land and atmosphere data assimilation systems on environment and climate change Canada's global deterministic prediction system. *Weather and Forecasting*, *34*(6), 1741–1758. https://doi.org/10.1175/waf-d-19-0073.1

Shangguan, W., Dai, Y., Liu, B., Ye, A., & Yuan, H. (2012). A soil particle-size distribution dataset for regional land and climate modelling in China. *Geoderma*, *171–172*, 85–91. https://doi.org/10.1016/j.geoderma.2011.01.013

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*(13), 3088–3111. https://doi.org/10.1175/jcli3790.1

Shi, C., Xie, Z., Qian, H., Liang, M., & Yang, X. (2011). China land soil moisture EnKF data assimilation based on satellite remote sensing data. *Science China Earth Sciences*, *54*(9), 1430–1440. https://doi.org/10.1007/s11430-010-4160-3

Shi, P., Wang, B., He, Y., Lu, H., Yang, K., Xu, S., et al. (2022). Contributions of weakly coupled data assimilation–based land initialization to interannual predictability of summer climate over Europe. *Journal of Climate*, *35*(2), 517–535. https://doi.org/10.1175/JCLI-D-20-0506.1

Sini, F., Boni, G., Caparrini, F., & Entekhabi, D. (2008). Estimation of large-scale evaporation fields based on assimilation of remotely sensed land temperature. *Water Resources Research*, *44*(6). https://doi.org/10.1029/2006WR005574

Smith, P. J., Thornhill, G. D., Dance, S. L., Lawless, A. S., Mason, D. C., & Nichols, N. K. (2013). Data assimilation for state and parameter estimation: Application to morphodynamic modelling. *Quarterly Journal of the Royal Meteorological Society*, *139*(671), 314–327. https://doi.org/10.1002/qj.1944

Strebel, L., Bogena, H., Vereecken, H., Andreasen, M., Aranda, S., & Hendricks Franssen, H. J. (2023). Evapotranspiration prediction for European forest sites does not improve with assimilation of in-situ soil water content data. *EGUsphere*, 1–38. https://doi.org/10.5194/egusphere-2023-366

Su, H., Yang, Z.-L., Dickinson, R. E., Wilson, C. R., & Niu, G.-Y. (2010). Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. *Journal of Geophysical Research*, *115*(D10), D10104. https://doi.org/10.1029/2009JD013035

Sun, C., Walker, J. P., & Houser, P. R. (2004). A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research*, *109*(D8), D08108. https://doi.org/10.1029/2003JD003765

Sun, L., Nistor, I., & Seidou, O. (2015). Streamflow data assimilation in SWAT model using Extended Kalman Filter. *Journal of Hydrology*, *531*, 671–684. https://doi.org/10.1016/j.jhydrol.2015.10.060

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, *61*(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., et al. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, *115*(12), 3517–3529. https://doi.org/10.1016/j.rse.2011.08.014

Tang, M., Liu, Y., & Durlofsky, L. J. (2021). Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, *376*, 113636. https://doi.org/10.1016/j.cma.2020.113636

Tangdamrongsub, N., Han, S.-C., Yeo, I.-Y., Dong, J., Steele-Dunne, S. C., Willgoose, G., & Walker, J. P. (2020). Multivariate data assimilation of GRACE, SMOS, SMAP measurements for improved regional soil moisture and groundwater storage estimates. *Advances in Water Resources*, *135*, 103477. https://doi.org/10.1016/j.advwatres.2019.103477

Tian, J., Qin, J., Yang, K., Zhao, L., Chen, Y., Lu, H., et al. (2022). Improving surface soil moisture retrievals through a novel assimilation algorithm to estimate both model and observation errors. *Remote Sensing of Environment*, *269*, 112802. https://doi.org/10.1016/j.rse.2021.112802

Tian, S., Renzullo, L. J., Lerat, J., & Pipunic, R. C. (2022). Spatio-temporal model variance-covariance approach to assimilating streamflow observations into a distributed landscape water balance model. *Water Resources Research*, *58*(4), e2021WR031649. https://doi.org/10.1029/2021WR031649

Tian, X., Zhang, H., Feng, X., & Xie, Y. (2018). Nonlinear least squares En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation. *Monthly Weather Review*, *146*(1), 77–93. https://doi.org/10.1175/mwr-d-17-0050.1

Tolosana-Delgado, R., Egozcue, J. J., Sáchez-Arcilla, A., & Gómez, J. (2011). Wave height data assimilation using non-stationary kriging. *Computers & Geosciences*, *37*(3), 363–370. https://doi.org/10.1016/j.cageo.2010.05.019

Tong, J., Hu, B. X., & Yang, J. (2012). Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field. *Stochastic Environmental Research and Risk Assessment*, *26*(3), 467–478. https://doi.org/10.1007/s00477-011-0534-0

Trenberth, K., Fasullo, J., & Shepherd, T. (2015). Attribution of climate extreme events. *Nature Climate Change*, *5*(8), 725–730. https://doi.org/10.1038/nclimate2657

Troch, P., Paniconi, C., & McLaughlin, D. (2003). Catchment-scale hydrologic modeling and data assimilation. *Advances in Water Resources*, *26*(2), 131–135. https://doi.org/10.1016/S0309-1708(02)00087-8

Van Den Hurk, B. (2002). European LDAS established. *GEWEX News*, *12*(2), 9.

Van Der Merwe, R. (2004). *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. Oregon Health & Science University. Retrieved from https://dl.acm.org/doi/abs/10.5555/1037398

Vanella, D., Busato, L., Boaga, J., Cassiani, G., Binley, A., Putti, M., et al. (2016). Micro 3D ERT tomography for data assimilation modelling of active root zone (p. EPSC2016-4608).

Van Leeuwen, P. J. (2017). Particle Filters for nonlinear data assimilation in high-dimensional systems. *Annales de la Faculté des Sciences de Toulouse, Mathématiques*, *26*(4), 1051–1085. https://doi.org/10.5802/afst.1560

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., et al. (2022). Soil hydrology in the Earth system. *Nature Reviews Earth & Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6

Vereecken, H., Huisman, J.-A., Hendricks Franssen, H.-J., Brüggemann, N., Bogena, H. R., Kollet, S., et al. (2015). Soil hydrology: Recent methodological advances, challenges, and perspectives. *Water Resources Research*, *51*(4), 2616–2633. https://doi.org/10.1002/2014WR016852

Visweshwaran, R., Ramsankaran, R., Eldho, T. I., & Lakshmivarahan, S. (2021). Sensitivity based soil moisture assimilation for improved streamflow forecast using a novel Forward Sensitivity Method (FSM) approach. *Water Resources Research*. https://doi.org/10.1002/essoar.10507814.1

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2003). A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8). https://doi.org/10.1029/2002WR001642

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, *41*(1). https://doi.org/10.1029/2004WR003059

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T., & Bouten, W. (2006). Application of stochastic parameter optimization to the Sacramento soil moisture accounting model. *Journal of Hydrology*, *325*(1), 288–307. https://doi.org/10.1016/j.jhydrol.2005.10.041

Vrugt, J. A., Ter, B. C. J., Diks, C. G., & Schoups, G. (2013). Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, *51*, 457–478. https://doi.org/10.1016/j.advwatres.2012.04.002

Vrugt, J. A., & Ter Braak, C. J. (2011). DREAM (D): An adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*, *15*(12), 3701–3713. https://doi.org/10.5194/hess-15-3701-2011

Walker, J. P., & Houser, P. R. (2001). A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *Journal of Geophysical Research*, *106*(D11), 11761–11774. https://doi.org/10.1029/2001JD900149

Waller, J. A., Dance, S. L., Lawless, A. S., & Nichols, N. K. (2014). Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, *66*(1), 23294. https://doi.org/10.3402/tellusa.v66.23294

Wang, C., Duan, Q., Tong, C. H., Di, Z., & Gong, W. (2016). A GUI platform for uncertainty quantification of complex dynamical models. *Environmental Modelling & Software*, *76*, 1–12. https://doi.org/10.1016/j.envsoft.2015.11.004

Wang, D., & Cai, X. (2008). Robust data assimilation in hydrological modeling—A comparison of Kalman and H-infinity filters. *Advances in Water Resources*, *31*(3), 455–472. https://doi.org/10.1016/j.advwatres.2007.10.001

Wang, F., Huang, G. H., Fan, Y., & Li, Y. P. (2022). Development of a disaggregated multi-level factorial hydrologic data assimilation model. *Journal of Hydrology*, *610*, 127802. https://doi.org/10.1016/j.jhydrol.2022.127802

Wang, L., Koike, T., Yang, K., Jackson, T. J., Bindlish, R., & Yang, D. (2009). Development of a distributed biosphere hydrological model and its evaluation with the Southern Great Plains Experiments (SGP97 and SGP99). *Journal of Geophysical Research*, *114*(D8), D08107. https://doi.org/10.1029/2008JD010800

Wang, S., Liang, X., & Nan, Z. (2011). How much improvement can precipitation data fusion achieve with a Multiscale Kalman Smoother-based framework? *Water Resources Research*, *47*(3). https://doi.org/10.1029/2010WR009953

Wang, W., Cui, W., Wang, X., & Chen, X. (2016). Evaluation of gldas-1 and gldas-2 forcing data and Noah model simulations over China at the monthly scale. *Journal of Hydrometeorology*, *17*(11), 2815–2833. https://doi.org/10.1175/jhm-d-15-0191.1

Wang, W., Ren, K., Duan, B., Zhu, J., Li, X., Ni, W., et al. (2024). A four-dimensional variational constrained neural network-based data assimilation method. *Journal of Advances in Modeling Earth Systems*, *16*(1), e2023MS003687. https://doi.org/10.1029/2023MS003687

Wang, Y., Shi, L., Xu, T., Zhang, Q., Ye, M., & Zha, Y. (2021). A nonparametric sequential data assimilation scheme for soil moisture flow. *Journal of Hydrology*, *593*, 125865. https://doi.org/10.1016/j.jhydrol.2020.125865

Ward, J., Evans, A., & Malleson, N. (2016). *Dynamic calibration of agent-based models using data assimilation* (Vol. 3, p. 150703). Royal Society Open Science. https://doi.org/10.1098/rsos.150703

Wiltshire, A. J., Burke, E. J., Chadburn, S. E., Jones, C. D., Cox, P. M., Davies-Barnard, T., et al. (2021). JULES-CN: A coupled terrestrial carbon–nitrogen scheme (JULES vn5. 1). *Geoscientific Model Development*, *14*(4), 2161–2186. https://doi.org/10.5194/gmd-14-2161-2021

Wongchuig-Correa, S., de Paiva, R. C. D., Biancamaria, S., & Collischonn, W. (2020). Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models. *Journal of Hydrology*, *590*, 125473. https://doi.org/10.1016/j.jhydrol.2020.125473

Wu, G., Bo, D., & Zheng, X. (2016). Soil moisture assimilation using a modified ensemble transform Kalman filter based on station observations in the Hai river basin. *Advances in Meteorology*, *2016*, 1–12. https://doi.org/10.1155/2016/4569218

Xia, Y., Ek, M. B., Mocko, D., Peters-Lidard, C. D., Sheffield, J., Dong, J., & Wood, E. F. (2014). Uncertainties, correlations, and optimal blends of drought indices from the NLDAS multiple land surface model ensemble. *Journal of Hydrometeorology*, *15*(4), 1636–1650. https://doi.org/10.1175/JHM-D-13-058.1

Xia, Y., Hao, Z., Shi, C., Li, Y., Meng, J., Xu, T., et al. (2019). Regional and global land data assimilation systems: Innovations, challenges, and prospects. *Journal of Meteorological Research*, *33*(2), 159–189. https://doi.org/10.1007/s13351-019-8172-4

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research*, *117*(D3), D03109. https://doi.org/10.1029/2011JD016048

Xie, X., & Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources*, *33*(6), 678–690. https://doi.org/10.1016/j.advwatres.2010.03.012

Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., & Liu, Y. (2019). Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. *Journal of Hydrology*, *568*, 758–768. https://doi.org/10.1016/j.jhydrol.2018.11.038

Xu, J., & Shu, H. (2014). Assimilating MODIS-based albedo and snow cover fraction into the Common Land Model to improve snow depth simulation with direct insertion and deterministic ensemble Kalman filter methods. *Journal of Geophysical Research: Atmospheres*, *119*(18), 10684–10701. https://doi.org/10.1002/2014JD022012

Xu, L., Abbaszadeh, P., Moradkhani, H., Chen, N., & Zhang, X. (2020). Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. *Remote Sensing of Environment*, *250*, 112028. https://doi.org/10.1016/j.rse.2020.112028

Xu, T., Bateni, S. M., Liang, S., Entekhabi, D., & Mao, K. (2014). Estimation of surface turbulent heat fluxes via variational assimilation of sequences of land surface temperatures from Geostationary Operational Environmental Satellites. *Journal of Geophysical Research: Atmospheres*, *119*(18), 10780–10798. https://doi.org/10.1002/2014JD021814

Xu, T., He, X., Bateni, S. M., Auligne, T., Liu, S., Xu, Z., et al. (2019). Mapping regional turbulent heat fluxes via variational assimilation of land surface temperature data from polar orbiting satellites. *Remote Sensing of Environment*, *221*, 444–461. https://doi.org/10.1016/j.rse.2018.11.023

Xu, X. Y., Zhang, X. S., Fang, H. W., Lai, R. X., Zhang, Y. F., Huang, L., & Liu, X. B. (2017). A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. *Environmental Modelling & Software*, *88*, 151–167. https://doi.org/10.1016/j.envsoft.2016.11.010

Yang, F., Lu, H., Yang, K., He, J., Wang, W., Wright, J. S., et al. (2017). Evaluation of multiple forcing data sets for precipitation and shortwave radiation over major land areas of China. *Hydrology and Earth System Sciences*, *21*(11), 5805–5821. https://doi.org/10.5194/hess-21-5805-2017

Yang, K., Koike, T., Kaihotsu, I., & Qin, J. (2009). Validation of a dual-pass microwave land data assimilation system for estimating surface soil moisture in semiarid regions. *Journal of Hydrometeorology*, *10*(3), 780–793. https://doi.org/10.1175/2008jhm1065.1

Yang, K., Watanabe, T., Koike, T., Li, X., Fujii, H., Tamagawa, K., et al. (2007). Auto-calibration system developed to assimilate amsr-e data into a land surface model for estimating soil moisture and the surface energy budget. *Journal of the Meteorological Society of Japan. Series II*, *85A*, 229–242. https://doi.org/10.2151/jmsj.85A.229

Yang, K., Zhu, L., Chen, Y., Zhao, L., Qin, J., Lu, H., et al. (2016). Land surface model calibration through microwave data assimilation for improving soil moisture simulations. *Journal of Hydrology*, *533*, 266–276. https://doi.org/10.1016/j.jhydrol.2015.12.018

Ye, N., Walker, J. P., Wu, X., Jeu, R. D., Gao, Y., Jackson, T. J., et al. (2021). The soil moisture active passive experiments: Validation of the smap products in Australia. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(4), 2922–2939. https://doi.org/10.1109/TGRS.2020.3007371

Yeung, Y. H., Barajas-Solano, D. A., & Tartakovsky, A. M. (2022). Physics-informed machine learning method for large-scale data assimilation problems. *Water Resources Research*, *58*(5). https://doi.org/10.1029/2021WR031023

Yin, J., Zhan, C., & Ye, W. (2016). An experimental study on evapotranspiration data assimilation based on the hydrological model. *Water Resources Management*, *30*(14), 5263–5279. https://doi.org/10.1007/s11269-016-1485-5

Zaitchik, B. F., & Rodell, M. (2009). Forward-looking assimilation of modis-derived snow-covered area into a land surface model. *Journal of Hydrometeorology*, *10*(1), 130–148. https://doi.org/10.1175/2008jhm1042.1

Zhang, C., Abbaszadeh, P., Xu, L., Moradkhani, H., Duan, Q., & Gong, W. (2021). A combined optimization-assimilation framework to enhance the predictive skill of community land model. *Water Resources Research*, *57*(12). https://doi.org/10.1029/2021WR029879

Zhang, D., Madsen, H., Ridler, M. E., Kidmose, J., Jensen, K. H., & Refsgaard, J. C. (2016). Multivariate hydrological data assimilation of soil moisture and groundwater head. *Hydrology and Earth System Sciences*, *20*(10), 4341–4357. https://doi.org/10.5194/hess-20-4341-2016

Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., & Franssen, H.-J. H. (2018). Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model. *Advances in Water Resources*, *111*, 224–238. https://doi.org/10.1016/j.advwatres.2017.11.003

Zhang, H., Xue, J., Zhu, G., Zhuang, S., Wu, X., & Zhang, F. (2004). Application of direct assimilation of ATOVS microwave radiances to typhoon track prediction. *Advances in Atmospheric Sciences*, *21*(2), 283–290. https://doi.org/10.1007/BF02915715

Zhang, M., & Zhang, F. (2012). E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Monthly Weather Review*, *140*(2), 587–600. https://doi.org/10.1175/mwr-d-11-00023.1

Zhang, Q., Shi, L., Holzman, M., Ye, M., Wang, Y., Carmona, F., & Zha, Y. (2019). A dynamic data-driven method for dealing with model structural error in soil moisture data assimilation. *Advances in Water Resources*, *132*, 103407. https://doi.org/10.1016/j.advwatres.2019.103407

Zhang, Y., Hou, J., Gu, J., Huang, C., & Li, X. (2017). Swat-based hydrological data assimilation system (swat-hdas): Description and case application to river basin-scale hydrological predictions. *Journal of Advances in Modeling Earth Systems*, *9*(8), 2863–2882. https://doi.org/10.1002/2017MS001144

Zhang, Y., Hou, J., Gu, J., Huang, C., & Li, X. (2019). Soil moisture data assimilation to estimate irrigation water use. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3670–3690. https://doi.org/10.1029/2019MS001797

Zhang, Y.-F., Hoar, T. J., Yang, Z.-L., Anderson, J. L., Toure, A. M., & Rodell, M. (2014). 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

Zhao, L., Yang, Z. L., & Hoar, T. J. (2016). Global soil moisture estimation by assimilating AMSR-E brightness temperatures in a coupled CLM4–RTM–DART system. *Journal of Hydrometeorology*, *17*(9), 2431–2454. https://doi.org/10.1175/jhm-d-15-0218.1

Zheng, X. (2009). An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, *26*(1), 154–160. https://doi.org/10.1007/s00376-009-0154-5

Zhu, G., Li, X., Ma, J., Wang, Y., Liu, S., Huang, C., et al. (2018). A new moving strategy for the sequential Monte Carlo approach in optimizing the hydrological model parameters. *Advances in Water Resources*, *114*, 164–179. https://doi.org/10.1016/j.advwatres.2018.02.007

Zhu, G., Li, X., Su, Y. H., Zhang, K., Bai, Y., Ma, J., et al. (2014). Simultaneously assimilating multivariate data sets into the two-source evapotranspiration model by Bayesian approach: Application to spring maize in an arid region of northwestern China. *Geoscientific Model Development*, *7*(4), 1467–1482. https://doi.org/10.5194/gmd-7-1467-2014

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. *Journal of Hydrology*, *549*, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

Zupanski, D., Zhang, S. Q., Zupanski, M., Hou, A. Y., & Cheung, S. H. (2011). A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations. *Journal of Hydrometeorology*, *12*(1), 118–134. https://doi.org/10.1175/2010jhm1271.1

## Erratum

The originally published version of this article omitted the contributions of coauthor Jinliang Hou from the Author Contributions list. Dr. Hou contributed to Conceptualization, Investigation, Visualization, Writing – Original Draft Preparation, and Writing – Review & Editing. The errors have been corrected, and this may be considered the authoritative version of record.