AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research®

RESEARCH ARTICLE
10.1029/2025WR041898

On the Importance of the Jacobian of a DNN Observation Operator in Land Data Assimilation

Xu Shan[1,2], **Susan Steele-Dunne**[1], **Sebastian Hahn**[3], **Wolfgang Wagner**[3], **Bertrand Bonan**[4], and **Ou Ku**[5]

[1]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft, Delft, The Netherlands, [2]Department of Water Management, Faculty of Civil Engineering and Geosciences, TU Delft, Delft, The Netherlands, [3]Department of Geodesy and Geoinformation (GEO), Vienna University of Technology, Vienna, Austria, [4]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France, [5]Netherlands eScience Center, Amsterdam, The Netherlands

Key Points:

- Assimilating C-band backscatter into the ISBA-A-gs LSM using a DNN as an observation operator improves estimates of surface soil moisture
- Comparison between DNN and water cloud model shows that the assimilation using DNN is sub-optimal although surface soil moisture is improved
- The primary causes of inconsistent Jacobians of the DNN are the model bias and redundant input variables

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Shan,
x.shan-2@tudelft.nl

Citation:

Shan, X., Steele-Dunne, S., Hahn, S., Wagner, W., Bonan, B., & Ku, O. (2026). On the importance of the Jacobian of a DNN observation operator in land data assimilation. *Water Resources Research*, *62*, e2025WR041898. https://doi.org/10.1029/2025WR041898

Received 13 AUG 2025
Accepted 30 JUN 2026

Author Contributions:

Conceptualization: Xu Shan, Susan Steele-Dunne, Wolfgang Wagner
Formal analysis: Xu Shan, Susan Steele-Dunne, Wolfgang Wagner
Funding acquisition: Susan Steele-Dunne
Investigation: Xu Shan, Susan Steele-Dunne
Methodology: Xu Shan, Susan Steele-Dunne, Wolfgang Wagner, Ou Ku
Project administration: Susan Steele-Dunne
Resources: Susan Steele-Dunne, Sebastian Hahn, Wolfgang Wagner

Abstract A recent study (Shan et al., 2024, https://doi.org/10.1016/j.rse.2024.114167) showed that using a Deep-Neural-Network (DNN)-based observation operator in land data assimilation (DA) does not guarantee improvements in surface soil moisture (WG2). Here, we conduct a synthetic experiment to explain this outcome by testing whether DNN-based operators reproduce physically consistent Jacobians (sensitivities) required by DA. The ISBA-A-gs land surface model is perturbed to generate "synthetic true" WG2 and leaf area index (LAI), and a Water Cloud Model (WCM) is used to generate synthetic backscatter. Two DNNs are trained taking simulated states from the open loop run as input and the synthetic backscatter as output. The first is trained on WG2 and LAI while the second uses LAI and soil moisture from multiple layers, that is including redundant inputs. The synthetic observations are then assimilated using the two DNNs and the WCM as observation operators. Results suggest that the assimilation using a DNN-based observation operator improves the estimates of WG2. However, DNN Jacobians are no longer physically plausible when the model simulations used as training inputs contain errors relative to the "true" data, or when redundant input variables are included. This confirms that a strong predictive performance does not automatically reflect accurate representation of underlying physical sensitivities which prove problematic in subsequent DA. This study has important implications for deep-learning-based DA and Earth system models (ESM): DA or ESM trained on reanalysis or simulations may learn inaccurate Jacobians if input contains errors and redundant variables. This potential limitation could remain hidden when evaluation focuses primarily on predictive performance.

Plain Language Summary Machine learning is increasingly used to help constraining land surface models (LSMs) with satellite measurements. However, Shan et al. (2024, https://doi.org/10.1016/j.rse.2024.114167) showed that using a Deep Neural Network (DNN) in a data assimilation (DA) system does not always lead to better SSM estimates. Here we conduct a synthetic experiment where the true relationship between soil moisture, vegetation, and microwave backscatter is known. This enables the evaluation of DNN-derived Jacobians—something that is not feasible in real-world scenarios. We found that good prediction performance alone was not enough: the DNN could match the observations well while still represent the underlying relationships incorrectly, which limited how much DA improved SSM estimates. This is because DNN "overfits" when model simulations used as training input have error compared to "true" data. This may raise concerns for generalizability of deep-learning-based DA and ESMs trained on model simulations or reanalysis data. We also show that including extra inputs such as deeper soil moisture can slightly improve prediction performance but can also contaminate the sensitivities that are needed for assimilation. The results show that machine learning in DA and ESM applications should be judged not only by prediction performance, but also by whether they capture physically meaningful relationships.

1. Introduction

Land data assimilation schemes assimilate satellite observations into land surface models (LSMs) to constrain the soil and vegetation dynamics (Kumar et al., 2022; Lannoy et al., 2022). Several studies show that the assimilation of soil moisture products retrieved from active (Advanced SCATterometer (ASCAT) (Albergel et al., 2017; Wagner et al., 1999), Sentinel-1 (Lievens, Reichle, et al., 2017)) and passive microwave remote sensing

Software: Xu Shan, Susan Steele-Dunne, Bertrand Bonan, Ou Ku
Supervision: Susan Steele-Dunne
Validation: Xu Shan
Visualization: Xu Shan
Writing – original draft: Xu Shan
Writing – review & editing: Xu Shan, Susan Steele-Dunne, Sebastian Hahn, Wolfgang Wagner, Bertrand Bonan, Ou Ku

observations (Advanced Microwave Scanning Radiometer for Earth observation science (AMSR-E) (Draper et al., 2012; Imaoka et al., 2010), the Soil Moisture Ocean Salinity (SMOS) mission (Kerr et al., 2010; Martens et al., 2016) and Soil Moisture Active Passive (SMAP) mission (Entekhabi et al., 2010; Seo et al., 2021)) improves the estimates of surface soil moisture (Aires et al., 2021; Albergel et al., 2012; Brocca et al., 2010; Seo et al., 2021), runoff (Draper et al., 2011) and evapotranspiration (Albergel et al., 2017).

However, retrieved products might be physically inconsistent with the LSMs due to the use of radiative transfer or empirical models with underlying assumptions of parameter, land surface, and climate characteristics that may deviate from that of the LSMs (Kumar et al., 2022; Lannoy et al., 2022). Therefore, directly assimilating radar backscatter (σ^o) or radiometer brightness temperature (Tb) offers an alternative means to constrain the LSMs (Balsamo et al., 2006; Crow & Wood, 2003; Han et al., 2014; Lievens, Martens, et al., 2017; Reichle et al., 2002). This approach can mitigate the need for CDF-matching and bias correction (Fairbairn et al., 2017; Leroux et al., 2018) and also avoids potential cross-correlated errors between retrievals and model simulations (De Lannoy & Reichle, 2016; Lievens, Martens, et al., 2017). Additionally, radar observables contain information on soil water dynamics and vegetation phenology (Steele-Dunne et al., 2019), allowing for the simultaneous updates of multiple state variables (surface soil moisture and vegetation states).

The advanced Scatterometer (ASCAT) on the Metop series of satellite provides data in one polarization (VV). ASCAT VV backscatter provides valuable insights to vegetation dynamics across a wide range of cover types (Frolking et al., 2005; Jarlan et al., 2002; Petchiappan et al., 2021; Schroeder et al., 2016; Steele-Dunne et al., 2019). The assimilation of σ^o and Tb requires an observation operator which links the states simulated by LSMs to the microwave remote sensing observables. Radiative transfer models (RTMs), like the water cloud model (WCM, Attema and Ulaby (1978)), the $\tau - \omega$ model and the L-band Microwave Emission from the Biosphere radiative transfer model (L-MEB, Wigneron et al. (2007)), are often used as the observation operator. Lievens, Martens, et al. (2017) compared the assimilation of ASCAT normalized backscatter (σ^o_{40}) and SMOS Tb into the Global Land Evaporation Amsterdam Model (GLEAM). They calibrated the WCM and L-MEB to forward model ASCAT σ^o and SMOS Tb. The study demonstrated that assimilating ASCAT σ^o_{40} improves the model skill in estimating surface soil moisture and land evaporation. Baguis et al. (2022) assimilated ASCAT σ^o_{40} in the hydrological model SCHEME using the WCM as an observation operator. They showed that assimilation of ASCAT σ^o_{40} reduces the RMSE in the estimates of streamflow, compared to the assimilation of the retrieved H-SAF H07 Surface Soil Moisture (SSM) product. However, they also found that the performance was still limited by bias in ensemble simulations. Modanesi et al. (2021) calibrated the WCM with simulations of SSM and LAI from LSM Noah-MP and then assimilated Sentinel-1 backscatter into Noah-MP to improve the SSM and LAI estimates (Modanesi et al., 2022).

In passive microwave remote sensing, De Lannoy and Reichle (2016) assimilated SMOS Tb into the GEOS-5 Catchment Land Surface Model using the $\tau - \omega$ model as the observation operator and compared the results to those obtained by assimilating SMOS retrieval products. The assimilation of Tb showed larger innovations and local differences in performance, which were attributed to differences in how the Tb and SM observations were masked, and to the different assumptions made in the retrieval model versus the forward model. A similar DA system is used for the SMAP Level 4 soil moisture product (Reichle et al., 2017, 2021).

However, RTMs require moisture content or dielectric properties of soil and vegetation cover, as well as descriptions of size, shape, orientation, and distributions of scatterers in the canopy. Unfortunately, these parameters are generally not simulated by LSM or available globally. Furthermore, despite their complexity, these models still have a highly idealized and simplified representation of vegetation and the distribution of water within it (Kumar et al., 2022; Lannoy et al., 2022). Therefore, many studies proposed the use of machine learning (ML) algorithms as a surrogate model to forward model the σ^o and Tb (de Roos et al., 2023; Shamambo et al., 2019; Shan et al., 2022). ML can model the indirect relationships between LSM states and microwave observables more efficiently, without explicitly parameterizing the complex interactions (Lannoy et al., 2022; Shan et al., 2022). Aires et al. (2021) investigated applying a neural network (NN) for retrieval versus forward modeling to assimilate ASCAT σ^o_{40} into an LSM. Corchia et al. (2023) assimilated ASCAT σ^o_{40} into ISBA-A-gs by using an NN as the observation operator. In this study, the NN was trained on surface soil moisture and surface soil temperature from ISBA-A-gs, and LAI observations from the PROBA-V satellite.

10.1029/2025WR041898

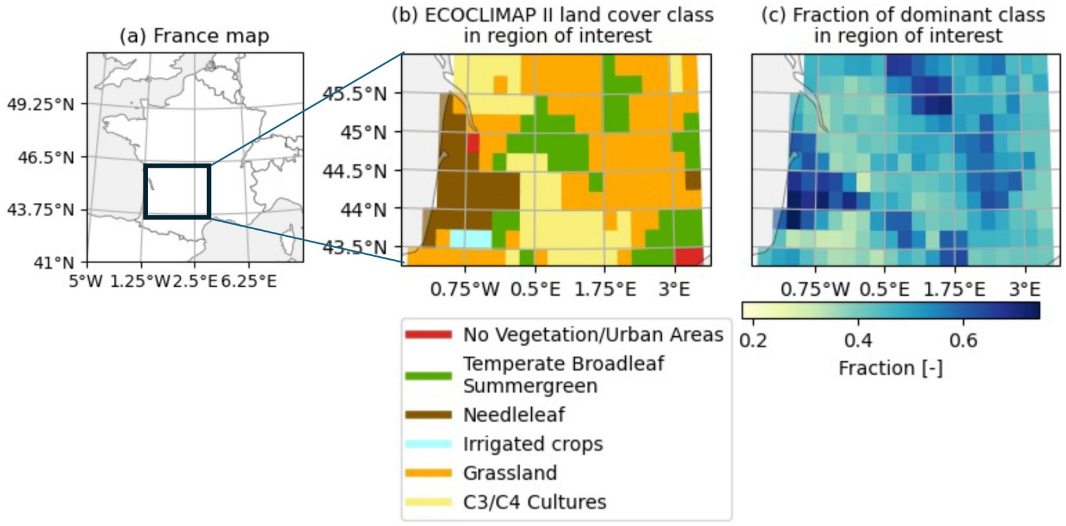

Figure 1. (a) Location of the study region of interest; (b) the dominant vegetation types in each 0.25° × 0.25° GPI; (c) the fraction of the dominant vegetation types in each GPI based on ECOCLIMAP II data.

A recent study demonstrated that the Jacobian terms of machine-learning-based observation operator might limit the efficiency of the DA (Shan et al., 2024). In particular, inconsistencies among the terms of the Jacobian of ASCAT observables to ISBA-A-gs land surface states led to the neutral impact of the DA of ASCAT normalized backscatter and slope. This highlighted the value of examining the physical plausibility and consistency of the machine-learning-based observation operator in terms of the sensitivities of microwave remote sensing observables to land surface states (Shan et al., 2024). The current study extends the work of Shan et al. (2024) by running synthetic data assimilation experiments to improve our understanding of the influence of choices related to the training of the DNN on its performance in the assimilation framework. ML approaches are frequently careless with regards to the number/types of inputs considered. More is generally considered better. This study considers the potential danger in that attitude with regard to observation operators, where more inputs could lead to incorrect sensitivities.

Specifically, the two DNNs are trained with different input choices, to test the hypothesis that assimilating synthetic backscatter using DNNs as the observation operator can lead to suboptimal performance if (a) the OL data used to train the DNN is biased compared to synthetic truth, and/or (b) the presence of redundant input variables introduces inconsistencies across the terms of the DNN-based Jacobians. In particular, introducing land surface variables indirectly linked to the training target may improve the performances, but may also introduce inconsistencies in the DNN-based Jacobian terms, that is a "better" estimation performance does not yield a "better" DA performance or Jacobian which is more consistent with scattering mechanisms.

2. Data

2.1. Study Area

The experiment is conducted over a region in southwestern France as shown in Figure 1a, extending from 43°N to 46°N and 1°W to 3.5°E. The domain contains 211 grid points (GPIs), each of which represents an area of 0.25° by 0.25°. Figure 1 maps the dominant land cover type (b) and the fraction of that cover type in the GPI (c). Most GPIs in this region are dominated by C3/C4 agriculture and grasslands, with the remainder primarily covered by Broadleaf and Needleleaf forests. The soil type information from ECOCLIMAP II data can be found in Figure S1 in Supporting Information S1. Most grid points in this region are classed as sandy soil.

2.2. ASCAT Normalized Backscatter

ASCAT is on board the Metop series of satellites operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) (Hahn et al., 2017). It operates a frequency of 5.255 GHz in C-band and provides (VV) co-polarized radar backscatter. The ASCAT sensors have two sets of antennas with different incidence angles (34° to 65° and 22° to 55° (Hahn et al., 2017)). The measured backscatter is normalized to a reference incidence angle of 40° by using a second-order polynomial describing the relationship between incidence angle and backscatter (Hahn et al., 2017; Melzer, 2013; Wagner et al., 1999). The ASCAT normalized backscatter is available at 0.25° resolution. The normalized backscatter data set consists of data from ascending (09:30 a.m.) and descending orbits (09:30 p.m.).

3. Methodology

3.1. ISBA-A-gs Model

The ISBA-A-gs (Interactions between Soil, Biosphere and Atmosphere (Noilhan & Mahfouf, 1996; Noilhan & Planton, 1989)) land surface model (LSM) simulates water, energy, and carbon exchanges at the land surface (Albergel et al., 2017). This study uses the CO_2-responsive version of ISBA-A-gs within the offline SURFEX platform (version 8.1) (Albergel et al., 2017, 2018; Masson et al., 2013). The simulations of dynamics of plant physiological states in land-atmospheric interactions are run with the “NIT” plant biomass monitoring option (Calvet et al., 1998, 2004, 2007). The current ISBA-A-gs model utilizes Jacob's biochemical A-gs model (Jacobs et al., 1996) to simulate the dynamics of photosynthetic processes responding to changing atmospheric conditions (Albergel et al., 2017). A multi-layer diffusion scheme is used to simulate the dynamics of soil water in different layers (Albergel et al., 2017; Leroux et al., 2018). The hydraulic parameters used in ISBA-A-gs are included in Text S1 and S2 in Supporting Information S1. More details can be found in Decharme et al. (2011). The ISBA-A-gs parameters are defined for 19 generic land surface patches as described in Table S2 in Supporting Information S1.

The model was forced by the latest ERA5 (ECMWF Reanalysis v5) atmospheric reanalysis (Hersbach et al., 2020) from 1996 to 2019. The hourly meteorological forcing data are available on 0.25° × 0.25° grid and include precipitation (Prcp), 2 m air temperature, 2 m specific humidity, wind speed, wind direction, surface pressure, downward direct shortwave radiation, downward diffuse radiation, downward longwave radiation, snowfall rate, and CO_2 concentration. All ERA5 atmospheric variables were interpolated using bilinear interpolation to match the grid points of ISBA. This step was necessary because the center points of the ERA5 grid do not match those of the ISBA-A-gs simulations or ASCAT data. The ISBA parameters are those used in Parrens et al. (2014) and Decharme et al. (2011). To initialize the model, the spinup is done by running the model using 20 repetitions of 1996 forcing data. The land-surface variables simulated at the end of this spinup are then used as the initial conditions for the open-loop run. The open-loop simulation was obtained using the ERA5 forcing data from 1997 to 2019. Then the data assimilation scheme was run from 1 Jan 2017 to 31 Dec 2019.

3.2. Water Cloud Model

In the Water Cloud Model (WCM) (Attema & Ulaby, 1978), the backscatter from a vegetated soil surface is represented by the sum of vegetation backscatter (σ^o_{veg}) and the soil backscatter (σ^o_{soil}) which is subject to two-way attenuation by the vegetation (Lievens, Martens, et al., 2017).

$$\sigma^o = \sigma^o_{\text{veg}} + T^2 \sigma^o_{\text{soil}} \tag{1}$$

with

$$\sigma^o_{\text{soil (dB)}} = 10 \log_{10} \sigma^o_{\text{soil}} = C + D \times \text{SSM} \tag{2}$$

The vegetation transmissivity (T) depends on the incidence angle (θ) and the vegetation optical depth (VOD):

$$T = \exp\left(-\frac{\text{VOD}}{\cos\theta}\right) \tag{3}$$

The vegetation is represented as a layer with uniformly distributed water droplets (Bindlish & Barros, 2001), characterized by a bulk vegetation descriptor V_1:

$$\sigma^o_{\text{veg}} = A \times V_1 \times \left(1 - T^2\right) \times \cos\theta \tag{4}$$

with V_1 being the bulk vegetation descriptor accounting for canopy backscatter and A being a fitting parameter (Lievens, Martens, et al., 2017) and VOD $= B \times V_2$ (Shamambo et al., 2019). The VOD can be related to the vegetation water content (VWC) or to proxies of the vegetation water content (VWC) like LAI. In this study, both V_1 and V_2 are set as LAI following (Modanesi et al., 2021).

The WCM contains the parameters A, B, C, and D which vary per GPI. The A and B parameters are related to vegetation properties while C and D are related to soil properties (Shamambo et al., 2019). Parameter A quantifies the effect of direct vegetation scattering while B is the measure of the vegetation attenuation effect (Lievens, Martens, et al., 2017). Parameters A and B are dimensionless and vary with canopy types and the radar configuration. Parameter C (dB) is the value of the backscatter for perfectly dry soil and is affected by the surface roughness and incidence angle. Parameter D (dB/m^3/m^3) describes the sensitivity of backscatter to soil moisture. A, B, C, and D also depend on the radar configuration.

3.3. Simplified Extended Kalman Filter

The Simplified Extended Kalman Filter (SEKF) is used to assimilate the ASCAT observables into the LSM ISBA-A-gs. The update equation of SEKF for a single GPI follows:

$$\widetilde{\mathbf{x}}^a_k = \widetilde{\mathbf{x}}^f_k + \widetilde{\mathbf{K}}_k\left(\mathbf{y}^o_k - \mathcal{H}\left(\mathbf{x}^f_k\right)\right) \tag{5}$$

where subscripts a, f, o indicate the analysis, forecast, and observation, respectively. The term $\widetilde{\mathbf{x}}_k$ represents the control vector or state vector of dimension (nbp, nbv) computed at time step k, in which nbp indicates the number of patches ISBA-A-gs simulates and nbv is the number of states. The term $\mathbf{x}$ represents the state vector aggregated over different patches which contains LAI, WG2 (0.01–0.04 m), WG3 (0.04–0.1 m), WG4 (0.1–0.2 m), WG5 (0.2–0.4 m), WG6 (0.4–0.6 m), WG7 (0.6–0.8 m), WG8 (0.8–1.0 m) as described in Table S2 in Supporting Information S1. $\mathbf{y}^o$ denotes the observation vector of dimension nbo. $\mathcal{H}$ denotes the non-linear observation operator, WCM or DNN as illustrated in Section 3.4.3 and 3.4.4.

The Kalman gain $\widetilde{\mathbf{K}}_k$ is computed at time k as follows:

$$\widetilde{\mathbf{K}}_k = \begin{pmatrix} \mathbf{K}_{k,[1]} \\ \mathbf{K}_{k,[2]} \\ \cdots \\ \mathbf{K}_{k,[12]} \end{pmatrix} = \widetilde{\mathbf{B}}\widetilde{\mathbf{J}}^T_k\left(\widetilde{\mathbf{J}}_k\widetilde{\mathbf{B}}\widetilde{\mathbf{J}}^T_k + \mathbf{R}\right)^{-1} \tag{6}$$

where $\mathbf{K}_{k,[p]}$ satisfies the update equation for $\mathbf{x}_{[p]}$:

$$\mathbf{x}^a_{k,[p]} = \mathbf{x}^f_{k,[p]} + \mathbf{K}_{k,[p]}\left(\mathbf{y}^o_k - \mathcal{H}\left(\mathbf{x}^f_k\right)\right) \tag{7}$$

and $\widetilde{\mathbf{J}}$ is the Jacobian of the model ($\mathbf{M}_{k,[p]}$) and the linearized observation operator ($\mathbf{H}_{k,[p]}$) at time step k for patch p, which follows:

$$\widetilde{\mathbf{J}} = \begin{pmatrix} \mathbf{J}_{k,[1]} \\ \mathbf{J}_{k,[2]} \\ \cdots \\ \mathbf{J}_{k,[12]} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{k,[1]}\mathbf{M}_{k,[1]} \\ \mathbf{H}_{k,[2]}\mathbf{M}_{k,[2]} \\ \cdots \\ \mathbf{H}_{k,[12]}\mathbf{M}_{k,[2]} \end{pmatrix} \tag{8}$$

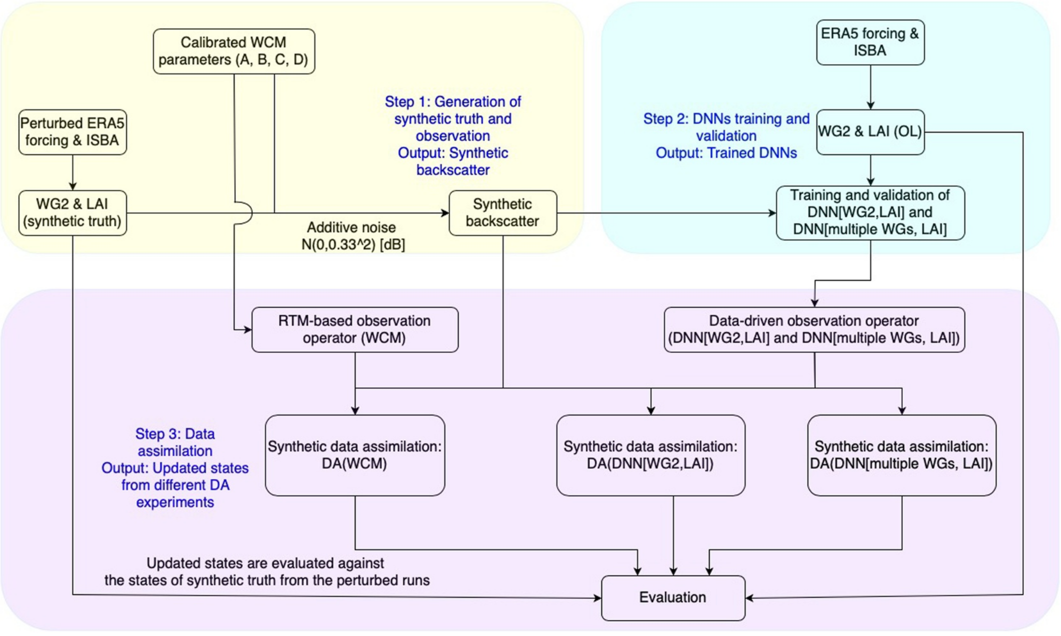

Figure 2. Experiments setup.

and $\widetilde{\mathbf{B}}$ is a block diagonal matrix which is "patch-dependently" defined as:

$$\widetilde{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_{k,[1]} & 0 & \dots & 0 \\ 0 & \mathbf{B}_{k,[2]} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \mathbf{B}_{k,[12]} \end{pmatrix} \tag{9}$$

The control vector evolution from time k to the end of the 24 hr assimilation window (k+1) follows:

$$\widetilde{\mathbf{x}}^{f}_{k+1,[p]} = \mathcal{M}_{k,[p]}[\widetilde{\mathbf{x}}^{a}_{k,[p]}] = \widetilde{\mathbf{M}}_{k,[p]}\widetilde{\mathbf{x}}^{a}_{k,[p]} \tag{10}$$

where $\mathcal{M}_{k,[p]}$ denotes the LSM ISBA-A-gs at time k at patch p. The term $\widetilde{\mathbf{M}}_{k,[p]}$ is the linearization of $\mathcal{M}_{k,[p]}$ at patch p.

3.4. Experiment Setup

The detailed experiment setup is illustrated in Figure 2. There are four steps in total: (a) calibration of the WCM, (b) generation of synthetic truth and observations, (c) DNN training and validation, and (d) data assimilation experiments. These are discussed in Sections 3.4.1 to 3.4.4 respectively. In step (2), a perturbed run of the model is used to generate the synthetic "true" surface soil moisture (WG2) and leaf area index (LAI). These synthetic true values of WG2 and LAI are used to generate synthetic backscatter using the calibrated WCM from step (a). Additive noise with zero mean and a standard deviation of 0.33 dB (Lievens, Martens, et al., 2017) is then added to these WCM outputs to account for observation error (Equation 12), producing the "true" radar backscatter observations for assimilation. The motivation to compare the WCM and the two DNNs using a synthetic experiment is:

- We can ensure that the dynamics of the radar backscattering coefficient are determined entirely by WG2 and LAI. There is no unexplained variance in the synthetic true observations. This allows us to evaluate the

Table 1
Perturbation Applied on the ERA5 Forcing Data, Following Reichle et al. (2002); De Lannoy and Reichle (2016)

Variables	Type	Std dev	Temporal correlation	Cross correlation with perturbations in white noise		
				Pcp	SW	LW
Pcp	M	0.5	24 h	1	−0.8	0.5
Downward SW	M	0.3	24 h	−0.8	1	0.5
Downward LW	A	20 W/m^2	24 h	0.5	−0.5	1

Jacobian of the respective observation operators, since the Jacobian of the "true" observation operator is perfectly known.

- Validating against "true" states and observations at the resolution of the model/assimilation grid eliminates any scale mismatch between the true and estimated states, which contributes to a representativeness error.

3.4.1. Calibration of WCM

The parameters A, B, C, and D of the WCM are determined by calibrating the ISBA-A-gs OL model simulations against the actual ASCAT normalized backscatter. The rationale is to provide a reasonable set of parameters to WCM to generate synthetic true observations described in Section 4.2. Therefore, the model simulations take WG2 and LAI from the ISBA-A-gs OL runs as the input. The calibration is performed from 2007 to 2016. The calibration process is done per GPI using the Shuffled Complex Evolution algorithm (SCE-UA) (Duan et al., 1992, 1994). The objective loss function J minimizes the root mean square error (RMSE) between WCM simulated backscatter σ^o_{WCM} and ASCAT backscatter $\sigma^o_{\mathrm{ASCAT}}$ at a given GPI. J is defined as:

$$J = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(\sigma^o_{\mathrm{WCM}} - \sigma^o_{\mathrm{ASCAT}}\right)^2} + W_\alpha \frac{1}{N_\alpha}\sum_{i}^{N_\alpha}\frac{(\alpha_{0,i} - \alpha_i)^2}{\sigma^2_{\alpha_{0,i}}} \tag{11}$$

where N is the number of WCM simulations and ASCAT observations available, N_α is the number of calibrated WCM parameters ($N_\alpha = 4$ in our case), W_α the weight-factor given to the parameter penalty term, α_i the ith parameter value, and $\alpha_{0,i}$ the prior value of the ith parameter. W_α is set as 0.1 to ensure that the minimization more strongly constrains the RMSE between WCM simulations and ASCAT observations than by the parameter penalty term. Following Lievens, Martens, et al. (2017), the parameter deviations from their initial values are scaled by the variance of a uniform distribution with boundaries $[\alpha_{\min}, \alpha_{\max}]$. The prior values and ranges of parameters are based on those used by Lievens, Martens, et al. (2017) and Shamambo et al. (2019). The Water Cloud Model is used here to generate synthetic backscattering coefficient data that are used as a training target for the DNN. The Water Cloud Model is implemented following the methodology outlined by previous studies (de Roos et al., 2023; Lievens, Martens, et al., 2017, Lievens, Reichle, et al., 2017; Modanesi et al., 2021; Shamambo, 2020; Shamambo et al., 2019).

3.4.2. Generation of Synthetic Truth and Observations

A perturbed run of the LSM ISBA-A-gs serves as the "synthetic true" solution and is meant to represent nature (Reichle et al., 2002). A "perturbed" run is generated starting from a prescribed initial condition on 1 January 1996. Following Reichle et al. (2002) and De Lannoy and Reichle (2016), perturbations are introduced to (a) ERA5 Forcing data to represent imperfect knowledge about the forcing data (Reichle et al., 2002); (b) the prognostic state variables like soil moisture in different layers, to mimic the errors in model structure and model parameters (De Lannoy & Reichle, 2016).

The precipitation (Pcp) and short wave radiation (SW) are perturbed with multiplicative (denoted as "M") lognormal perturbations. Long wave radiation receives additive normal perturbations, as listed in Table 1. Forcing perturbations have a temporal correlation length of 1 day. Additionally, moderate cross-correlations between the perturbed forcings are imposed to ensure physical consistency between the forcing errors. The perturbations to prognostic states like soil moisture in different layers are prescribed as additive Gaussian noise with zero mean and a standard deviation of $\lambda\left(w_{fc} - w_{wilt}\right)$ for surface soil moisture (SSM), with $\lambda = 0.5$ for soil moisture in layer 2 (1–4 cm depth), 0.2 for SM in layer 3 (4–10 cm depth), 0.05 for SM in layer 4 (10–20 cm depth) and 0.02 for SM in deeper layers. The variable λ is a dimensionless scaling coefficient for the calibration of the background errors (Fairbairn et al., 2015). These values are in line with Fairbairn et al. (2015) and Bonan et al. (2020).

Synthetic observations of backscatter (at 40° incidence angle) are generated by taking the WG2 and LAI from the synthetic truth and running the WCM with the calibrated parameters $(\mathrm{A_{cal}}, \mathrm{B_{cal}}, \mathrm{C_{cal}}, \mathrm{D_{cal}})$:

$$\sigma^{o}_{\mathrm{true}} = \mathrm{WCM}(\mathrm{A_{cal}}, \mathrm{B_{cal}}, \mathrm{C_{cal}}, \mathrm{D_{cal}}, \mathrm{WG2_{true}}, \mathrm{LAI_{true}}) + \epsilon_{\mathrm{true}} \tag{12}$$

Observation error is represented through ϵ_{true}, an additive Gaussian noise with zero mean and a standard deviation of 0.33 dB following Lievens, Martens, et al. (2017).

A second model integration, this time without perturbations (in forcing and states), represents the best model estimate of the "truth" without the benefit of data assimilation and is called the "open loop" (OL) simulation (following Reichle et al. (2008)). Next, synthetic observations generated from the "true" WG2 and LAI are subsequently assimilated back into the ISBA-A-gs OL using different observation operators. The synthetic true WG2 and LAI are unknown to the training of the DNN-basedobservation operator. Finally, the resulting assimilation estimates are compared against the synthetic truth data set. In the remainder of this section, we provide more detailed descriptions of the observation operators and synthetic data assimilation experiments.

3.4.3. DNN Training and Validation

Two DNNs are trained following the methodology described by Shan et al. (2022). The DNN consists of an input layer, multiple hidden layers, and an output layer. Each layer contains several neurons connected to neurons in adjacent layers through weight and bias matrices (Goodfellow et al., 2016). During the training process, the weights and biases are updated via backpropagation using the Adam optimizer (Kingma & Ba, 2014), a stochastic gradient-based algorithm that minimizes the loss function (or equivalently, maximizes accuracy) between the DNN outputs and the training targets. To prevent overfitting, early stopping is applied during the training process. The input data of DNN comes from the OL experiment of ISBA-A-gs from 2007 to 2019. Similar to the calibration of WCM, the DNN is trained independently per grid point. The training target of DNN comes from the synthetic observations of backscatter (at 40° incidence angle) generated from Equation 12. The structure of the DNN is not fixed. The DNN's structural hyperparameters—including the number of layers, the number of neurons per layer, the batch size, the activation function, and the learning rate—are tuned via Bayesian optimization (Snoek et al., 2012).

The DNNs are trained and tested using data from 2007 to 2016. To improve the robustness of the Jacobian, the data from 2007 to 2016 are randomly shuffled and sampled 50 times (Shan et al., 2024). This process creates 50 unique training-testing splits, ensuring a comprehensive examination across diverse data scenarios. Subsequently, each split is used to train an individual DNN, and the final estimates of the DNN are the ensemble mean of the outputs from each of these models trained on the $k - th$ data split ($k = 1, \ldots, 50$). The final Jacobian estimates are also obtained by computing the ensemble mean of the Jacobian estimates from each model.

The statistical inference of the DNN Jacobian is an ill-posed problem (Vapnik, 2013) because multiple solutions can be found for similar output statistics (Aires et al., 2004). The parameters of DNNs are very unstable and may vary drastically from one solution to another (Aires et al., 2004). The spread of Jacobians of multiple DNN solutions (trained on same data set) is referred to as the error of Jacobians (Krasnopolsky, 2007). Following Krasnopolsky (2007), an ensemble of DNNs was created for each GPI, each trained on different samples of the training data. The model architectures vary because the number of layers and nodes is optimized using Bayesian optimization, resulting in different weights for each DNN and different hyperparameters. Krasnopolsky (2007) showed that the ensemble mean of Jacobian estimates of all individual NN emulations can reduce the simulation errors and the magnitudes of the extreme outliers, as well as the errors of the Jacobian.

In this study, we consider three DNNs. First two are trained using the OL as input and the synthetic backscatter as the training target. For the first DNN (DNN[WG2, LAI]), the input land surface variables are surface soil moisture (WG2) and LAI. This first DNN therefore has the same inputs as the WCM. The second DNN (DNN[multiple WGs, LAI]) is trained with LAI and soil moisture in different layers from OL as the input. Soil moisture values in other layers are included because of the indirect link between root zone soil moisture and vegetation water content (Petchiappan et al., 2021; Shan et al., 2022) which affects radar backscatter (Khabbazan et al., 2022; Petchiappan et al., 2021). As a result, normalized backscatter could be indirectly influenced by root zone soil moisture.

The third DNN, DNN[$\mathrm{WG2_{true}}$, $\mathrm{LAI_{true}}$], is trained using synthetic truth values of WG2 and LAI as inputs. This will allow us to determine whether a DNN trained on the "true" input variables produces Jacobian estimates that are consistent with Jacobian of WCM. By comparing the Jacobians of DNN[$\mathrm{WG2_{true}}$, $\mathrm{LAI_{true}}$], WCM, and DNN [WG2, LAI], we test the hypothesis that discrepancies between DNN and WCM Jacobians stem from difference

Table 2
Details About Different Model Runs, Including OL, Perturbed Run, and Different DA Experiments

	OL	Perturbed run	DA(WCM)	DA(DNN[WG2, LAI])	DA(DNN[multiple WGs, LAI])
Forcing	ERA5	Perturbed ERA5	ERA5	ERA5	ERA5
Observation	–	–	Synthetic backscatter (σ^o_{true})	σ^o_{true}	σ^o_{true}
States	–	–	LAI, WG2, …,WG8	LAI, WG2, …,WG8	LAI, WG2, …,WG8
Period	1996–2019	2007–2019	2017–2019	2017–2019	2017–2019
Observation operator	–	–	WCM	DNN trained on WG2, LAI from OL	DNN trained on WG2,…,WG8, LAI

between OL and synthetic land surface variables (WG2 and LAI), instead of from the nonlinearities of the radiative transfer model used to generate the training targets (normalized backscatter).

The calculation of the Jacobians of the DNN models follows the methodology of Shan et al. (2022). Perturbation is added to the DNN input data. It is applied to one DNN input state at a time step in order to calculate the Jacobian. Then the estimates of the perturbed input are used to calculate the first order sensitivity of estimation to the input data. The magnitude of the perturbation is ±5% of the range of input data.

3.4.4. Data Assimilation Experiments

The synthetic backscatter observations generated in Section 4.2 are assimilated into the ISBA-A-gs using the SEKF to constrain the ISBA-A-gs model from 2017 to 2019. ISBA-A-gs is run on the same 0.25° × 0.25° grid as the ASCAT data. The synthetic backscatter is assimilated once every 24 hr at 10 a.m. (Shan et al., 2024) because ISBA-A-gs can only update LAI once a day. Backscatter at evening is not assimilated and model state is not updated at 10 p.m. either. Table 2 summarizes the details of different model runs, including OL, perturbed run, and different DA experiments. Three DA experiments were conducted using the same synthetic observations and the same parameter settings (background error and observation error) so that the only difference is the choice of the observation operator. DA(WCM) uses the WCM as an observation operator. DA(DNN[WG2, LAI]) is the DNN trained on surface soil moisture (WG2) and LAI, the same inputs required by the WCM. Finally, DA(DNN [multiple WGs, LAI]) is the DNN trained on the longer list of input variables (WG2, …, WG8, LAI). The DA (WCM) is the reference case because it uses the same observation operator that is used to simulate the synthetic observations.

In the SEKF, the background error matrix is defined as "patch-dependent" because the model uncertainty for LAI is calculated based on the value of LAI in the different patches. The values of $\mathbf{B}_{k,[p]}$, $p = 1,\ldots,12$ are defined following Mahfouf et al. (2009), Draper et al. (2011), and Albergel et al. (2017). The background errors of soil moisture are assumed to be proportional to the dynamic range (the difference between the volumetric field capacity w_{fc} and the wilting point w_{wilt}), which is determined by the soil texture (Noilhan & Mahfouf, 1996). In this study, the perturbation applied to the states of soil moisture in the calculation of the Jacobian matrix was assigned as $1 \times 10^{-4} \times \left(w_{fc} - w_{wilt}\right)$ following Albergel et al. (2017), and 0.001 × LAI for LAI. For volumetric surface soil moisture, a mean standard deviation (SD) error of 0.04 $m^3\ m^{-3}$ is prescribed. For soil moisture in deeper layers, SD error of 0.02 $m^3\ m^{-3}$ is used following Mahfouf et al. (2009), Draper et al. (2011), Barbu et al. (2011), and Barbu et al. (2014). The observation error for σ^o_{40} is set as 0.33 dB following Lievens, Martens, et al. (2017).

4. Results

4.1. WCM Calibration

The calibrated values of parameters A, B, C, and D are shown in Figure S1 in Supporting Information S1. High values of A and C are obtained in the urban areas of Toulouse and Bordeaux which is consistent with Shamambo et al. (2019). The B parameter exhibits higher values in needleleaf forest regions where vegetation attenuation effects are stronger than those in agricultural areas and grasslands (Lievens, Martens, et al., 2017). The distribution of parameter D is similar to those reported by (Modanesi et al., 2021), with a peak at 15 $dB/m^3/m^3$. Additionally, high values for both parameters C and D are found in the Landes forest. Shamambo et al. (2019) found the same and attributed this to the sandy soil (Danquechin Dorval et al., 2016), leading to dominant volume scattering, and the flat topography.

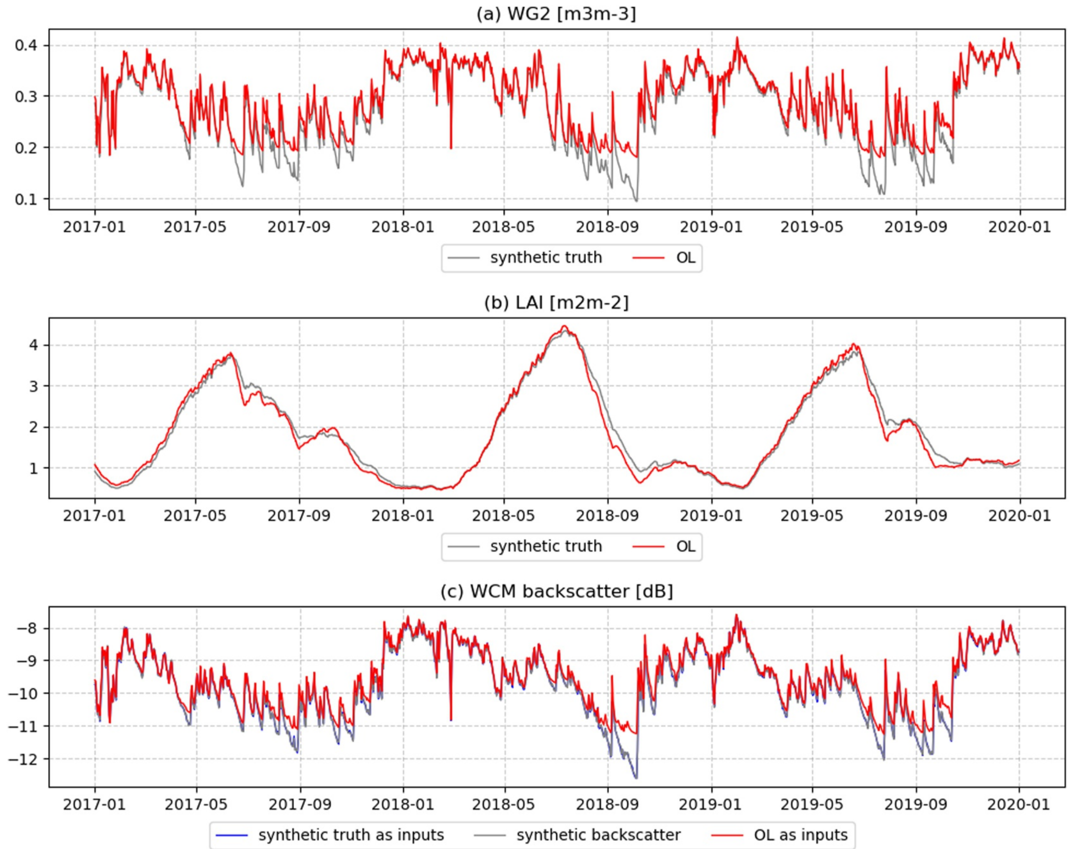

Figure 3. Time series of (a) surface soil moisture (WG2) and (b) LAI from the OL and perturbed run (synthetic truth), (c) observation equivalent (backscatter) estimated using WCM based on inputs from open loop (OL, the red line), synthetic truth of WG2 and LAI (blue line) and synthetic backscatter (σ^o_{true}, dark gray color) from 2017 to 2019. The time series are averaged over agricultural GPIs While the influence of observation noise is clear in the time series for a single GPI, it is not visible in this averaged time series.

4.2. Synthetic Truth and Observations

Figures 3a and 3b compare WG2 and LAI estimated using the OL to the synthetic true values from the perturbed run used to generate the synthetic observations. The synthetic true WG2 exhibits a consistent dry bias of approximately 0.015 $\mathrm{m}^3\ \mathrm{m}^{-3}$ with respect to the OL, even though the perturbations applied to forcing and prognostic states are bias-free. Fairbairn et al. (2015) and Bonan et al. (2020) also observed a significant negative bias in the perturbed ISBA-A-gs simulated WG2 compared to the OL when perturbations were applied to precipitation for the ensemble square root filter. They found that the magnitude of the perturbation bias in WG2 is inversely correlated with precipitation amount. Fairbairn et al. (2015) linked this negative bias to the nonlinear behavior of evapotranspiration and drainage. The negative bias occurs when the WG2 from the OL is near the wilting point during summer and autumn. Negative increments have no impact on the state. However, the extra water from the positive increments is removed rapidly by transpiration (Fairbairn et al., 2015). Fairbairn et al. (2015) also argue that this effect is also partly linked to the phenology. In particular, that under water-stressed conditions, the vegetation roots readily absorb excess water that becomes available, which increases the transpiration and LAI. For LAI, the difference between the OL and synthetic truth are generally very small, with largest biases (0.14 $\mathrm{m}^2\ \mathrm{m}^{-2}$) from August to November. Therefore, there is limited potential for ASCAT assimilation to improve the estimate of LAI.

Figure 3c compares the (synthetic) true normalized backscatter (σ^o_{true}) to that simulated using WCM given the OL or synthetic true inputs (from Figures 3a and 3b). The influence of the dry bias in WG2 between the OL and synthetic truth is clear in Figure 3c. Recall that σ^o_{true} is generated by WCM simulations using perturbed WG2 and

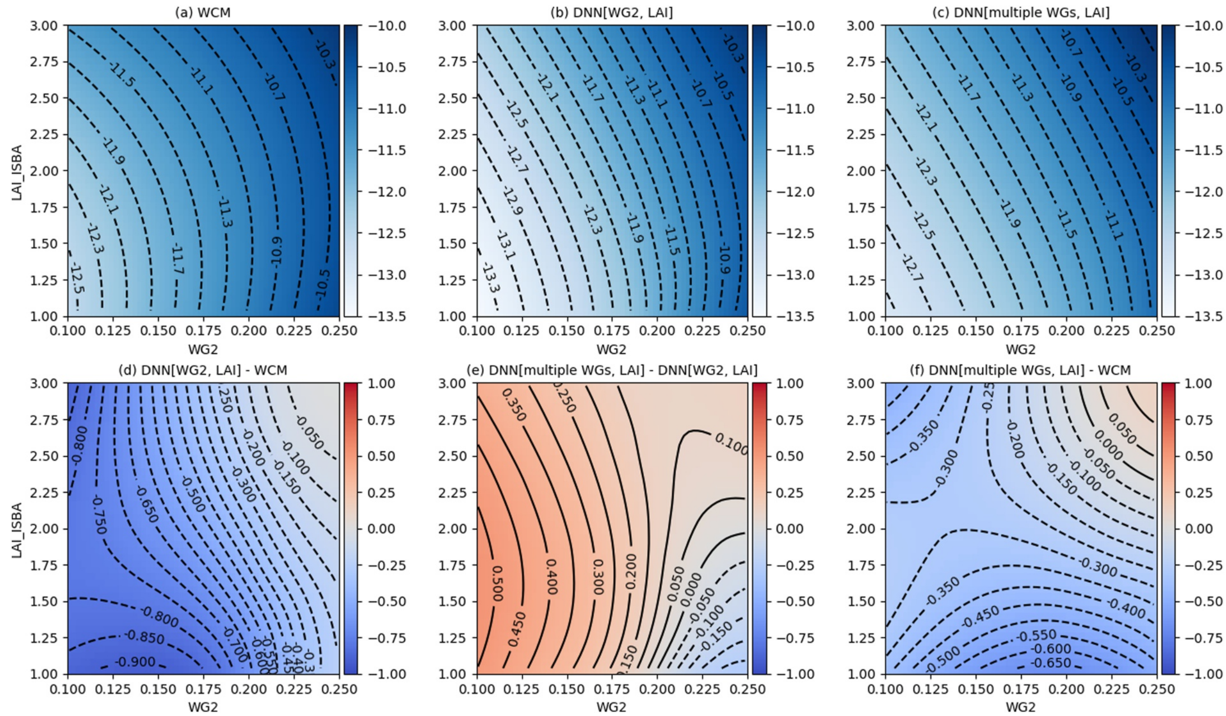

Figure 4. Contour plot of estimated backscattering coefficient (dB) based on different observation operators. The estimate is averaged over observation operators trained on all agricultural GPIs. The x and y axes represent the input variables, WG2 and LAI. Results are shown for the three observation operators, in subplots (a) WCM, (b) DNN[WG2, LAI], and (c) DNN[multiple WGs, LAI]. Subplot (d), (e) and (f) show the difference between estimated backscattering coefficient using different observation operators. The ranges of WG2, LAI, and other states correspond to the values simulated in July, August, September, and October.

LAI and then adding noise. The synthetic true backscatter is therefore lower than that simulated using the OL, and the only difference between the blue and gray lines is the addition of the assumed observation error.

4.3. Comparison of Different Observation Operators

Figure 4 shows the simulated σ^o based on different observation operators as a function of WG2 and LAI. Only observation operators over agricultural GPIs are considered, as the differences in estimation performance among the methods are most pronounced for agricultural areas compared to other land cover types (Figure S3 in Supporting Information S1). The ranges for WG2 (0.10–0.25 $m^3\ m^{-3}$) and LAI (1–3 $m^2\ m^{-2}$) are chosen to approximate the range of these states from the OL during July, August, September, and October. This period is particularly relevant because the largest differences between OL and synthetic truth were observed in this period as shown in Figure 3. The similarity between Figures 4b and 4d, or equivalently (c) and (f), demonstrates that there is some difference in the simulated σ^o from the two DNNs. This difference is quantified in Figure 4e. The maximum difference between two DNNs is about 0.35 dB when WG2 reaches 0.15 $m^3\ m^{-3}$) while the max difference between DNN and WCM is about −0.7 dB) as shown in Figure 4c. The gradients of contour lines in Figures 4a, 4b, and 4d correspond to the sensitivity of each of the observation operators to WG2 and LAI, and therefore provide insight into the similarity of the Jacobians of the three models. It is clear from Figures 4b and 4a that DNN[WG2, LAI] overestimates the sensitivity of σ^o to WG2 when LAI is low (around 1.5 $m^2\ m^{-2}$). The contour line is denser and the directions are different in Figures 4b and 4d than they are in Figure 4a. Further, the gradients of contour lines in Figure 4c show the difference in sensitivities in Figures 4a and 4b. It is clear that the sensitivity of DNN[WG2, LAI] to WG2 is overestimated while the sensitivity to LAI is underestimated. Therefore, the simulated backscatter is more sensitive to the dynamics of soil moisture than in the WCM.

Figure 5 maps the agreement between the synthetic true backscatter and the simulated backscatter from each of the observation operators given the open loop LSM states as inputs. Recall that the calibrated WCM parameters are used to generate synthetic σ^o_{true} which serves as the training target for the two DNN models. Therefore, WCM's

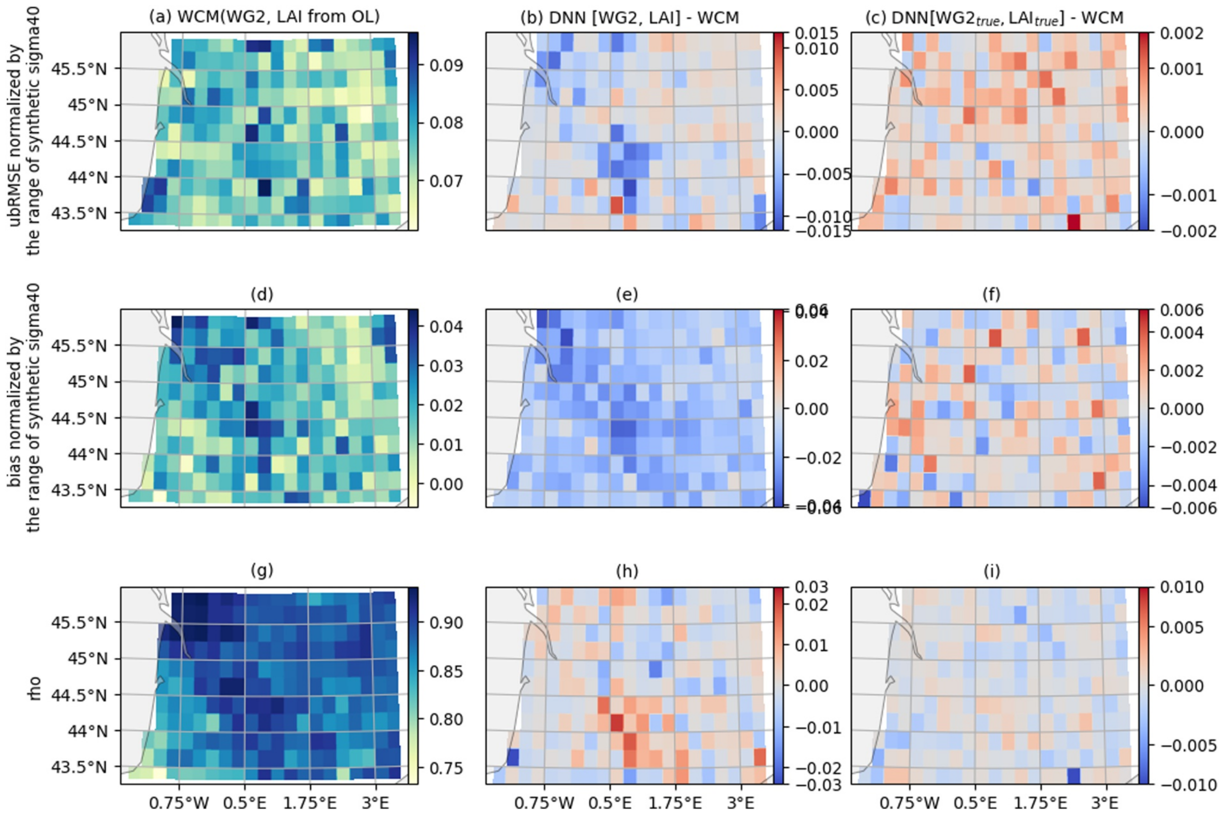

Figure 5. Performance metrics of observation equivalents (backscatter) estimated using different observation operators based on WG2 and LAI from open loop (OL) from 2017 to 2019, evaluated against synthetic backscatter. Specifically, the first column calculates the metrics between σ^o_{true} (from Equation 12) and WCM($A_{cal}, B_{cal}, C_{cal}, D_{cal}, WG2_{OL}, LAI_{OL}$). Top, middle, and bottom rows show unbiased RMSE (ubRMSE, in dB), bias (dB), and correlation coefficient. The ubRMSE (dB) and bias (dB) at individual GPIs are normalized by the range of synthetic backscatter (σ^o_{true}) at that GPI. The first column contains the metrics for the WCM, while the second and third columns show the differences between performances of different DNNs and the WCM.

performance represents the best estimate of σ^o_{true} with the given WG2 and LAI inputs from the OL. The domain average value of the ubRMSE of WCM is 7.47% of the range of σ^o_{true}. In the forested area of Les Landes, ρ is low, which may be attributed to the effects of cyclone Klaus, as discussed by Corchia et al. (2023) and Shan et al. (2024). Cyclone Klaus struck Les Landes in January 2009, resulting in forest degradation from 2009 to 2012 and regeneration from 2013 to at least 2017 (Shamambo et al., 2019; Teuling et al., 2017). Damage due to the storm itself and the subsequent recovery of the forest means that interannual variability is high, rendering the observations difficult to predict (Shan et al., 2022). Both DNN[WG2, LAI] and DNN[multiple WGs, LAI] have a lower normalized ubRMSE compared to WCM in agricultural GPIs.

Figure 6 shows the seasonal variation in the performance of different observation operators taking the OL states as inputs to estimate the synthetic observations, averaged across agricultural GPIs as shown in Figure 1. All observation operators exhibit the highest ubRMSE in autumn. Similar to Figure 5, the two DNNs have smaller ubRMSE than the WCM. The DNN[multiple WGs, LAI] achieves the lowest ubRMSE. The WCM bias peaks in August and September at 0.4 dB due to the difference between the OL and synthetic true WG2 and LAI. However, the bias of two DNNs is close to zero. Figure 7 shows that the normalized backscatter simulated with the DNNs with OL as inputs is closer to σ^o_{true} (0.057 dB averaged in August, September, and October) compared to WCM estimation in autumn (0.37 dB), using OL as inputs. However, when the synthetic truth is used as input to the DNNs, the simulated backscatter is also much lower than σ^o_{true} in August and September. This is because synthetic WG2 is generally drier than the WG2 from OL during this period. The difference in monthly performances between the DNNs and the WCM suggests that the DNNs are overfitting. The smaller bias in autumn occurs because the DNN is trained taking the OL as inputs and the synthetic observations as target variables.

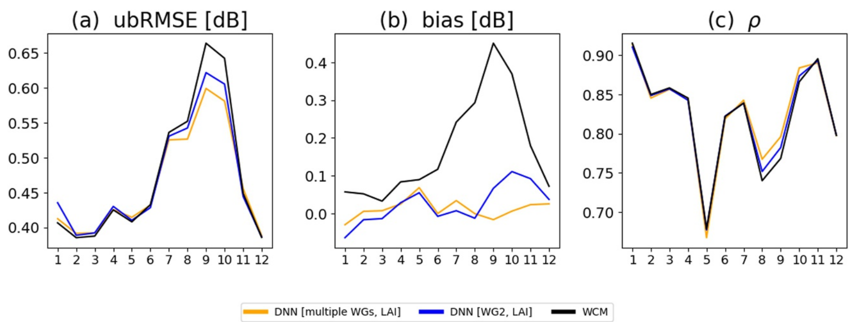

Figure 6. Monthly performance of observation equivalent (backscatter) estimated using different observation operators based on inputs from open loop (OL) from 2017 to 2019, evaluated against synthetic backscatter. The performances are averaged among agricultural GPIs as shown in Figure 1.

4.4. Jacobians of Different Observation Operators

Figure 8 shows the time series of the Jacobian terms of the different observation operators to WG2 and LAI calculated using the method described by Shan et al. (2022). The $J(\sigma^o_{\mathrm{true}}, \mathrm{WG2})$ of WCM represents the "true" sensitivity, that is how simulated backscatter is sensitive to WG2 and LAI from the OL. The two DNNs generally reproduce the seasonal cycle of the WCM Jacobian terms. However, there are large biases in $J(\sigma^o_{\mathrm{true}}, \mathrm{WG2})$ of DNN[WG2, LAI] in autumn. DNN[WG2, LAI] overestimates the sensitivity to WG2 compared to WCM which is in consistent with Figure 4c.

WCM is used to generate the synthetic backscatter, providing a mathematical mapping from the synthetic truth of WG2 and LAI to estimate synthetic backscatter. Therefore, the synthetic backscatter and synthetic truth share similar distributions. However, DNN[WG2, LAI] is trained on WG2 and LAI from the OL to estimate synthetic backscatter. There are differences between the distributions of the OL and synthetic truth (or synthetic backscatter) due to the perturbations. Consequently, any difference between the mathematical mappings of DNN [WG2, LAI] and WCM is due to the difference between WG2 from the OL and synthetic backscatter (or synthetic truth of WG2). WG2 from the OL is larger than the synthetic truth in autumn, so DNN[WG2, LAI] overestimates the sensitivity of σ^o_{true} to WG2 as shown in Figure 9. The magnitude of the difference between WG2 of the synthetic truth and OL is closely related to the difference in the Jacobian terms from WCM and DNN[WG2, LAI].

The inclusion of soil moisture in other layers (WG3,…,WG8) in the list of inputs leads to more overfitting. Figure 8a shows that $J(\sigma^o_{\mathrm{true}}, \mathrm{WG2})$ of DNN[multiple WGs, LAI] is smaller than the corresponding term of the Jacobian of WCM or DNN[WG2, LAI], which uses fewer inputs. The sum of $J(\sigma^o_{\mathrm{true}}, \mathrm{WG2})$, …$J(\sigma^o_{\mathrm{true}}, \mathrm{WG8})$ (the

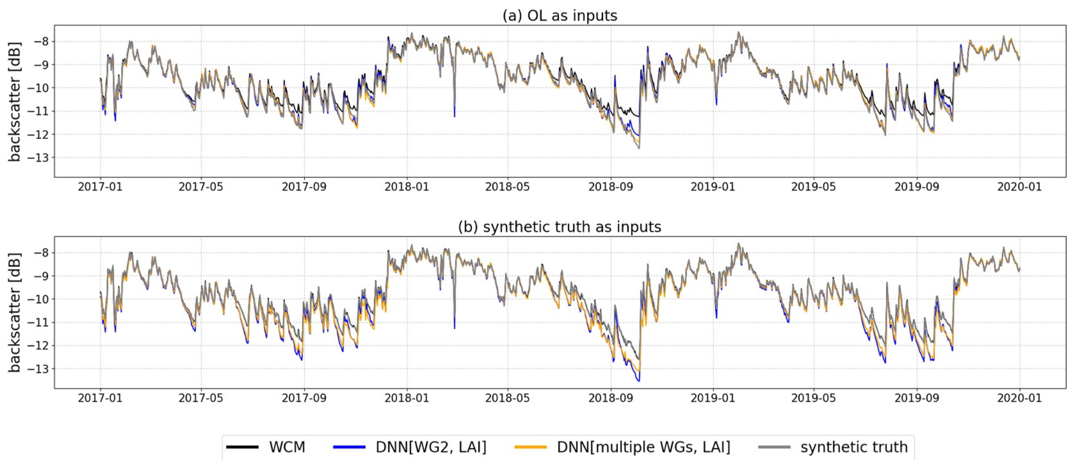

Figure 7. Time series of observation equivalent (backscatter) estimated using different observation operators based on inputs from (a) open loop (OL) (b) synthetic truth of WG2 and LAI (blue line) and synthetic backscatter (σ^o_{true}, dark gray color) from 2017 to 2019. Time series are averaged over agricultural GPIs as shown in Figure 1.

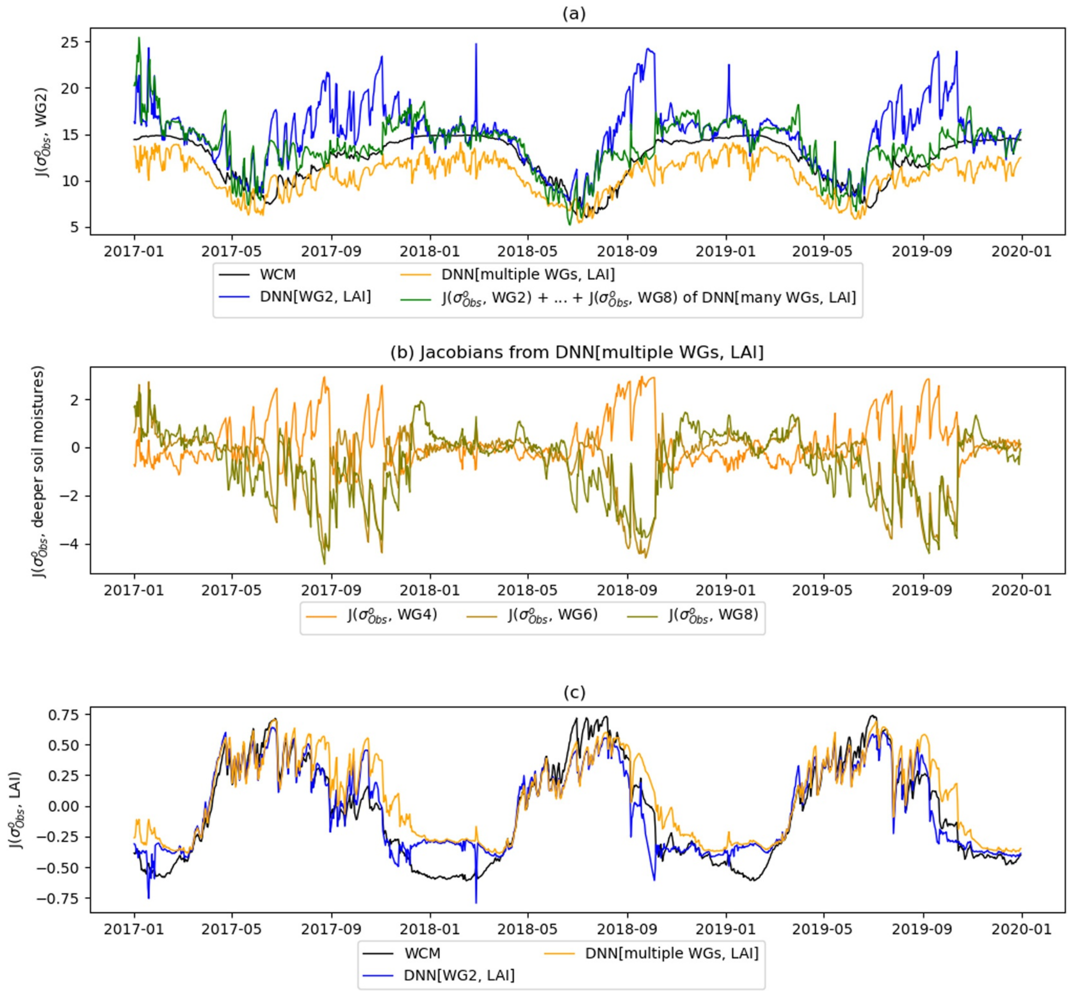

Figure 8. Time series of Jacobian terms related to WG2 and LAI from different observation operators, including WCM (black), DNN[WG2, LAI] (blue), and DNN[multiWGs, LAI] (orange). Units are dB m^{-3} m^3 for $J(\sigma^o_{\text{obs}}, \text{WG2})$ and dB m^{-2} m^2 for $J(\sigma^o_{\text{obs}}, \text{LAI})$. Subplot (a) shows $J(\sigma^o_{\text{obs}}, \text{WG2})$ from each observation operator, along with the sum of Jacobians with respect to multiple soil moisture layers, that is, $J(\sigma^o_{\text{obs}}, \text{WG2}) + \cdots + J(\sigma^o_{\text{obs}}, \text{WG8})$, from DNN[multiWGs, LAI] (green). Subplot (b) shows the individual Jacobians of σ^o_{obs} with respect to deeper soil moisture layers (WG4, WG6, WG8) from DNN[multiWGs, LAI]. Subplot (c) shows $J(\sigma^o_{\text{obs}}, \text{LAI})$ from WCM, DNN[WG2, LAI], and DNN[multiWGs, LAI]. All Jacobians are computed using open-loop (OL) values of WG2 and LAI as input, and the time series are averaged over agricultural GPIs.

green line in Figure 8a) is much closer to the corresponding term of the Jacobian of WCM. This indicates that the DNN[multiple WGs, LAI] arbitrarily assigns the sensitivity to soil moisture in different layers. For example, the sensitivity to WG4 has an opposite sign compared to the sensitivity to WG6 and WG8, which is not physically plausible.

Regarding the sensitivity to vegetation dynamics, $J(\sigma^o_{\text{true}}, \text{LAI})$ of both DNNs are much closer to zero in winter. This is because of the low values of LAI during this season. The low LAI values lead the DNN models to lose sensitivity to LAI. Additionally, $J(\sigma^o_{\text{true}}, \text{LAI})$ of DNN[multiple WGs, LAI] is larger in autumn compared to DNN [WG2, LAI]. This suggests that the inclusion of redundant soil moisture states in the inputs to the DNN has a detrimental effect of the estimation of the sensitivity to vegetation dynamics.

Table 3 provides a summary of the differences between WCM and two different DNNs in terms of performance skills and Jacobians. Unlike the equations in WCM, the DNNs are trained using the OL data. This results in smaller ubRMSE and/or higher ρ when estimating σ^o_{true} with inputs from the OL. There are notable differences in the Jacobians. The comparison between DNN[WG2, LAI] and WCM shows that, even with the correct choice of

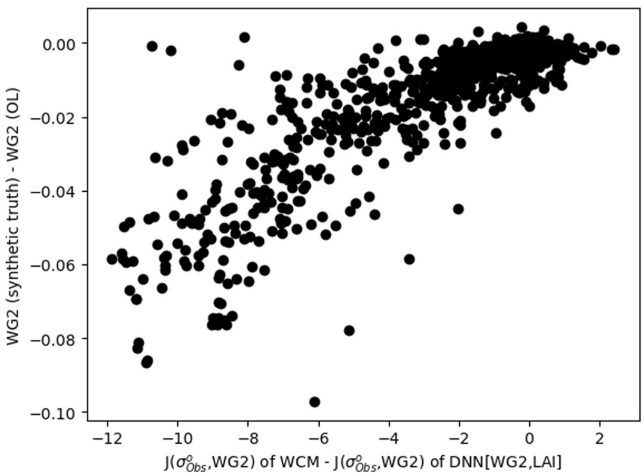

Figure 9. Scatter plot about the differences between the WG2 from synthetic truth and OL (y-axis) versus the differences between the $J(\sigma^o_{\text{true}}, \text{WG2})$ of WCM and DNN[WG2, LAI]. The Jacobians takes the WG2 and LAI from the OL as inputs. The data are averaged over agricultural GPIs.

input, DNN[WG2, LAI] tends to overfit, and achieve better performance skills than the WCM. Furthermore, the comparison between DNN[WG2, LAI] and DNN[multiple WGs, LAI] illustrates that even when the state estimates are quite similar, the Jacobian terms can vary significantly. This variation is due to the presence of redundant information in the input data. In the case of DNN[multiple WGs, LAI], the inclusion of several additional soil moisture states (WG3 to WG8) as input allows DNN[multiple WGs, LAI] overfitting.

4.5. DNN Trained on the Synthetic Data

DNN[WG2_{true}, LAI_{true}] is compared to DNN[WG2, LAI] and WCM to examine how different input sources affect the learning. Figure S4 in Supporting Information S1 shows the σ^o simulated by DNN[WG2_{true}, LAI_{true}] closely matches the WCM, with a difference smaller than 0.06 dB. The contour line directions are visually similar. These indicate that DNN[WG2_{true}, LAI_{true}] can successfully emulate WCM, despite the non-linearities present in σ^o_{true}.

Figure S5 in Supporting Information S1 shows the comparison of Jacobian terms of DNN[WG2_{true}, LAI_{true}], DNN[WG2, LAI], and WCM. The difference between $J(\sigma^o_{\text{true}}, \text{WG2})$ of DNN[WG2_{true}, LAI_{true}] and WCM is very small. This confirms that the DNN can accuratelly capture the sensitivities of normalized backscatter to WG2 if trained on synthetic true data without noise. Given the same training target (σ^o_{true}), it is clear that the difference between $J(\sigma^o_{\text{true}}, \text{WG2})$ of DNN[WG2, LAI] (blue line) and WCM (black line) is due to the difference between OL and synthetic true data. These differences arise because DNN[WG2, LAI] is optimized to minimize the error between σ^o_{true} and biased OL inputs, without accounting for uncertainty in the input data. This reflects a form of overfitting, that is following the errors or noise in training data too closely (James et al., 2013) or memorizing the pecularities in input data, compromising its ability to find a general predictive rule (Dietterich, 1995). This issue parallels classical overfitting in machine learning, where a model learns the training data (in-sample) too well, including irrelevant details and noise, leading to poor performance on new, unseen (out-sample) data (Ying, 2019). The expansion from in-sample to out-sample data relies on a correct representation of sensitivities of training target to input variables. The current way to design the machine learning training workflow seeks the "optimal" loss during the training process. However, this design becomes less "optimal" regarding the representation of Jacobians.

Similar issues also exist in other data-driven algorithms like linear regression, but with random noise. When predictor variables are measured with random error, the conventional least squares regression analysis (or logistic regression (Rosner et al., 1989)) would create biased regression coefficient (MacMahon et al., 1990) (referred to as "regression dilution" in the univariate case (Frost & Thompson, 2000)). Therefore, machine learning studies in land data assimilation should use data that are more representative of the "truth", or include error quantification and physical constraints related to sensitivities into the training process to avoid potential overfitting or any influence of regression dilution on the Jacobians.

4.6. Data Assimilation Results

Figure 10 shows the agreement between states including LAI, WG2, WG4 and WG6 estimated in the different DA experiments and the synthetic truth. Detailed statistics are reported in Table S3 in Supporting Information S1. The domain median value of ubRMSE of WG2 from the OL is 0.0265 $m^3\ m^{-3}$. WG2 from the OL has a wet bias of

Table 3
Summary of the Performance of the DNNs

	DNN[WG2, LAI]	DNN[multiple WGs, LAI]
Performances based on OL	Better than WCM	Similar as DNN[WG2, LAI]
$J(\sigma^o_{\text{true}}, \text{WG2})$ based on OL	Overestimated in autumn	Underestimated
$J(\sigma^o_{\text{true}}, \text{LAI})$ based on OL	Lose sensitivity in winter	Lose sensitivity in winter

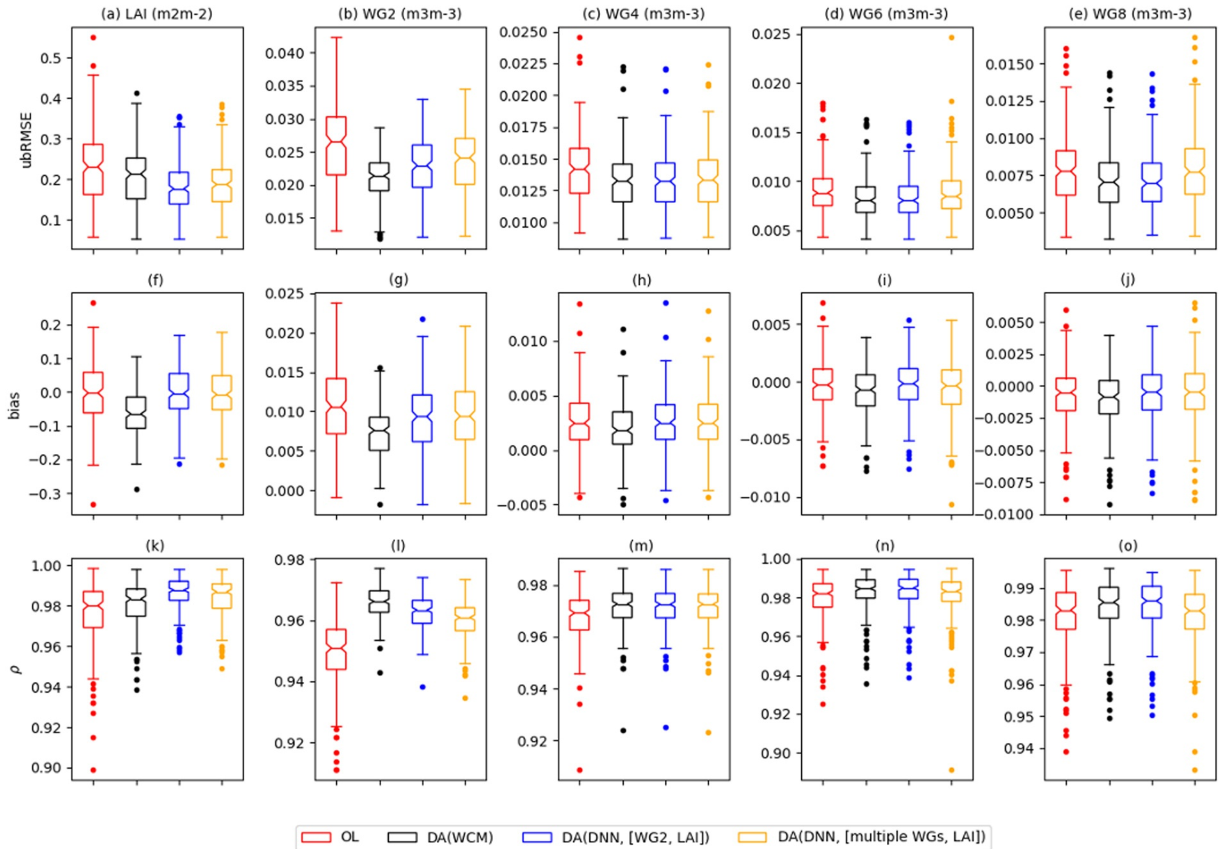

Figure 10. Performances of land surface states (LAI, WG2, WG4, WG6, WG8) from different experiments like OL, DA(WCM), DA(DNN[WG2, LAI]) and DA(DNN [multiple WGs, LAI]) DA(WCM). Different rows show different performance metrics (unbiased RMSE from (a–e), bias from (f–j), and correlation coefficient from (k–o). Different columns show performances of different variables. Different colors show different experiments.

0.0106 $m^3 m^{-3}$ (domain median value). The magnitude of ubRMSE (ρ) of WG2 from the OL is generally lower (higher) than what is typically observed in real data assimilation experiments (Shan et al., 2024). This is due to the fact that the perturbations applied to forcing and state variables are small and fail to represent the complex nature of precipitation errors, including the probability of missed precipitation or false alarms (Fairbairn et al., 2015). However, the magnitude reported here is in agreement with the findings presented in the previous synthetic data assimilation experiment (Fairbairn et al., 2015). Future studies should explore more sophisticated error models to better capture the errors in forcings and model.

The DA(WCM) achieves the largest improvements, reducing the domain median ubRMSE of WG2 to 0.0213 $m^3 m^{-3}$ and improving the ρ from 0.951 to 0.966. The DA(DNN[WG2, LAI]) reduces the ubRMSE of WG2 to 0.0229 $m^3 m^{-3}$. The update of WG2 from the DA(DNN[multiple WGs, LAI]) is the smallest among the three DA experiments. For soil moisture in deeper layers (WG4, WG6 and WG8), all 3 experiments reduce the domain median values of ubRMSE and improve the ρ. However, the performance of DA(DNN[multiple WGs, LAI]) is closest to the OL. Regarding LAI, the DA(DNN[WG2, LAI]) shows the most substantial reduction in ubRMSE, even better than the DA(WCM). This is because the DA(WCM) updates LAI to a smaller value in July, August and September 2018, as shown in Figure 11b. Recall that the DNN[WG2, LAI] and DNN[multiple WGs, LAI] have smaller updates due to the lower sensitivities to LAI compared to WCM during the same period as shown in Figure 8.

In terms of monthly performance, Figure 11d shows that the ubRMSE of WG2 from OL peaks in autumn. Meanwhile, the update due to the assimilation of σ^o_{true} is largest in the same period. The monthly ubRMSE of WG2 from DA(DNN[WG2, LAI]) is smaller than the DA(DNN[multiple WGs, LAI]). Additionally, Figures 11j and 11m shows that the DA(DNN[multiple WGs, LAI]) increases the ubRMSE of WG6 and WG8. This is because DNN[multiple WGs, LAI] is overfitting and arbitrarily assigns the sensitivity of the estimated σ^o_{true} to soil

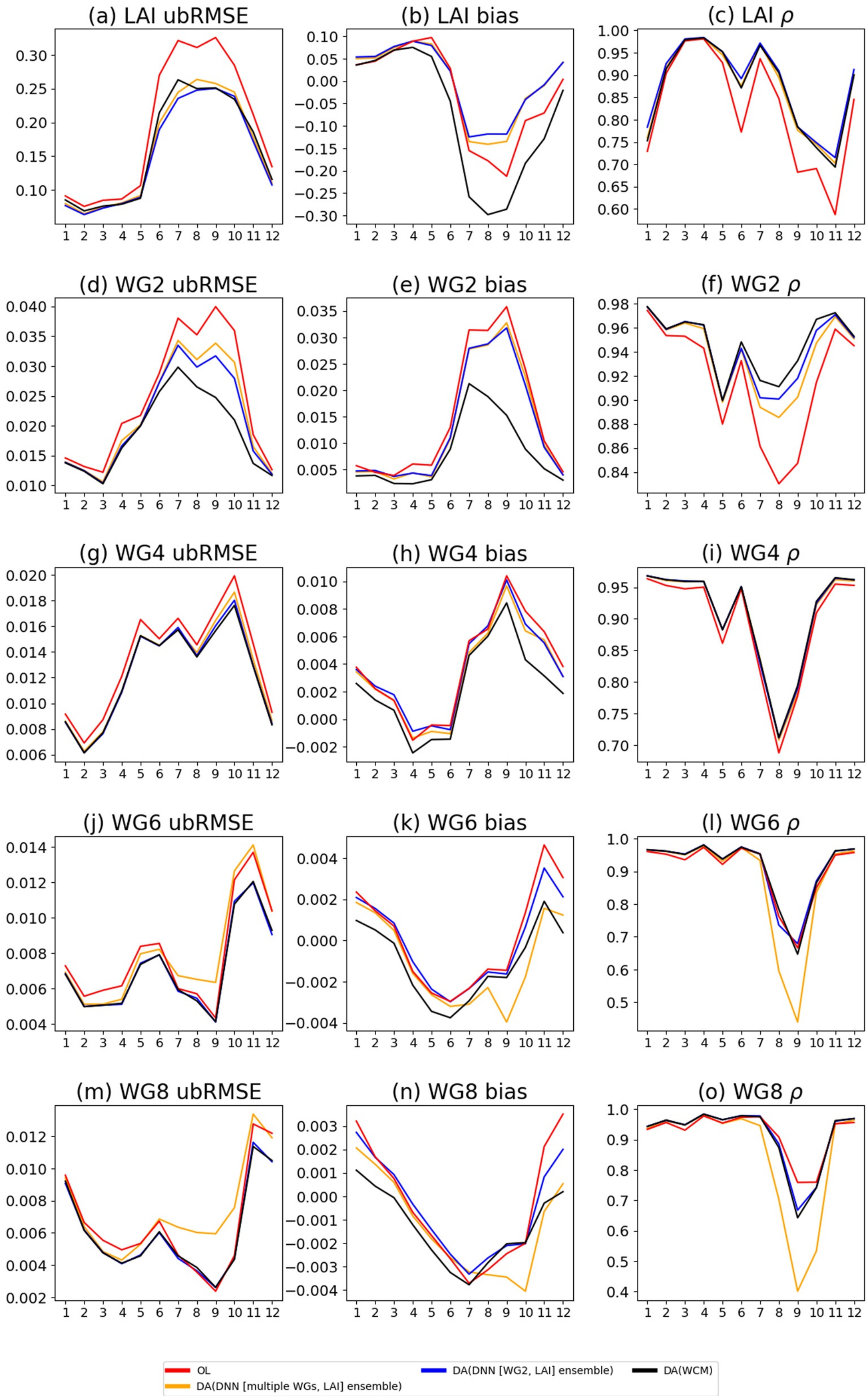

Figure 11. Monthly performances of land surface states (LAI, WG2, WG4, WG6, WG8) from different experiments like OL, DA(WCM), DA(DNN[WG2, LAI]) and DA(DNN[multiple WGs, LAI]) DA(WCM). Data are averaged among agricultural GPIs.

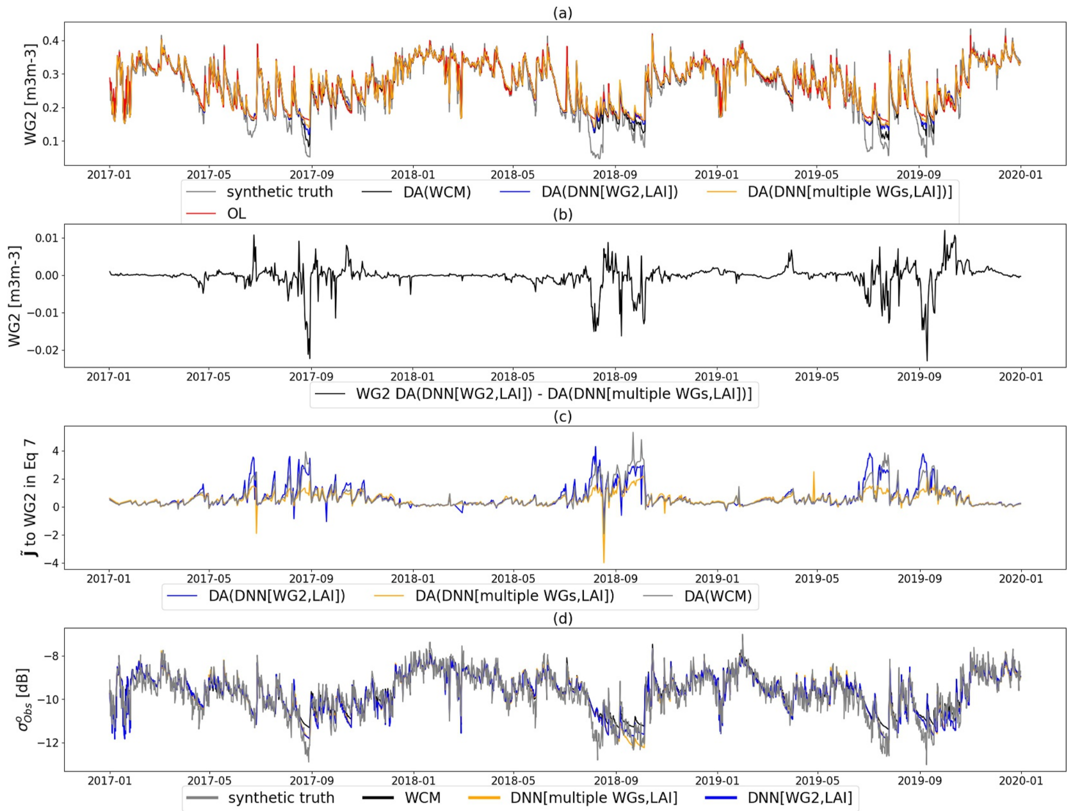

Figure 12. Time series on one GPI (lat 43.875, lon 2.125) of (a) WG2 from OL, synthetic truth, DA(DNN[WG2, LAI]), and DA(DNN[multiple WGs, LAI]); (b) the difference of WG2 between DA(DNN[WG2, LAI]) and DA(DNN[multiple WGs, LAI]); (c) the $\widetilde{\mathbf{J}}$ to WG2 in Equation 7 aggregated on patches; (d) the observation equivalent of different observation operators based on OL.

moisture across various layers—from WG3 to WG8. This allocation leads to large values of components of $\widetilde{\mathbf{J}}$ matrix in Equation 7 for WG6 and WG8, which results in the false updates in WG6 and WG8.

Figure 12 shows the time series of WG2 from DA(DNN[WG2, LAI]) and DA(DNN[multiple WGs, LAI]) for one agricultural GPI (lat 43.875, lon 2.125). The GPI is covered by agriculture crops (40%) and grasslands (27%). The largest difference in updates occurs in August, September, and October as shown in Figure 12b. This is the time when WG2 from the OL deviates most from the synthetic truth. DA(DNN[WG2, LAI]) updates WG2 to a much drier estimate while the update from DA(DNN[multiple WGs, LAI]) is smaller. Figure 12d shows that the DNN [WG2, LAI] and DNN[multiple WGs, LAI] have similar prediction skills of σ^o_{true} in the same period using the same input (WG2 and LAI from OL). They both fit the synthetic truth of σ^o_{true} well, while the estimation from WCM based on OL input exhibits a positive bias.

The similar prediction skills suggest the difference in the DA update is due to the difference in $\widetilde{\mathbf{J}}$ to WG2 in Equation 7 (as illustrated in Figure 12c). The $\widetilde{\mathbf{J}}$ matrix quantifies the sensitivity of observation equivalent to the state in the forecast. Larger absolute values of $\widetilde{\mathbf{J}}$ indicate a larger sensitivity. The DA(DNN[multiple WGs, LAI]) has a smaller value of $\widetilde{\mathbf{J}}$ because the DNN[multiple WGs, LAI] is overfitting and allocates the sensitivity to soil moisture in multiple layers (suggested in Figure 8). This leads to smaller updates of WG2 from DA(DNN [multiple WGs, LAI]).

5. Discussion

The motivation to use the states from the OL as training input is to minimize the potential bias in the climatology between the model and the predicted observations (Forman & Xue, 2016; Xue et al., 2018). The comparison between DNN[WG2, LAI] and WCM shows that even with the correct choice of inputs, a DNN model trained on

OL data may overfit. DNN[WG2, LAI] trained on OL data achieves lower ubRMSE and higher ρ evaluated against synthetic σ^o_{true}. A consequence of this overfitting is the poor estimation of the Jacobian. So, while the DNN model may be very successful in estimating the target variable, the sensitivity to geophysical variables may not be consistent with the true sensitivity, or the sensitivity one would obtain using a physically-based model.

The results suggest that the DNN limitation identified here should not be interpreted as a failure of neural networks per se. Rather, it arises from a mismatch between the state distribution used for training and the synthetic truth distribution used for assimilation. This issue is relevant beyond the present synthetic experiment, because many machine-learning-based DA and Earth system modeling applications are trained using model simulations or reanalysis products. If these training data contain biases or errors, the underlying state-observation sensitivities may differ from those of the true system. This may remain hidden when evaluation focuses primarily on predictive performance. Therefore, machine-learning-based observation operators should be evaluated not only by their prediction skill, but also, where possible, by whether their learned sensitivities are physically meaningful.

The comparison between DNN[WG2, LAI] and WCM highlights the need to train ML observation operators with inputs with smaller bias and errors. One unavoidable challenge is that the real truth is unknown. One solution would be to train the ML-based observation operator on satellite data, similar to Corchia et al. (2023)'s use of CGLS LAI rather than the model LAI. This approach is suitable if satellite data are available for each of the required DNN inputs, and if these satellite data are an adequate proxy for the true state, in terms of both climatology and anomalies. If the satellite data have low precision (i.e., low correlation with the truth), one possible solution may be to use reanalysis or assimilated products that incorporate external observations (e.g., site-level measurements) to reduce errors.

Beyond using training data with smaller bias and errors, two potential solutions could mitigate this Jacobian issue under biased or uncertain state representations: (a) a learning setup that explicitly accounts for data uncertainty and distributional mismatch, to reduce overfitting to noisy samples; (b) a different network architecture which incorporates physics-based constraints to regularize the learned sensitivites.

For solution (a), one choice might be uncertainty-aware training that better represents state-distribution mismatch. The DNNs can be trained on an ensemble of LSM runs to learn the proxy of uncertainty or probabilistic distribution of the states. Therefore, the pre-trained machine-learning-based observation operator would not rely on single open-loop trajectory of model simulations. The ensemble simulations may include different stochastic perturbations on forcing and states (representing random errors between LSMs and “truth”), as well as different model subschemes for single key processes (representing systematic errors between LSMs and “truth”). However, the ensemble of LSM runs require significant computational costs. Therefore, distribution-driven data assimilation is proposed to use machine learning algorithms to emulate and constrain the state and parameter distributions (S. Li et al., 2025; L. Li et al., 2024; Bhouri & Gentine, 2023), without integrating numerous model runs, thereby reducing the computational cost. Future studies could explore how machine-learning-based observation operator behave in the distribution-driven data assimilation framework.

For solution (b), one choice might be to embed microwave scattering physics directly into the training of ML-based observation operators and then systematically evaluate how this affects DNN-derived Jacobians and DA performances. Future work can introduce physics-informed constraints based on radiative transfer and vegetation–soil scattering mechanisms (Nikaein & Lopez-Dekker, 2025), so that the learned operator remains consistent with known sensitivities to key state variables (e.g., soil moisture, vegetation water content, and biomass).

Additionally, the comparison between DNN[WG2, LAI] and DNN[multiple WGs, LAI] highlights that it is crucial to look beyond the agreement with the true states when evaluating ML models; the Jacobian is also important. As discussed in Section 4.4, although the inclusion of more input data may improve estimation performances evaluated during an independent validation period), it can also lead to incorrect sensitivities, which are critical in data assimilation and earth system modeling applications. In ML, training optimizes parameter sets within given input and output spaces, searching for one of the local optima or the global optima. However, this study shows that achieving high model performance does not necessarily ensure correct underlying inferences, such as Jacobians or sensitivities consistent with physical understanding. These challenges may limit the applicability of machine learning algorithms in data assimilation.

Regarding the following DA experiments, the current design follows the best practice on perturbation used in prior synthetic and real-data assimilation studies of land surface models, such as the Catchment model (Reichle et al., 2002) and ISBA (Bonan et al., 2020; Fairbairn et al., 2015). However, the error introduced in forcing and state variables remains somewhat idealized and may not fully capture real-world variability. One outcome is the small domain-averaged difference between synthetic truth and OL data. While the magnitude of the difference in WG2 from synthetic truth and OL is consistent with previous findings (Fairbairn et al., 2015), more comprehensive error models can be taken into account. For example, a larger precipitation error models could be used as recommended by Fairbairn et al. (2015). Additionally, cross-correlated perturbations could be added into the soil moisture in each layer to assess the impact of such correlations on DNN training and subsequent DA results. However, this would require prior knowledge of the coupling strength between surface and root-zone soil moisture (Sawada et al., 2015), which is known to vary with both climate and soil type (Kumar et al., 2009). The coupling strength could also be incorporated into the training of DNN[multiWGs, LAI] to reduce the dimension of input features. In that case, additional research would be needed to understand how surface–root zone soil moisture coupling could influence the training of DNN[multiWGs, LAI] and the resulting DA performance.

The observation and background errors are based on previous studies (Bonan et al., 2020; Lievens, Martens, et al., 2017; Shan et al., 2024). The usage of 0.33 dB as the prescribed ASCAT backscatter error has been investigated in Shan et al. (2024) by diagnosing the innovations and residuals of the DA system. The magnitudes of pre-assumed model and observation errors tend to overestimate the magnitudes of prescribed model errors. This overestimation varies with land cover type, but is less pronounced in agricultural GPIs. In Shan et al. (2024), an additional DA experiment with reduced prescribed observation errors does not improve the DA performances (Shan et al., 2024), indicating that the Jacobian-related issues lie more in the DNN training process than in the settings of the DA system including prescribed observation error.

While this study focuses on the accuracy of DNN-estimated Jacobians and their impact in SEKF system, the implication extend beyond Jacobian-based filters. The reported degradation reflects a broader misrepresentation of state observation sensitivities that manifest differently across data assimilation methods. In SEKF, this sensitivity is represented explicitly by local derivatives (Jacobians). In EnKF, Jacobians are not explicitly required, but the update still depends on the same underlying sensitivity structure through ensemble-derived state-observation cross covariances. These perturbation-based mappings still depend on the underlying state–observation sensitivity structure. Therefore, physically implausible local derivatives in SEKF correspond to spurious state–observation cross-covariances in EnKF, which can project observational information onto inappropriate state dimensions. In EnKF implementations, this sensitivity inconsistency may appear as spurious state–observation correlations, potentially causing excessive vertical information propagation (e.g., false increments in deeper soil layers) and localization-dependent artifacts. These issues thus degrade the assimilation performance. An additional consideration is that sensitivity inconsistencies in EnKF may also arise from the perturbations used to generate the ensemble covariance matrix. Since EnKF relies on perturbed model simulations to estimate state-observation cross-covariances, biases introduced by these perturbations may be reflected in the inferred sensitivity structure. Thus, in EnKF systems, part of the sensitivity bias may be intrinsically linked to the perturbation-based estimation of the covariance matrix.

From a data assimilation perspective, one of the core risks is that statistically informative but physically weak-causal (or non-causal) inputs may corrupt the learned sensitivity structure of observation operators. In our case, including soil moisture from deeper soil layers into the training may improve apparent predictive skill through correlation, but it can also induce sensitivity leakage across state dimensions that are not directly observed by the microwave signal. As a result, the learned operator may attribute observational variations to the wrong state components. This issue has direct consequences across different DA frameworks: in Jacobian-based systems (e.g., SEKF), it appears as physically inconsistent local derivatives; in ensemble-based systems, it appears as spurious state–observation cross-covariances. In both cases, observational information is projected into inappropriate parts of the state vector, degrading DA performances even when fitting performance metrics of DNN are acceptable. This suggests a general cautionary principle: feature selection for learning-based observation operators should be guided not only by statistical informativeness, but also by sensitivity (or response of observation to state vectors).

6. Conclusions

This study compares three synthetic data assimilation experiments using different observation operators. Performance skills and Jacobians of DNNs are evaluated against the WCM which serves as the synthetic true observation operator. Data assimilation performances using different observation operators are also evaluated. Although real ASCAT backscatter is not assimilated in this study, the synthetic experiment enhances our understanding about the DNN-based observation operator. When using the OL as input to estimate σ^o_{true}, the two DNNs have smaller ubRMSE and bias and larger ρ compared to the WCM. Given the assumption that the WCM is perfect in our synthetic experiment, this suggests that two DNNs are overfitting because they are trained on the OL data. The $J(\sigma^o_{\text{true}},\text{WG2})$ of DNN[WG2, LAI] is overestimated in autumn. This is due to the large monthly bias between WG2 from OL and synthetic truth in August and September. In winter, both DNNs overestimate $J(\sigma^o_{\text{true}},\text{LAI})$ in a period when LAI is close to zero and σ should primarily be governed by soil moisture. The comparison between two DNNs and WCM highlights the need to train ML-based observation operators with inputs with smaller bias and errors. The comparison between DNN[WG2, LAI] and DNN[multiple WGs, LAI] shows that including soil moisture in other layers slightly improves the performance skills but yields a worse estimate of $J(\sigma^o_{\text{true}},\text{WG2})$. DNN[multiple WGs, LAI] distributes the sensitivity to soil water dynamics to soil moisture in different layers. These underestimates of $J(\sigma^o_{\text{true}},\text{WG2})$ leads to a smaller $\widetilde{\mathbf{J}}$ to WG2 in Kalman gain.

The results from DA(DNN[WG2, LAI]) show that the DA using a ML-based observation operator improves the estimates of ISBA. The performance of WG2 from DA(DNN[multiple WGs, LAI]) is worse than DA(DNN[WG2, LAI]) because of the underestimates of $J(\sigma^o_{\text{true}},\text{WG2})$. Additionally, the non-zero sensitivities of DNN[multiple WGs, LAI] to soil moisture in deeper layers enlarge the ubRMSE of deeper soil moisture (like WG6, WG8) from DA(DNN[multiple WGs, LAI]) in autumn. The suboptimal DA performance of DNN[multiple WGs, LAI] underscores the importance of the choice of input data used to train the DNN model.

This study highlights the importance of examining the physical plausibility and robustness of the ML-based observation operators, as well as their prediction performance. This is particularly relevant given the growing interest in including machine learning models as surrogates for physical models in data assimilation.

Conflict of Interest

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

Availability Statement

Data of ISBA-A-gs open loop and ASCAT observables are available at https://data.4tu.nl/private_datasets/0w4LR9qOApQK3Kwjh-evuuY5pVQzzfS-bTPKnmvvqc4. ISBA-A-gs software can be found at https://www.umr-cnrm.fr/surfex/spip.php?rubrique10.

Acknowledgments
The work of Xu Shan and Susan C. Steele-Dunne was supported by the Dutch Research Council (NWO) under Grants ALWGO.2018.036 and ENW.GO.001.034, as part of the "Use of space infrastructure for Earth observation and planetary research" research programme, and by the Netherlands eScience Center under Grant NLeSC C19.007. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

References

Aires, F., Prigent, C., & Rossow, W. B. (2004). Neural network uncertainty assessment using bayesian statistics: A remote sensing application. *Neural Computation*, *16*(11), 2415–2458. https://doi.org/10.1162/0899766041941925

Aires, F., Weston, P., Rosnay, P., & Fairbairn, D. (2021). Statistical approaches to assimilate ASCAT soil moisture Information—I. Methodologies and first assessment. *Quarterly Journal of the Royal Meteorological Society*, *147*(736), 1823–1852. https://doi.org/10.1002/qj.3997

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., et al. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, *118*, 215–226. https://doi.org/10.1016/j.rse.2011.11.017

Albergel, C., Munier, S., Bocher, A., Bonan, B., Zheng, Y., Draper, C., et al. (2018). LDAS-Monde sequential assimilation of satellite derived observations applied to the contiguous US: An ERA-5 driven reanalysis of the land surface variables. *Remote Sensing*, *10*(10), 1627. https://doi.org/10.3390/rs10101627

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., et al. (2017). Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the euro-mediterranean area. *Geoscientific Model Development*, *10*(10), 3889–3912. https://doi.org/10.5194/gmd-10-3889-2017

Attema, E. P. W., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. *Radio Science*, *13*(2), 357–364. https://doi.org/10.1029/rs013i002p00357

Baguis, P., Carrassi, A., Roulin, E., Vannitsem, S., Modanesi, S., Lievens, H., et al. (2022). Assimilation of backscatter observations into a hydrological model: A case study in belgium using ASCAT data. *Remote Sensing*, *14*(22), 5740. https://doi.org/10.3390/rs14225740

Balsamo, G., Mahfouf, J.-F., Bélair, S., & Deblonde, G. (2006). A global root-zone soil moisture analysis using simulated l-band brightness temperature in preparation for the hydros satellite mission. *Journal of Hydrometeorology*, *7*(5), 1126–1146. https://doi.org/10.1175/jhm525.1

Barbu, A., Calvet, J.-C., Mahfouf, J.-F., Albergel, C., & Lafont, S. (2011). Assimilation of soil wetness index and leaf area index into the isba-a-gs land surface model: grassland case study. *Biogeosciences*, *8*(7), 1971–1986. https://doi.org/10.5194/bg-8-1971-2011

Barbu, A., Calvet, J.-C., Mahfouf, J.-F., & Lafont, S. (2014). Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: A land data assimilation application over France. *Hydrology and Earth System Sciences*, *18*(1), 173–192. https://doi.org/10.5194/hess-18-173-2014

Bhouri, M. A., & Gentine, P. (2023). Memory-based parameterization with differentiable solver: Application to lorenz'96. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *33*(7), 073116. https://doi.org/10.1063/5.0131929

Bindlish, R., & Barros, A. P. (2001). Parameterization of vegetation backscatter in radar-based, soil moisture estimation. *Remote Sensing of Environment*, *76*(1), 130–137. https://doi.org/10.1016/s0034-4257(00)00200-5

Bonan, B., Albergel, C., Zheng, Y., Barbu, A. L., Fairbairn, D., Munier, S., & Calvet, J.-C. (2020). An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the land data assimilation system LDAS-monde: Application over the euro-mediterranean region. *Hydrology and Earth System Sciences*, *24*(1), 325–347. https://doi.org/10.5194/hess-24-325-2020

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences*, *14*(10), 1881–1893. https://doi.org/10.5194/hess-14-1881-2010

Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., & Piguet, B. (2007). In situ soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network. In *2007 IEEE international geoscience and remote sensing symposium*. IEEE. https://doi.org/10.1109/igarss.2007.4423019

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., & Wigneron, J.-P. (1998). An interactive vegetation SVAT model tested against data from six contrasting sites. *Agricultural and Forest Meteorology*, *92*(2), 73–95. https://doi.org/10.1016/s0168-1923(98)00091-4

Calvet, J.-C., Rivalland, V., Picon-Cochard, C., & Guehl, J.-M. (2004). Modelling forest transpiration and CO2 fluxes—Response to soil moisture stress. *Agricultural and Forest Meteorology*, *124*(3–4), 143–156. https://doi.org/10.1016/j.agrformet.2004.01.007

Corchia, T., Bonan, B., Rodríguez-Fernández, N., Colas, G., & Calvet, J.-C. (2023). Assimilation of ASCAT radar backscatter coefficients over southwestern France. *Remote Sensing*, *15*(17), 4258. https://doi.org/10.3390/rs15174258

Crow, W. T., & Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble kalman filtering: A case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, *26*(2), 137–149. https://doi.org/10.1016/s0309-1708(02)00088-x

Danquechin Dorval, A., Meredieu, C., & Danjon, F. (2016). Anchorage failure of young trees in sandy soils is prevented by a rigid central part of the root system with various designs. *Annals of Botany*, *118*(4), 747–762. https://doi.org/10.1093/aob/mcw098

Decharme, B., Boone, A., Delire, C., & Noilhan, J. (2011). Local evaluation of the interaction between soil biosphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *Journal of Geophysical Research*, *116*(D20), D20126. https://doi.org/10.1029/2011jd016002

De Lannoy, G. J. M., & Reichle, R. H. (2016). Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences*, *20*(12), 4895–4911. https://doi.org/10.5194/hess-20-4895-2016

de Roos, S., Busschaert, L., Lievens, H., Bechtold, M., & Lannoy, G. J. D. (2023). Optimisation of AquaCrop backscatter simulations using sentinel-1 observations. *Remote Sensing of Environment*, *294*, 113621. https://doi.org/10.1016/j.rse.2023.113621

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys*, *27*(3), 326–327. https://doi.org/10.1145/212094.212114

Draper, C. S., Mahfouf, J.-F., Calvet, J.-C., Martin, E., & Wagner, W. (2011). Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences*, *15*(12), 3829–3841. https://doi.org/10.5194/hess-15-3829-2011

Draper, C. S., Reichle, R. H., Lannoy, G. J. M. D., & 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., 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

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the sce-ua global optimization method for calibrating watershed models. *Journal of Hydrology*, *158*(3–4), 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Entekhabi, D., Njoku, E. G., ONeill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/jproc.2010.2043918

Fairbairn, D., Barbu, A. L., Mahfouf, J.-F., Calvet, J.-C., & Gelati, E. (2015). Comparing the ensemble and extended kalman filters for in situ soil moisture assimilation with contrasting conditions. *Hydrology and Earth System Sciences*, *19*(12), 4811–4830. https://doi.org/10.5194/hess-19-4811-2015

Fairbairn, D., Barbu, A. L., Napoly, A., Albergel, C., Mahfouf, J.-F., & Calvet, J.-C. (2017). The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France. *Hydrology and Earth System Sciences*, *21*(4), 2015–2033. https://doi.org/10.5194/hess-21-2015-2017

Forman, B. A., & Xue, Y. (2016). Machine learning predictions of passive microwave brightness temperature over snow-covered land using the special sensor microwave imager (SSM/I). *Physical Geography*, *38*(2), 176–196. https://doi.org/10.1080/02723646.2016.1236606

Frolking, S., Fahnestock, M., Milliman, T., McDonald, K., & Kimball, J. (2005). Interannual variability in north American grassland biomass/productivity detected by SeaWinds scatterometer backscatter. *Geophysical Research Letters*, *32*(21). https://doi.org/10.1029/2005gl024230

Frost, C., & Thompson, S. G. (2000). Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society - Series A: Statistics in Society*, *163*(2), 173–189. https://doi.org/10.1111/1467-985x.00164

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Hahn, S., Reimer, C., Vreugdenhil, M., Melzer, T., & Wagner, W. (2017). Dynamic characterization of the incidence angle dependence of backscatter using metop ASCAT. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 2348–2359. https://doi.org/10.1109/jstars.2016.2628523

Han, X., 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

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

Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., & Oki, T. (2010). Instrument performance and calibration of amsr-e and amsr2. International archives of the photogrammetry. *remote sensing and spatial information science*, *38*(8), 13–18.

Jacobs, C., van den Hurk, B., & de Bruin, H. (1996). Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions. *Agricultural and Forest Meteorology*, *80*(2–4), 111–134. https://doi.org/10.1016/0168-1923(95)02295-3
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. *Springer*, *112*(1). Retrieved from https://link.springer.com/book/10.1007/978-1-0716-1418-1
Jarlan, L., Mougin, E., Frison, P., Mazzega, P., & Hiernaux, P. (2002). Analysis of ERS wind scatterometer time series over Sahel (mali). *Remote Sensing of Environment*, *81*(2–3), 404–415. https://doi.org/10.1016/s0034-4257(02)00015-9
Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., et al. (2010). The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proceedings of the IEEE*, *98*(5), 666–687. https://doi.org/10.1109/jproc.2010.2043032
Khabbazan, S., Steele-Dunne, S., Vermunt, P., Judge, J., Vreugdenhil, M., & Gao, G. (2022). The influence of surface canopy water on the relationship between l-band backscatter and biophysical variables in agricultural monitoring. *Remote Sensing of Environment*, *268*, 112789. https://doi.org/10.1016/j.rse.2021.112789
Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
Krasnopolsky, V. M. (2007). Reducing uncertainties in neural network Jacobians and improving accuracy of neural network emulations with nn ensemble approaches. *Neural Networks*, *20*(4), 454–461. https://doi.org/10.1016/j.neunet.2007.04.008
Kumar, S. V., Kolassa, J., Reichle, R., Crow, W., de Lannoy, G., de Rosnay, P., et al. (2022). An agenda for land data assimilation priorities: Realizing the promise of terrestrial water, energy, and vegetation observations from space. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003259. https://doi.org/10.1029/2022ms003259
Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., & Peters-Lidard, C. D. (2009). Role of subsurface physics in the assimilation of surface soil moisture observations. *Journal of Hydrometeorology*, *10*(6), 1534–1547. https://doi.org/10.1175/2009jhm1134.1
Lannoy, G. J. M. D., Bechtold, M., Albergel, C., Brocca, L., Calvet, J.-C., Carrassi, A., et al. (2022). Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication. *Frontiers in Water*, *4*, 981745. https://doi.org/10.3389/frwa.2022.981745
Leroux, D., Calvet, J.-C., Munier, S., & Albergel, C. (2018). Using satellite-derived vegetation products to evaluate LDAS-monde over the euro-mediterranean area. *Remote Sensing*, *10*(8), 1199. https://doi.org/10.3390/rs10081199
Li, L., Carver, R., Lopez-Gomez, I., Sha, F., & Anderson, J. (2024). Generative emulation of weather forecast ensembles with diffusion models. *Science Advances*, *10*(13), eadk4489. https://doi.org/10.1126/sciadv.adk4489
Li, S., Zheng, T., Farchi, A., Bocquet, M., & Gentine, P. (2025). Probabilistic data assimilation for ensemble distribution projections with generative machine learning: A lorenz'96 proof-of-concept. *Geophysical Research Letters*, *52*(12), e2024GL112523. https://doi.org/10.1029/2024gl112523
Lievens, H., Martens, B., Verhoest, N., Hahn, S., Reichle, R., & Miralles, D. (2017). Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates. *Remote Sensing of Environment*, *189*, 194–210. https://doi.org/10.1016/j.rse.2016.11.022
Lievens, H., Reichle, R. H., Liu, Q., De Lannoy, G. J., Dunbar, R. S., Kim, S., et al. (2017). Joint sentinel-1 and smap data assimilation to improve soil moisture estimates. *Geophysical Research Letters*, *44*(12), 6145–6153. https://doi.org/10.1002/2017gl073904
MacMahon, S., Peto, R., Collins, R., Godwin, J., Cutler, J., Sorlie, P., & Stamler, J. (1990). Blood pressure, stroke, and coronary heart disease: Part 1, prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *The Lancet*, *335*(8692), 765–774.
Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, F., Taillefer, F., & Taseva, L. (2009). A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *Journal of Geophysical Research*, *114*(D8). https://doi.org/10.1029/2008jd011077
Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., & Verhoest, N. (2016). Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. *International Journal of Applied Earth Observation and Geoinformation*, *48*, 146–162. https://doi.org/10.1016/j.jag.2015.09.012
Masson, V., Moigne, P. L., Martin, E., Faroux, S., Alias, A., Alkama, R., et al. (2013). The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes. *Geoscientific Model Development*, *6*(4), 929–960. https://doi.org/10.5194/gmd-6-929-2013
Melzer, T. (2013). Vegetation modelling in warp 6.0. In *Proceedings of the eumetsat meteorological satellite conference* (pp. 16–20). austria.
Modanesi, S., Massari, C., Bechtold, M., Lievens, H., Tarpanelli, A., Brocca, L., et al. (2022). Challenges and benefits of quantifying irrigation through the assimilation of sentinel-1 backscatter observations into noah-MP. *Hydrology and Earth System Sciences*, *26*(18), 4685–4706. https://doi.org/10.5194/hess-26-4685-2022
Modanesi, S., Massari, C., Gruber, A., Lievens, H., Tarpanelli, A., Morbidelli, R., & De Lannoy, G. J. M. (2021). Optimizing a backscatter forward operator using sentinel-1 data over irrigated land. *Hydrology and Earth System Sciences*, *25*(12), 6283–6307. https://doi.org/10.5194/hess-25-6283-2021
Nikaein, T., & Lopez-Dekker, P. (2025). Physics-guided machine learning based forward-modeling of radar observables: A case study on sentinel-1 observations of corn-fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
Noilhan, J., & Mahfouf, J.-F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, *13*(1–4), 145–159. https://doi.org/10.1016/0921-8181(95)00043-7
Noilhan, J., & Planton, S. (1989). A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, *117*(3), 536–549. https://doi.org/10.1175/1520-0493(1989)117⟨0536:aspols⟩2.0.co;2
Parrens, M., Mahfouf, J.-F., Barbu, A., & Calvet, J.-C. (2014). Assimilation of surface soil moisture into a multilayer soil model: Design and evaluation at local scale. *Hydrology and Earth System Sciences*, *18*(2), 673–689. https://doi.org/10.5194/hess-18-673-2014
Petchiappan, A., Steele-Dunne, S. C., Vreugdenhil, M., Hahn, S., Wagner, W., & Oliveira, R. (2021). The influence of vegetation water dynamics on the ASCAT backscatter-incidence angle relationship in the Amazon. https://doi.org/10.5194/hess-2021-406
Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007wr006357
Reichle, R. H., Lannoy, G. J. M. D., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W. T., et al. (2017). Global assessment of the SMAP level-4 surface and root-zone soil moisture product using assimilation diagnostics. *Journal of Hydrometeorology*, *18*(12), 3217–3237. https://doi.org/10.1175/jhm-d-17-0130.1
Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002). Extended versus ensemble kalman filtering for land data assimilation. *Journal of Hydrometeorology*, *3*(6), 728–740. https://doi.org/10.1175/1525-7541(2002)003⟨0728:evekff⟩2.0.co;2
Reichle, R. H., Zhang, S. Q., Liu, Q., Draper, C. S., Kolassa, J., & Todling, R. (2021). Assimilation of SMAP brightness temperature observations in the GEOS land–atmosphere data assimilation system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 10628–10643. https://doi.org/10.1109/jstars.2021.3118595

Rosner, B., Willett, W., & Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, *8*(9), 1051–1069. https://doi.org/10.1002/sim.4780080905

Sawada, Y., Koike, T., & Walker, J. P. (2015). A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics. *Journal of Geophysical Research: Atmospheres*, *120*(12), 5910–5930. https://doi.org/10.1002/2014jd022895

Schroeder, R., McDonald, K. C., Azarderakhsh, M., & Zimmermann, R. (2016). ASCAT MetOp-a diurnal backscatter observations of recent vegetation drought patterns over the contiguous u.s.: An assessment of spatial extent and relationship with precipitation and crop yield. *Remote Sensing of Environment*, *177*, 153–159. https://doi.org/10.1016/j.rse.2016.01.008

Seo, E., Lee, M.-I., & Reichle, R. H. (2021). Assimilation of SMAP and ASCAT soil moisture retrievals into the JULES land surface model using the local ensemble transform kalman filter. *Remote Sensing of Environment*, *253*, 112222. https://doi.org/10.1016/j.rse.2020.112222

Shamambo, D. C. (2020). Satellite data assimilation for water resource monitoring in the euro-mediterranean area. https://doi.org/10.70675/b1920716z24a9z4ab2za86az46445d9dc863 (Doctoral thesis).

Shamambo, D. C., Bonan, B., Calvet, J.-C., Albergel, C., & Hahn, S. (2019). Interpretation of ASCAT radar scatterometer observations over land: A case study over southwestern France. *Remote Sensing*, *11*(23), 2842. https://doi.org/10.3390/rs11232842

Shan, X., Steele-Dunne, S., Hahn, S., Wagner, W., Bonan, B., Albergel, C., et al. (2024). Assimilating ascat normalized backscatter and slope into the land surface model isba-a-gs using a deep neural network as the observation operator: Case studies at ismn stations in western Europe. *Remote Sensing of Environment*, *308*, 114167. https://doi.org/10.1016/j.rse.2024.114167

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., et al. (2022). Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a deep neural network. *Remote Sensing of Environment*, *279*, 113116. https://doi.org/10.1016/j.rse.2022.113116

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951–2959).

Steele-Dunne, S. C., Hahn, S., Wagner, W., & Vreugdenhil, M. (2019). Investigating vegetation water dynamics and drought using metop ASCAT over the north American grasslands. *Remote Sensing of Environment*, *224*, 219–235. https://doi.org/10.1016/j.rse.2019.01.004

Teuling, A. J., Taylor, C. M., Meirink, J. F., Melsen, L. A., Miralles, D. G., van Heerwaarden, C. C., et al. (2017). Observational evidence for cloud cover enhancement over Western European forests. *Nature Communications*, *8*(1), 14065. https://doi.org/10.1038/ncomms14065

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment*, *70*(2), 191–207. https://doi.org/10.1016/s0034-4257(99)00036-x

Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., et al. (2007). L-band microwave emission of the biosphere (l-MEB) model: Description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, *107*(4), 639–655. https://doi.org/10.1016/j.rse.2006.10.014

Xue, Y., Forman, B. A., & Reichle, R. H. (2018). Estimating snow mass in North America through assimilation of advanced microwave scanning radiometer brightness temperature observations using the catchment land surface model and support vector machines. *Water Resources Research*, *54*(9), 6488–6509. https://doi.org/10.1029/2017wr022219

Ying, X. (2019). An overview of overfitting and its solutions. In *Journal of Physics: Conference Series* (Vol. 1168, p. 022022), https://doi.org/10.1088/1742-6596/1168/2/022022

References From the Supporting Information

Boone, A., Samuelsson, P., Gollvik, S., Napoly, A., Jarlan, L., Brun, E., & Decharme, B. (2017). The interactions between soil–biosphere–atmosphere land surface model with a multi-energy balance (ISBA-MEB) option in SURFEXv8 – Part 1: Model description. *Geoscientific Model Development*, *10*(2), 843–872. https://doi.org/10.5194/gmd-10-843-2017

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water Resources Research*, *14*(4), 601–604. https://doi.org/10.1029/wr014i004p00601

Green, W. H., & Ampt, G. (1911). Studies on soil phyics. *The Journal of Agricultural Science*, *4*(1), 1–24. https://doi.org/10.1017/s0021859600001441

Noilhan, J., & Lacarrere, P. (1995). Gcm grid-scale evaporation from mesoscale modeling. *Journal of Climate*, *8*(2), 206–223. https://doi.org/10.1175/1520-0442(1995)008<0206:ggsefm>2.0.co;2