A new perspective on vegetation water dynamics

Assimilating ASCAT observations to constrain soil and vegetation states using a data-driven observation operator

Shan, X.

DOI
10.4233/uuid:95a9b453-6d46-42f9-9e04-dfe9efae54ae

Publication date
2024

Document Version
Final published version

Citation (APA)
Shan, X. (2024). *A new perspective on vegetation water dynamics: Assimilating ASCAT observations to constrain soil and vegetation states using a data-driven observation operator* . [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:95a9b453-6d46-42f9-9e04-dfe9efae54ae

A New Perspective on Vegetation Water Dynamics

Assimilating ASCAT observations to constrain soil and vegetation states using a data-driven observation operator

A New Perspective on Vegetation Water Dynamics

Assimilating ASCAT observations to constrain soil and vegetation states using a data-driven observation operator

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op Woensdag 30, October, 2024 om 15:00 o'clock

door

Xu SHAN

Bachelor of Engineering in Hydraulic and Hydropower Engineering,
Bachelor of Science in Pure and Applied Mathematics,
Tsinghua University, Beijing, China
geboren te Shanxi Province, China

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. ir. S. C. Steele-Dunne
copromotor: Dr. ir. F. J. Lopez Dekker

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. S. C. Steele-Dunne,	Technische Universiteit Delft
Dr. ir. F. J. Lopez Dekker,	Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. ir. N. C. van de Giesen,	Technische Universiteit Delft
Prof. dr. D. Entekhabi,	Massachusetts Institute of Technology, USA
Prof. dr. W. Wagner,	Technische Universität Wien, Austria
Dr. J.-C. Calvet,	Météo France, France
Dr. M. Scholze,	Lund University, Sweden

Overig lid:

Prof. dr. ir. H. W. J. Russchenberg,	Technische Universiteit Delft

Keywords: ASCAT, Land surface model, Machine learning, Land data assimilation, Soil moisture, Vegetation
Printed by: Proefschriftspecialist.nl
Front & Back: Xu Shan, YIOU

ISBN 978-94-6384-658-5

An electronic copy of this dissertation is available at
https://repository.tudelft.nl/.

To my parents

CONTENTS

Summary xi

Samenvatting xiii

Preface xv

1. Introduction 1
1.1. Vegetation water dynamics in land surface model . . . 2
1.2. Radar backscatter from vegetation and soil . . . 3
1.3. ASCAT backscatter incidence-angle dependence . . . 4
1.4. Data assimilation of ASCAT data . . . 6
1.5. Machine-learning-based observation operators . . . 8
1.6. Research outline . . . 9

2. Background 11
2.1. Land surface model ISBA-A-gs . . . 12
2.2. ASCAT dynamic vegetation parameters . . . 14

3. Modeling ASCAT backscatter and slope with a Deep Neural Network 19
3.1. Introduction . . . 20
3.2. Method and materials . . . 20
3.2.1. Study area . . . 20
3.2.2. ASCAT Backscatter-Incidence angle dependence . . . 20
3.2.3. Interactions between Soil, Biosphere and Atmosphere (ISBA) Model . 22
3.2.4. Deep Neural Network . . . 22
3.2.5. Normalized Sensitivity Coefficient . . . 25
3.3. Results and Discussion . . . 26
3.3.1. Model performance . . . 26
3.3.2. Model sensitivity: A qualitative analysis . . . 30
3.3.3. Model sensitivity: Normalized Sensitivity Coefficient . . . 33
3.4. Conclusions . . . 35

4. Assimilating ASCAT normalized backscatter and slope into ISBA 39
4.1. Introduction . . . 40
4.2. Data and Methodology . . . 40
4.2.1. Data . . . 40
4.2.2. Methodology . . . 41
4.3. Results . . . 46
4.3.1. Performance . . . 46

4.3.2. Data Assimilation Diagnostics . . . 52
4.4. Discussion . . . 57
4.5. Conclusions . . . 61

5. Evaluating the ML-based observation operator in land data assimilation framework 63
5.1. Introduction . . . 64
5.2. Data . . . 64
5.2.1. ASCAT normalized backscatter . . . 64
5.2.2. Study area . . . 64
5.3. Methodology . . . 65
5.3.1. Water Cloud Model . . . 65
5.3.2. Experiment setup . . . 66
5.4. Results . . . 70
5.4.1. WCM calibration . . . 70
5.4.2. Synthetic truth and observations . . . 71
5.4.3. Comparison of different observation operators . . . 71
5.4.4. Jacobians of different observation operators . . . 78
5.4.5. Data assimilation results . . . 79
5.5. Discussion . . . 86
5.6. Conclusions . . . 86

6. Conclusion 89
6.1. Knowledge generated and original contributions . . . 89
6.1.1. On estimating radar observables using ISBA states . . . 89
6.1.2. On land data assimilation of ASCAT observables . . . 90
6.1.3. On the evaluation of machine-learning based observation operator . 90
6.1.4. On constraining land surface model using ASCAT observables . . . 91
6.2. Future research . . . 91
6.2.1. Reducing uncertainty of plant hydraulic schemes using ASCAT observables . . . 91
6.2.2. Estimating ASCAT slope and curvature using radiative transfer models 92
6.2.3. Improving the robustness of the Jacobian of the machine-learning based observation operator . . . 92
6.2.4. Correcting the bias and errors in land surface model simulations . . . 93
6.2.5. Understanding the value of ASCAT slope and curvature for carbon cycle science . . . 94
6.2.6. Assimilating multi-source earth observations into earth system models . . . 94
6.2.7. Using ASCAT dynamic vegetation parameters to monitor climate extremes . . . 94

A. Supplementary material 121

Acknowledgements 141

Curriculum Vitæ 143

List of Publications **145**

SUMMARY

In the current generation, most land surface models (LSMs) do not explicitly model the plant hydraulic states or fluxes, which limits the ability of LSMs to model evapotranspiration [1], stomatal conductance [2], and monitor and predict drought [3]. Therefore it is necessary to constrain the canopy water dynamics in LSMs. Advanced SCATterometer (ASCAT) provides a long record of C-band backscatter since 2007. A key advantage of the ASCAT instrument is the ability to obtain measurements of the Earth's surface from different incidence angles. The dependence of ASCAT backscattering coefficient (hereafter referred to as backscatter) on incidence angle provides valuable information about vegetation water dynamics via normalized backscatter (σ^o_{40}), and vegetation parameters (slope (σ'), and curvature (σ'')) of the Taylor expansions of backscatter to incidence angle [4–7]. In this thesis, the ASCAT normalized backscatter and slope are assimilated into the "Interactions between soil, biosphere and atmosphere" (ISBA-A-gs, hereafter referred to as ISBA) LSM.

In order to assimilate ASCAT observables, an observation operator is needed to link between LSM states to radar observations. Radiative transfer models (RTMs) are often used to assimilate radar backscatter into LSM [8, 9]. However, RTMs require moisture content or dielectric properties of soil and vegetation cover which are not simulated by the LSM. Therefore, to directly link land surface states and ASCAT observables, a Deep Neural Network (DNN) was trained and validated in Chapter 3. The performances and sensitivity of the DNN were evaluated to make sure the observation operator is physically plausible.

In Chapter 4, ASCAT σ^o_{40} and σ' are assimilated into the ISBA-A-gs LSM using the so-called Simplified Extended Kalman Filter (SEKF). The DNN trained and validated in Chapter 3 was used as the observation operator. The data assimilation (DA) performances were evaluated against the International Soil Moisture Network (ISMN) in-situ soil moisture measurements. The data assimilation (DA) performances are neutral compared to the open loop (OL), evaluated against in-situ soil moisture measurements. Chapter 4 presented several reasons that might limit the DA performances, like representativeness errors, error specifications, as well as the robustness of Jacobians of machine-learning-based observation operators.

Motivated by the fact that the DA performances are limited by the robustness of Jacobians of machine-learning-based observation operators, a synthetic data assimilation experiment was performed to compare the DNN to the water cloud model (WCM) in the context of data assimilation. Results and discussions were presented in Chapter 5. The synthetic data assimilation experiment proved that assimilating normalized backscatter using a DNN as the observation operator improves the estimation of surface soil moisture. However, the improvement is sub-optimal because using the ISBA OL data (which is biased compared to synthetic truth) as input training data produces Jacobians that

differ from the true Jacobians. In addition, the inclusion of soil moisture in deeper layers may also contribute to errors in the Jacobians though the estimation performances are slightly improved. Therefore, it is recommended to carefully examine the observation operator in the land data assimilation system.

This thesis demonstrates that it is feasible to constrain the land surface model by assimilating normalized backscatter using a data-driven observation operator without explicitly modeling scattering from the vegetation and soil. However, it also points out that this approach might be suboptimal due to the overfitting from the redundant input variables (indirectly related to training targets) and the bias in the training data, both of which influence the Jacobians. A key takeaway is that the evaluation of Jacobians requires more attention when data-driven observation operators are used. Possible solutions may include training the DNN with independent satellite observations, or providing error information of the LSM by training the DNN on an ensemble of the LSM, or LSMs. Careful evaluation of the input choice also matters to avoid potential overfitting. Future work could focus on transferring the methodology developed in this thesis to other land surface models, and other satellite observations. The use of radiative transfer modeling is also recommended to further investigate the information in the ASCAT slope and its potential benefit in assimilation. The efforts made in this thesis will also support future studies aiming to assimilate multiple earth observations to better constrain earth system models.

SAMENVATTING

De huidige generatie van landoppervlaktemodellen (LSM) modelleert de hydraulische staten of fluxen van planten niet expliciet, wat de mogelijkheid van LSM's beperkt om evapotranspiratie [1], stomatale geleidbaarheid [2] en droogte te monitoren en te voorspellen [3]. Daarom is het noodzakelijk om de waterdynamiek van het bladerdak in LSM's te beperken. De Advanced SCATterometer (ASCAT) levert sinds 2007 een lange reeks van C-band terugkaatsing. Een belangrijk voordeel van het ASCAT-instrument is het vermogen om metingen van het aardoppervlak te verkrijgen vanuit verschillende invalshoeken. De afhankelijkheid tussen ASCAT-terugkaatsing en de invalshoek biedt waardevolle informatie over de waterdynamiek van de vegetatie via genormaliseerde terugkaatsing (σ^o_{40}), helling (σ') en kromming (σ''), afgeleid uit de Taylor-uitbreidingen van de relatie tussen terugkaatsing en invalshoek [4–7]. In dit proefschrift worden de genormaliseerde terugkaatsing en helling van ASCAT geassimileerd in het LSM ISBA-A-gs (hierna ISBA genoemd).

Om ASCAT-waarnemingen te assimileren, is een waarnemingsoperator nodig om een link te leggen tussen de LSM-staten en radarwaarnemingen. Stralingstransfermodellen worden vaak gebruikt om radar terugkaatsing in LSM te assimileren [8, 9]. Echter, RTM's vereisen informatie over vochtgehaltes of diëlektrische eigenschappen van de bodem en vegetatiebedekking die niet door LSM's worden gesimuleerd. Daarom werd in Hoofdstuk 3 een diep neuraal netwerk (DNN) getraind en gevalideerd om landoppervlaktetoestanden en ASCAT-waarnemingen rechtstreeks met elkaar te verbinden. De prestaties en gevoeligheid van het DNN werden geëvalueerd om te controleren of de waarnemingsoperator fysiek aannemelijk is.

In Hoofdstuk 4 worden ASCAT σ^o_{40} en σ' geassimileerd in de ISBA-A-gs LSM met behulp van het zogenaamde Simplified Extended Kalman Filter (SEKF). De DNN die in Hoofdstuk 3 werd getraind en gevalideerd, werd gebruikt als waarnemingsoperator. De prestaties van de data-assimilatie (DA) werden geëvalueerd aan de hand van de in-situ bodemvochtmetingen van het International Soil Moisture Network (ISMN). De DA-prestaties zijn neutraal in vergelijking met de open loop (OL). Hoofdstuk 4 presenteert verschillende redenen die de DA-prestaties kunnen beperken, zoals representativiteitsfouten, foutspecificaties, en de robuustheid van Jacobianen van machine-learning-gebaseerde waarnemingsoperatoren.

Gemotiveerd door het feit dat de DA-prestaties beperkt kunnen zijn door de robuustheid van Jacobianen op waarnemingsoperatoren gebaseerd op machine learning, werd een synthetisch data-assimilatie experiment uitgevoerd om de DNN te vergelijken met het water cloud model (WCM) in de context van data-assimilatie. De resultaten en discussies zijn gepresenteerd in Hoofdstuk 5. Het experiment met synthetische DA heeft bewees dat het assimileren van genormaliseerde terugkaatsing met behulp van een DNN als waarnemingsoperator de schatting van het bodemvocht aan het oppervlak verbetert.

De verbetering is echter suboptimaal omdat het gebruik van ISBA OL-gegevens als invoertrainingsgegevens problematische Jacobianen oplevert. Bovendien kan de opname van bodemvocht in diepere lagen ook bijdragen aan fouten in de Jacobianen, hoewel de schattingsprestaties enigszins zijn verbeterd. Daarom wordt aanbevolen om de waarnemingsoperator in het landgegevensassimilatiesysteem zorgvuldig te onderzoeken.

Dit proefschrift laat zien dat het haalbaar is om het landoppervlakmodel te beperken door genormaliseerde terugkaatsing te assimileren met behulp van een machine-learning-gebaseerde waarnemingsoperator zonder expliciet weerkaatsing van de vegetatie en bodem te modelleren. Het wijst er echter ook op dat deze aanpak suboptimaal kan zijn vanwege overfitting door de overtollige inputvariabelen en de bias in de trainingsgegevens, die beide de Jacobianen beïnvloeden. Een belangrijke conclusie is dat de evaluatie van Jacobianen meer aandacht vereist wanneer data-gedreven waarnemingsoperatoren worden gebruikt. Een mogelijke oplossingen is het verstrekken van foutinformatie van de LSM door het DNN te trainen op een ensemble van de LSM, of LSM's. Zorgvuldige evaluatie van de inputkeuze is ook van belang om mogelijke overfitting te vermijden. Toekomstig werk kan zich richten op het overdragen van de in dit proefschrift ontwikkelde methodologie naar andere landoppervlakmodellen. Het gebruik van stralingsoverdrachtmodellen wordt ook aanbevolen om de informatie in de ASCAT helling en het potentiële voordeel bij assimilatie verder te onderzoeken. De inspanningen in dit proefschrift zullen ook toekomstige studies ondersteunen die gericht zijn op het assimileren van meerdere aard-observaties om de systeemmodellen van de aarde beter te kunnen beperken.

PREFACE

Land surface models (LSMs) simulate the interactions between the atmosphere and land by modeling the water cycle (evapotranspiration and precipitation etc) and carbon cycle (respiration and photosynthesis etc). However, LSMs suffer from uncertainties in model structure and external forcing. Therefore, it is necessary to constrain the LSM using independent satellite observations. ASCAT dynamic vegetation parameters provide unique opportunities to explore the vegetation water dynamics. One key advantage is that ASCAT offers long-term data records with high temporal frequency. Additionally, the signals are directly observed by radar instruments rather than being retrieved from satellite observables, which might introduce cross-correlated errors between retrievals and model simulations. Therefore, this thesis proposes to assimilate ASCAT dynamic vegetation parameters to constrain the land surface model ISBA.

This thesis is the first study to assimilate ASCAT observables (not retrieved products) into the LSM ISBA using a DNN-based observation operator. Our goal is to constrain the plant hydraulic dynamics without explicitly adding equations or submodels which might increase the model complexity. Using a DNN instead of the typical radiative transfer model avoids the need to explicitly model the variations in vegetation water dynamics. The DNN captures the dynamics in vegetation water content by including ISBA land surface variables related to vegetation water transport processes to estimate ASCAT observables which are dependent on plant hydraulics.

However, the DA performances showed neutral improvements compared to the open loop. Reasons include representativeness errors, error specification, and the robustness of Jacobians. This thesis argues that the key factor limiting the efficiency of the DA scheme is the correct estimation of the Jacobians which control the degree to which the observations update the states. A critical insight from this work is that even a well-trained data-driven observation operator does not guarantee successful DA performance, especially if the Jacobian calculations are affected by too many predictors.

Therefore we ran a synthetic experiment to evaluate the machine-learning-based observation operator. The synthetic experiment successfully demonstrated that using the DNN as the observation operator to assimilate normalized backscatter improves the LSM simulation. However, the impact of the DA is reduced compared to the DA using WCM as the observation operator. When training the machine-learning-based observation operator, it is suggested to use input data as close as to “truth" as possible. Although the data-driven observation operator mapping from the open loop to the observation avoids the need for CDF-matching or bias correction, it becomes problematic for the estimation of Jacobians. Additionally, the choice of input set also matters. Adding more variables may improve the model performances (even in an independent validation period) by including indirect relationships between inputs and outputs, however, it might lead to overfitting in terms of Jacobians.

This thesis proves that the land surface model can be constrained by assimilating normalized backscatter using DNN as the observation operator. However, more research needs to be done about constraining plant water dynamics schemes by incorporating ASCAT slope into the model.

I hope this thesis could contribute to the fields of land data assimilation and microwave remote sensing. It also serves as a reminder that things are not always as smooth as we expect, and the journey to achieving the final target can be longer and more complex. While this thesis may not fully achieve the goal of constraining vegetation dynamics in the land surface model, I hope it sheds light on the path forward.

Xu Shan
Delft, May 2024

1

INTRODUCTION

1

1.1. VEGETATION WATER DYNAMICS IN LAND SURFACE MODEL

Land surface models (LSMs) simulate surface fluxes such as latent heat, sensible heat, and ground heat fluxes, and predict the evolution of surface state variables. Reliable estimations of water and carbon related state variables depend on an adequate representation of the exchange of water and carbon fluxes between the canopy and the atmosphere. However, numerous studies have shown that LSMs are deficient in simulating the carbon and water fluxes during seasonal-scale droughts [3, 10, 11] due to limited parameterizations of vegetation water and carbon dynamics [1, 12, 13].

A key point of modeling the vegetation water and carbon dynamics is stomatal conductance (g_s), as leaf stomata regulate transpiration and photosynthesis [14]. There are two kinds of representations for g_s in LSMs: a) the “multiplicative model" (or Jarvis model [15]) which represents g_s as the product of several functions, and each function responds to one single environmental variable; b) the “coupled photosynthesis stomatal conductance" models [16, 17] which rely on the linear relationship between leaf net photosynthesis (A_n) and g_s (also known as the Ball-Berry model or BB model (Ball *et al.* [18]) with some essential mathematical transformations [19, 20]). The first model does not account for the carbon feedback and requires calibrating many parameters, making the second representation (A-g_s) more commonly used in LSMs [21].

In the absence of water stress, Jacob's A-g_s model assumes a linear relationship between the A_n and the associated leaf conductance to water vapor (g_s) [16, 17]. However, the A-g_s model does not simulate how water flows in and out of the plant water storage. The effect of water stress is accounted for by representing the slope of the A-g_s relationship as empirical functions of soil water content [22], or multiplying normalized soil moisture by A_n [21]. However, these approaches are not sufficient enough to represent the water dynamics during extreme drought [23, 24] although they work under mild to moderate drought conditions [25].

Some studies [1, 2, 26] found that modeling plant hydraulics improves the estimation of soil and vegetation water dynamics during drought. A soil–plant–atmosphere continuum (SPA) model was compared with the BB model used in the Community Land Model (CLM) 4.5 [26]. The SPA model simulates changes in leaf water potential and leaf-specific hydraulic conductance. They found that the performance of the SPA model is comparable to the BB model when soil water stress is absent but is significantly better under limited soil water availability. Anderegg *et al.* [2] found that incorporating leaf water potential as Weibull curve ($g_s(\Psi_{leaf})$) into the BB model reduces the biases in stomatal conductance observed during drought conditions in current empirical stomatal conductance models. This finding is consistent with published curves using dry-down or dehydration experiments [27]. Cochard *et al.* [28] formalized the flow of water between different plant organs and achieved good agreement with observed soil water balance. However, the model introduces a large number of parameters, and the computational cost is high [28].

Despite some basic plant hydraulics schemes being considered in some large-scale models [29], the information contained in plant hydraulics is still incomplete in current-generation land surface models. Additionally, implementing current vegetation models (e.g., the finite element tree crown hydrodynamics (FETCH) model [30]) in regional or global LSMs is challenging. Increased complexity raises the number of model parameters with substantial uncertainties, reducing model robustness [12, 14]. Many parameters also vary widely and non-linearly within one single plant function in response to water stress [2]. This challenge is further exacerbated by the extremely large variability in hydraulic traits [1]. Another factor is computational cost [14, 28], as solving a transient nonlinear diffusion equation for soil-plant continuum requires substantial computational resources [30]. Thus we need to find a more general way to incorporate vegetation water dynamics into land surface models.

1.2. RADAR BACKSCATTER FROM VEGETATION AND SOIL

Microwaves are defined as electromagnetic radiations with frequencies ranging from 0.3 to 300 GHz [31]. Microwave instruments include radars (active) and radiometers (passive). Radiometers measure the electromagnetic radiation emitted from objects, while radars transmit an electromagnetic wave and measure the backscatter. Satellite observations from both active and passive microwave remote sensing are sensitive to plant water content variations from sub-daily to interannual scales [32–38]. This sensitivity to water is central to the value of microwave remote sensing for vegetation monitoring in a wide range of applications [31, 39, 40] including the parameterization and validation of LSMs [41, 42]. This section provides a brief introduction to active microwave remote sensing for vegetation.

Radar backscatter from a vegetated surface includes contributions from vegetation, soil and the interactions between the vegetation and the underlying soil [43]. The backscatter depends on the frequency, polarization, incidence, and azimuth angles of the microwaves. Interactions between microwaves and the canopy depend on the dielectric properties, size, shape, orientation, and roughness of leaves and stems, as well as their distributions within the canopy. The dielectric properties of vegetation are primarily determined by their water content [31, 43]. Interactions between microwaves and soil are influenced by soil roughness and dielectric properties, which are dependent on soil water content.

Several studies have demonstrated that C-band (4-8 GHz) backscatter (e.g., Advanced SCATterometer (ASCAT) and Sentinel-1) contains valuable information about vegetation and soil water dynamics [43–46]. Comparison of backscatter and slope (the first derivative of backscatter to the incidence angle) from European Remote-Sensing Satellite (ERS) scatterometer to precipitation and NDVI suggests that variations in backscatter are influenced by soil moisture, and the seasonal variations of slope are linked to vegetation dynamics [44]. The largest diurnal differences in ERS 1/2 backscatter correlate with the onset of water stress [33]. A study using the Michigan Microwave Canopy Scattering (MIMICS) model found that diurnal variations may be attributed to changes in the water content of leaves and trunks

[34]. During the 2011 and 2012 droughts, Schroeder *et al.* [47] found a correlation between drought patterns and negative anomalies of ASCAT C-band backscatter during morning overpasses. Vegetation optical depth (VOD) can be retrieved from ASCAT C-band backscatter by using the water cloud model [48]. This shows the potential of the ASCAT data for vegetation monitoring.

Several tower-based experiments also demonstrated diurnal variations in C-band backscatter [49, 50] over crop canopies. These variations during the plant growth period are attributed to the loss of canopy moisture during the day due to transpiration [49]. A study in a maize canopy [35] shows that backscatter is very sensitive to variations in vegetation water content during water stress, and this sensitivity varies with polarization and incidence angle.

Backscatter of higher frequency (Ku-band) is more sensitive to the vegetation canopy than lower frequencies [49]. Ku-band backscatter is more related to canopy properties such as canopy structure, vegetation density, and canopy biomass [42, 51]. Saatchi *et al.* [52] identified the importance of the annual mean and standard deviation values of Ku-band backscatter from Quick Scatterometer (QSCAT) for estimating tropical forest tree biodiversity. There were negative anomalies in both morning-overpass backscatter from Ku-band SeaWinds scatterometer and accumulating water deficit during the 2005 Amazon drought [53]. The difference between the interannual variability of morning and afternoon backscatter is attributed to the inability of the canopy to recover leaf moisture content overnight [53]. An opposite seasonality between rainfall and 6:00 A.M. QSCAT HH backscatter was found in deciduous evergreen forests in Africa [42]. Notice that at around 6:00 A.M. plant water content reaches a pre-dawn or dawn maximum value. The dry season increase in HH Ku-band backscatter may be due to a combination of increases in transpiration [54, 55] and foliar biomass during drought periods (responding to increased radiation in the dry season) [56].

Microwave at lower frequency (L-band) has a larger crown transmissivity than C-band [34] and is more sensitive to soil and trunk moisture contents. Many field studies observed diurnal cycles in L-band backscatter [49, 50], and this cycle correlates well with changes in plant water content. Steele-Dunne *et al.* [34] also demonstrated this diurnal variation using MIMICS simulations.

1.3. ASCAT BACKSCATTER INCIDENCE-ANGLE DEPENDENCE

Since 2007, the Advanced Scatterometer (ASCAT), on board the Metop series of satellites operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), has been providing a long-term C-band backscatter data record. The unique design of the ASCAT instrument, which includes a set of three fan-beam antennas, allows for the nearly simultaneous measurement of backscatter (σ^o) at two distinct incidence angles [4]. Therefore, the incidence angle dependency of backscatter can be estimated from these backscatter triplets.

The incidence angle dependence means that, backscatter is modeled as a function with respect to the incidence angle [4] with the assumption that it is isotopic in the azimuth direction. The slope (σ') of the relationship between backscatter and

incidence angle contains valuable information about vegetation and soil conditions. Research about the ERS normalized backscatter and slope in the Iberian Peninsula [45] and Canadian Prairies [44] showed that the temporal variations of the slope are linked to the wet biomass of the vegetation, while the temporal variability of the normalized backscatter is mainly controlled by soil moisture. Slope and curvature were obtained based on the local slope data over several years, resulting in one value for each day of the year (DoY) [45, 57]. Therefore, only the climatology data of slope and curvature were available. In the TU Wien soil moisture retrieval algorithm [46], the climatology values of slope and curvature are used to normalize the backscatter to a reference angle and account for the vegetation effect. The TU Wien soil moisture retrieval algorithm was successfully applied on ERS-1/2 data [58] and ASCAT backscatter data [59, 60]. However, the interannual variability of vegetation cover effects was not considered until 2017. After 2013, a kernel smoother was developed to calculate the slope for the day d using all local slope values within a window with half-width λ centered at d [61]. Hahn *et al.* [4] tested and confirmed the robustness of the interannual estimation of the incidence angle dependence of backscatter. The availability of the time series of slope and curvature introduces more opportunities for understanding vegetation dynamics from microwave remote sensing observations, especially the backscatter incidence-angle dependence.

From the perspective of microwave remote sensing, the incidence angle dependence of backscatter depends on the relative dominance of surface, volume, and multiple scattering. Over bare soil, σ° decreases sharply as the incidence angle increases, indicating strong surface scattering. Changes in soil moisture lead to an increase in backscatter at all incidence angles. Increased volumetric scattering in the presence of vegetation reduces sensitivity to incidence angle, leading to a change in the slope and curvature. The degree to which volumetric scattering and multiple scattering between the vegetation and ground occur depends on vegetation structure, total water content, and the vertical distribution within the vegetation. A study in the North American grasslands [5] interpreted slope as a measure of vegetation density. Pfeil *et al.* [6] explored the sensitivity of ASCAT slope anomalies to the spring phenology of temperate deciduous broadleaf forests (DBF). They found a correspondence between the occurrence of DBF and a peak in the ASCAT slope time series between DOY 80 and 140. The observed shallow slope is explained as the maximum twigs and branches scattering, which occurs when the water content in the woody parts of the DBF is highest. Walraven [62] found that ASCAT slope anomalies correlated well with ISBA simulations and can be used to monitor drought in agricultural and grassland areas.

The curvature σ'' is affected by vegetation phenology [63–65]. Results from Steele-Dunne *et al.* [5] also suggested that in grasslands, curvature is a measure of the relative dominance of direct scattering from vertical vegetation constituents over a ground-bounce contribution, a phenomenon observed in earlier studies on wheat [63, 65, 66]. Stiles *et al.* [66] showed that mainly C-band VV backscatter gets a positive curvature while HH backscatter decreases monotonically with incidence angle, as reported in [63, 67, 68]. A more recent study by [7] showed that curvature may be linked to phenological changes such as leaf flushing and litterfall.

Therefore, ASCAT provides valuable information about vegetation water dynamics via normalized backscatter, slope, and curvature [4, 5]. ASCAT also offers a long data record with higher temporal frequency than leaf area index (LAI) which might help to better constrain the vegetation dynamics [69]. This thesis proposes to use the incidence dependency of ASCAT backscatter to monitor the variations in vegetation water content by combining the ASCAT observables with land surface models. Integrating the ASCAT observables into land surface models helps phenological and drought monitoring and improves our understanding of vegetation dynamics.

1.4. DATA ASSIMILATION OF ASCAT DATA

As summarized in Section 1.1, LSMs provide estimates of surface fluxes and land surface states of terrestrial water and carbon balance. However, the estimates of LSMs are susceptible to uncertainty in forcing data, parameter initialization, and land-atmosphere coupling parameterizations [70]. These uncertainties contribute to biases in temperature, humidity, and precipitation forecasts during periods of drought [13, 71]. Therefore, it is essential to use independent information from satellite observations to correct the errors in model simulations [72]. This approach is called data assimilation (DA). The resulting estimates of land surface states are calculated based on the errors and uncertainties of LSMs and satellite observations [70].

To improve the estimation of land surface variables (LSV) related to soil moisture and vegetation, there has been an increasing number of studies on assimilating remote sensing observations into LSMs [8, 73, 74]. Generally, land data assimilation systems constrain the LSMs with data products retrieved from satellites, such as surface soil moisture (SSM) retrieved from instruments like the Advanced Microwave Scanning Radiometer for Earth observation science (AMSR-E) [75–79], the Soil Moisture Ocean Salinity (SMOS) mission [80–83] and Soil Moisture Active Passive (SMAP) mission [84–87], as well as leaf area index (LAI) from Copernicus Global Land Service (CGLS) [74, 88] or station based observations [89, 90]. There are also studies exploring the direct assimilation of signals observed by satellites, such as radar backscatter (σ^o) observed by ASCAT [8, 44, 91] and Sentinel-1 [92–94] or passive temperature brightness (T_b) from SMOS [95, 96] or SMAP [92, 97–99].

Several studies have demonstrated the capability of ASCAT observations to accurately capture the temporal dynamics of in-situ and modeled soil moisture observations across Europe [100, 101]. Many studies have assimilated the ASCAT surface soil moisture product into LSMs to improve the estimates of surface soil moisture, root zone soil moisture, and other hydrological variables [87, 101–104]. Since the retrieved SSM products always have different climatologies compared to LSM simulations, cumulative density function (CDF)-matching and bias correction are used to diminish the difference between the climatology of SSM products and LSM states [105]. The assimilation of ASCAT SSM product was found to improve the estimates of surface and root-zone soil moisture over the continental U.S. [87] and Australia [106]. Mahfouf *et al.* [107] found that assimilating the ASCAT SSM product into a limited-area NWP model improves the forecast of relative humidity.

ASCAT SSM was then assimilated into the SIM hydrological model to improve the surface flux forecasts for drainage [102]. Additionally, ASCAT products have been jointly assimilated with LAI observations into LSMs to improve the estimates of soil moisture and vegetation biomass [69, 88, 108, 109]. Combining soil moisture and vegetation observations in the LSM provides more coherent information to reduce model uncertainties [109].

However, few studies concentrate on directly assimilating microwave observables (such as normalized backscatter and slope). Directly assimilating the microwave observations, such as σ^o and T_b, to constrain the states of LSMs [8, 110–113], obviates the need for CDF-matching and bias correction [114, 115] and also avoids potential cross-correlated errors between retrievals and model simulations [8, 95].

Lievens *et al.* [8] compared the assimilation of ASCAT normalized backscatter (σ^o_{40}) and SMOS T_b into the Global Land Evaporation Amsterdam Model (GLEAM). They calibrated the Water Cloud Model (WCM, [116]) and L-band Microwave Emission from the Biosphere radiative transfer model (L-MEB, [117]) to forward model ASCAT σ^o_{40} and SMOS T_b. The study demonstrated that assimilating ASCAT σ^o_{40} improves the model skills in surface soil moisture and land evaporation estimates. In passive microwave remote sensing, Lannoy *et al.* [95, 96] assimilated SMOS T_b into the GEOS-5 Catchment Catchment Land Surface Model with the $\tau - \omega$ model serving as the observation operator. The DA of SMOS T_b was compared to the DA of SMOS retrieval products. The assimilation of T_b showed larger innovations and local differences in performance, which were attributed to differences in how the T_b 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 [98, 118]. The diagnosis of the DA system was fully examined on a global scale to assess the impact of assimilating SMAP T_b. In the snow assimilation community, Forman *et al.* [119] and Xue *et al.* [120] trained a machine learning model to extract information from microwave observables to constrain the LSM. Then AMSR-E T_b was assimilated into the Catchment model to improve snow state estimation.

However, there are few studies directly assimilating ASCAT observables into LSMs. ASCAT observables are shown to contain valuable information about soil moisture [44], vegetation water content [5], and vegetation phenology [6]. The assimilation of ASCAT observables enables simultaneous updates of multiple land surface states (e.g., soil moisture and vegetation) more effectively than DA of the various associated individual retrieval products with cross-correlated errors [72].

The premise of joint assimilation of ASCAT normalized backscatter and slope is to simultaneously constrain the soil and vegetation dynamics in the LSM. Previous work achieved this by assimilating retrieved SSM and LAI products [69, 88, 108, 109, 121–123]. They showed that assimilating the vegetation products together with soil moisture observations improves the estimate of land surface variables like evapotranspiration, SSM, LAI, and CO2 fluxes [108, 109, 121], hydrological variables [114], and other variables [122]. This can lead to an improved ability to monitor and forecast drought [69, 109], and constrain the phenological cycle in the land surface model [108]. Analogous to VOD, dynamics in ASCAT slope have been shown

to reflect changes in biomass [5], phenology [6], and water status [7, 124]. One advantage of using ASCAT slope rather than VOD is that it is directly calculated from observations, thereby eliminating the need for assumptions or ancillary data. Several studies have demonstrated the assimilation of VOD along with SSM and other observations [125, 126] improves the estimation of various land surface variables such as evapotranspiration [126], soil moisture and GPP [125]. Therefore, joint assimilation ASCAT normalized backscatter and slope should be expected to improve the estimation of land surface (soil and vegetation) states.

1.5. MACHINE-LEARNING-BASED OBSERVATION OPERATORS

Directly assimilating microwave observables into LSMs requires a forward model to relate the land surface state (soil moisture and leaf area index) to the satellite observables [72]. Radiative transfer models (RTMs) are commonly used as observation operators to connect simulated states with microwave observations. The water cloud model (WCM) and $\tau-\omega$ model are the most frequently used observation operators for the assimilation of backscatter and brightness temperature.

The WCM was used as the observation operator to reconcile the information in ASCAT normalized backscatter with the states in the LSM Global Land Evaporation Amsterdam Model (GLEAM) by [8] and the hydrological model SCHEME [91]. Shamambo *et al.* [127] also proposed using WCM as an observation operator linking ISBA states to ASCAT normalized backscatter. Modanesi *et al.* [128] calibrated the WCM with open loop simulations of SSM and LAI from LSM Noah-MP. Then Sentinel-1 backscatter was assimilated into Noah-MP to improve the SSM and LAI estimates [94]. However, the improvement of the estimates of irrigation amounts is limited by the irrigation parameterizations [94].

To directly assimilate ASCAT normalized backscatter, slope, and curvature into an LSM, an observation operator is needed to predict these observables based on the outputs of the land surface model. However, the implicit assumption of WCM that water is uniformly distributed within the vegetation conflicts with the objective of understanding water transport through the canopy. Other sophisticated radiative transfer modeling approaches [129–131] implement double bounce terms that are neglected in WCM. However, they 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. In this study, machine learning will be employed to generate a surrogate model to circumvent the challenges and limitations of current radiative transfer modeling approaches.

The number and diversity of applications of machine learning have increased considerably in recent years [132]. In remote sensing, neural network (NN) approaches have proved valuable to "emulate" or reproduce the outputs of (e.g.) Monte Carlo radiative transfer code in a computationally efficient and fast way [133–136]. Similarly, they have been applied to emulate retrieval models to provide

products for assimilation [137]. Aires *et al.* [104] recently investigated the merits of applying an NN for retrieval versus forward modeling to assimilate ASCAT σ^o_{40} into an LSM. Corchia *et al.* [138] also assimilated ASCAT normalized backscatter into ISBA by using a neural network as the observation operator. Their observation operator is trained on surface soil moisture and surface soil temperature from ISBA, and LAI observations from the PROBA-V satellite. Their results show an overall improvement compared to the open-loop simulation, evaluated against PROBA-V LAI observations and in-situ soil moisture observations. In studies focused on snow processes, Forman *et al.* [119], Xue *et al.* [120], and Forman *et al.* [139] directly linked LSM states to passive microwave brightness temperatures by training a machine learning model to extract information from microwave observables to constrain the LSM.

In this thesis, we will follow the approach of previous studies [119, 120, 139, 140], where machine learning is used to provide an observation operator for the direct assimilation of microwave observables to constrain a land surface model. The innovation of the machine-learning-based observation operator is that the relationship between the LSM and the observables is indirect. Here, the ASCAT observables are known to depend on vegetation water content, which is not directly modeled in ISBA. Therefore, we hypothesize that dynamics in vegetation water content can be captured through the inclusion of ISBA land surface variables related to vegetation water transport processes as inputs to the observation operator.

1.6. RESEARCH OUTLINE

The goal of this thesis is to develop a machine-learning-based observation operator and use it to assimilate ASCAT normalized backscatter and slope into the land surface model ISBA-A-g_s.

Chapter 2 summarizes the theoretical overview of the land surface model ISBA and ASCAT observables. More details about ISBA and ASCAT observables are introduced in this chapter.

Chapter 3 tests the hypothesis that the machine learning algorithm can capture the dynamics in vegetation water content by including ISBA land surface variables related to vegetation water transport processes as inputs to the measurement operator. A DNN is trained to directly link the ASCAT dynamic vegetation parameters to the states from ISBA. The performances of the DNN are evaluated to assess the ability of the DNN to reproduce the dynamics of ASCAT observables. Then the normalized sensitivity coefficient is analyzed to ensure whether the DNN is physically plausible.

Chapter 4 tests the hypothesis that the joint assimilation of ASCAT normalized backscatter and slope improves the estimates of surface and root zone soil moisture as well as vegetation states such as LAI if given an accurate and physically plausible observation operator, along with well-defined model and observation errors. The estimates of the data assimilation and open loop runs are evaluated against the ISMN in-situ soil moisture measurements and independent LAI observations from CGLS. The data assimilation diagnostics are analyzed to test whether the model and observation errors are reasonable.

Chapter 5 runs a synthetic data assimilation experiment. ISBA is perturbed to

create synthetic truth of WG2 and LAI, as well as synthetic backscatter. Then Chapter 5 tests the hypothesis that assimilating synthetic backscatter using DNNs as the observation operator leads to suboptimal performances because a) DNN estimations are trained using open loop data (which is biased compared to synthetic truth) as input; b) the presence of redundant input variables or high collinearity makes inaccurate estimation of DNN-based Jacobians. The DNN is compared to the water cloud model in terms of estimation skills, Jacobians, and subsequent DA performances.

Chapter 6 synthesizes the key innovations and contributions of this thesis in constraining the LSM by assimilating ASCAT observables using a machine-learning-based observation operator. The findings are discussed in a broader view, e.g. the usage of the machine learning algorithms in the data assimilation community, linking the ASCAT observables to the carbon cycle, assimilating multi-source earth observation data, and using the ASCAT observations to monitor drought.

2

Background

2.1. LAND SURFACE MODEL ISBA-A-GS

In this thesis, the ISBA (Interactions between Soil, Biosphere and Atmosphere [141, 142]) land surface model is used to simulate water, energy, and carbon exchanges at the land surface. ISBA is a soil-vegetation-atmosphere transfer (SVAT) scheme developed by Meteo-France [141–143]. It uses a force-restore method to calculate soil water and heat dynamics. Land surface variables (surface temperature, surface soil volumetric moisture, and the canopy interception reservoir) are calculated using prognostic equations [144]. Land cover information is incorporated into ISBA using the “tile" or “patch" approach [145]. Within each grid point (GPI), ISBA simulates the evolution of prognostic variables for each land cover type. Surface parameters like soil texture, albedo, emissivity, surface roughness, and land cover fractions are provided by the ECOCLIMAP dataset. The ISBA parameters are defined for 19 generic land surface patches as described in Table 2.1.

The initial version of ISBA simulates the soil water and heat states in three layers [143]. The soil hydrology is represented by the surface layer, the rooting layer, and the deep layer [146], using the force-restore method [147]. Soil moisture simulations from ISBA show good agreement with ground observations at the SMOSREX station [90]. Sabater *et al.* [89] found that ISBA reproduced the root zone soil moisture over SMOSREX site from 2001 to 2004 with a Nash efficiency score of 0.79. Sabater *et al.* [89] also found that the daily average daily standard deviation of SSM is 0.03 m^3m^{-3}. Albergel *et al.* [148] evaluated the surface soil moisture from ISBA simulations in Southwestern France against ASCAT data and in-situ measurements from SMOSMANIA and SMOSREX sites (depth at 5 m). The RMSE between ISBA simulations and in-situ data ranges from 0.053 to 0.174 m^3m^{-3}. ASCAT data showed a stronger correlation with ISBA simulations, possibly because the surface layer in ISBA is more consistent with the sensing depth by ASCAT at C-band (0-0.02 m) than with the in-situ observations at a depth of 0.05 m [148]. Additionally, the LSM and forcing data may better represent the area observed in an ASCAT pixel compared to local observations [148].

Subsequently, the multi-layer diffusion scheme was developed to explicitly simulate the time evolution of soil moisture and temperature in multi-layers [145, 146]. In the multi-layer diffusion scheme, ISBA simulates soil moisture in 14 layers, with depths of 0.01 m, 0.04 m, 0.1 m, 0.2 m, 0.4 m, 0.6 m, 0.8 m, 1.0 m, 1.5 m, 2.0 m, 3.0 m, 5.0 m, 8.0 m, 12.0 m [88, 115]. Evaluations showed that the multi-layer diffusion scheme can reproduce the evolution of soil moisture dynamics and performs better in estimating surface energy fluxes when heterogeneous soil properties are considered [146]. Therefore, in this thesis, the multi-layer diffusion scheme of ISBA is chosen to simulate the exchanges of water and heat fluxes in the soil [145, 149].

The CO2-responsive version of ISBA (ISBA-A-g_s, here referred to as ISBA) is built within the SURFEX (Surface Externalisée, in French) platform (version 8.1) [69, 88, 151]. With the “NIT” plant biomass monitoring option, ISBA simulates plant physiological states and plant growth [22, 143, 152]. ISBA models the dynamics of photosynthetic processes by simulating stomatal conductance. Calvet *et al.* [143] replaced Jarvis’s parameterization of stomatal conductance [153] with

Table 2.1.: Description of land cover types in study domain. The reader is referred to Boone *et al.* [150] for a more complete description of the classes.

Number	Class	Abbreviation
1	no vegetation (smooth)	NO
2	no vegetation (rocks)	ROCK
3	permanent snow and ice	SNOW
4	temperate broadleaf cold-deciduous summergreen	TEBD (TREE)
5	boreal needleleaf evergreen	BONE (CONI)
6	tropical broadleaf evergreen	EVER
7	C3 cultures types	C3
8	C4 cultures types	C4
9	irrigated crops	IRR
10	grassland (C3)	GRAS
11	tropical grassland (C4)	TROG
12	peat bogs, parks and gardens (irrigated grass)	PARK
13	tropical broadleaf deciduous	TRBD (TREE)
14	temperate broadleaf evergreen	TEBE (TREE)
15	temperate needleleaf evergreen	TENE (CONI)
16	boreal broadleaf cold-deciduous summergreen	BOBD (TREE)
17	boreal needleleaf cold-deciduous summergreen	BOND (CONI)
18	boreal grass	BOGR (GRASS)
19	shrub	SHRB (TREE)

Jacob's formulation [16], i.e. A-g_s model. Jacob's biochemical A-g_s model simulates stomatal conductance by assuming a linear relationship between stomatal activities and photosynthetic carbon assimilation under well-watered conditions. Under water stress conditions, ISBA considers two different types of drought responses of plants: drought-tolerant and drought-avoiding strategies. The key difference is related to water use efficiency (WUE). In the drought-tolerant strategy, WUE remains constant or decreases in the early drought stage but increases under the drought-avoiding strategy.

The stomatal conductance is simulated as a function of the concentration of CO2 at the leaf surface, leaf temperature, saturation deficit at the leaf surface, global radiation, and plant type (C3 or C4). The leaf carbon assimilation is then given by the A-g_s model. ISBA also simulates growth and senescence processes to predict the evolution of the LAI. The bulk biomass (B) is controlled by the integrated canopy net assimilation (obtained from the atmospheric CO2) and mortality (resulting from a deficit of photosynthesis due to external factors) [143]. Then the LAI is given by a linear relationship between the LAI and B:

$$B_\alpha = \alpha \times LAI \tag{2.1}$$

where the α depends on vegetation type, nitrogen supply, and climate [108]. A minimum LAI value is set (0.3 m^2m^{-2} for the herbaceous PFTs and the broadleaf forests, 1 m^2m^{-2} for coniferous forests) to allow for the start of the growing season

[154]. Gibelin *et al.* [155] demonstrated that ISBA provides realistic LAI time series on a global scale. However, ISBA simulates a late start of the growing season due to limited phenology parameterizations. Calvet *et al.* [156] showed that ISBA obtained good simulations of interannual variability in grasslands. However, it is difficult for ISBA to represent the interannual variability of crops due to the lack of a specific crop representation in ISBA. ISBA was also compared to ORCHIDEE regarding LAI simulations [154]. Results showed that ISBA better represents the interannual variability especially for grasslands because of the weak constraints on the phonological cycle. However, ISBA tends to simulate a delayed seasonal cycle. This might be due to the fact that the leaf onset of ISBA is more sensitive to errors in model physics and atmospheric forcing [154, 157].

Studies have been conducted to assimilate station-based or remote-sensed observations into ISBA to improve simulations of SSM and LAI. Sabater *et al.* [89] assimilated SMOSREX observations of SSM and LAI into an offline version of ISBA. The assimilation improves the simulations of root zone soil moisture and vegetation biomass during growing and senescence seasons. Barbu *et al.* [108] found similar improvements at the SMOSREX grassland site. Regarding remote sensing observations, Albergel *et al.* [88] and Draper *et al.* [102] found that assimilating ASCAT soil moisture data and GEOV1 LAI observations benefits the carbon and water balances, improving GPP and ET estimates. Assimilation also reduced the one-month delay in the LAI seasonal cycle. Albergel *et al.* [69] found that assimilating LAI observations provides better initial conditions for early warning systems of agricultural drought and yield forecasts, with impacts lasting for several weeks or even months.

However, the assimilation of LAI is limited by the low availability of LAI products (every 10 days at best). Therefore, alternative microwave remote sensing observations with higher temporal frequency can be assimilated into ISBA to improve efficiency [69]. Many studies assimilated vegetation optical depth (VOD) into LSMs to constrain the vegetation dynamics [125, 126]. However, retrieved products use the land surface parameters and auxiliary information (like vegetation, texture, and temperature) which might be inconsistent with model simulations [69, 88]. If the same information is used in retrieval and model simulations, their errors may be cross-correlated [95]. This leads to the exploration of new microwave remote sensing observables that could better constrain land surface models.

2.2. ASCAT DYNAMIC VEGETATION PARAMETERS

The Advanced Scatterometer (ASCAT) on board the series of Metop satellites is a real aperture radar operating at 5.255 GHz (C-band). Metop-A was first launched in October 2006, followed by Metop-B in September 2012, and the third in November 2018. These satellites, operated by the EUMETSAT, are in a sun-synchronous orbit at 817 kilometers with an equator crossing Local Solar Time (LST) of 09:30 A.M. and P.M. in descending and ascending nodes, respectively [4].

ASCAT was primarily designed to observe the wind patterns at the ocean surface. However, its utility extends far beyond its initial design. The incidence

angle dependency of ASCAT backscatter contributes significantly to land surface observations, particularly in assessing soil moisture and vegetation properties.

ASCAT observes backscatter using two sets of three side-looking antennas, each illuminating a 550 km wide swath on either side of the satellite track. On each side, the antennas are oriented at 45° (fore), 90° (mid), and 135° (aft) to the satellite track, as shown in Fig 2.1. The incidence angles are 34–65° and 25–55° for fore-/aft- and mid-antennas, respectively. The measurements at multiple incidence angles allow ASCAT to effectively capture the incidence angle dependency of backscatter.

C-band backscatter as a function of incidence angle can be approximated by a second-order Taylor expansion [4, 46, 61].

$$\sigma^o(\theta) = \sigma^o(\theta_r) + \sigma'(\theta_r)(\theta - \theta_r) + \frac{1}{2}\sigma''(\theta_r)(\theta - \theta_r)^2, \qquad \text{[dB]} \tag{2.2}$$

where the zeroth order term $\sigma^o(\theta_r)$ is the normalised backscatter at reference angle ($\theta_r = 40°$), and the 1st and 2nd order coefficient $\sigma'(\theta_r)$ and $\sigma''(\theta_r)$ are referred to as slope and curvature [4]. The measured backscatter at any incidence angle can be extrapolated to the θ_r as follows given the values of $\sigma'(\theta_r)$ and $\sigma''(\theta_r)$:

$$\sigma^o(\theta_r) = \sigma^o(\theta) - \sigma'(\theta_r)(\theta - \theta_r) - \frac{1}{2}\sigma''(\theta_r)(\theta - \theta_r)^2. \tag{2.3}$$

Backscatter at 40 degrees is obtained using this equation and referred to as "normalized backscatter" or σ^o_{40}.

The slope and curvature values are estimated from backscatter triplets measured by Metop ASCAT. Backscatter triplets [σ_f, σ_{mid}, σ_a] (fore-, mid-, and aft-beam) at the incidence angles [θ_f, θ_{mid}, θ_a] are used to produce two local slope estimates:

$$\sigma'\left(\frac{\theta_{mid} - \theta_{a/f}}{2}\right) = \frac{\sigma^o_{mid}(\theta_{mid}) - \sigma^o_{a/f}(\theta_{a/f})}{\theta_{mid} - \theta_{a/f}}, \qquad \text{[dB/deg]} \tag{2.4}$$

A large number of local slope values, distributed over the entire incidence angle range [4], are combined to estimate the slope and curvature [4]. Following Melzer [61], the slope and curvature parameters for the day d are computed using all local slope values within a window with half-width λ centered at d. The value of λ is chosen to balance the bias and variance of estimates. Hahn *et al.* [4] tested the performance and robustness of the kernel smoother (KS) globally, by comparing the results obtained from ASCAT on-board Metop-A and Metop-B independently. Overall the results from Metop-A and Metop-B are in good agreement, confirming a robust interannual estimation of the incidence angle dependence of backscatter using KS.

In vegetated areas, the slope mainly depends on the relative contributions of the vegetation canopy (volume scattering) and the underlying soil surface (surface scattering). The volume scattering of vegetation canopy is nearly uniform over the range of incidence angles, while the backscatter from bare soil decreases rapidly with an increase in the incidence angle [45]. Therefore, in summer, there is more volume scattering from the vegetation canopy compared to surface scattering from the underlying soil due to the decrease in vegetation transitivity. As a result, the

backscatter decreases less rapidly with an increase in the incidence angle, making the slope in equation 2.2 larger. The transitivity of vegetation cover is closely related to the product of the volumetric water content of the vegetation and the plant height [116], linking the variation of slope to the wet biomass of the vegetation.

A study in the Iberian Peninsula [45] confirmed that the slope value from ERS data is smallest in winter (December and January) and largest in summer (July). The summer peak of the slope is slightly delayed compared to the NDVI time series. This may be because the variations of slope indicate the wet biomass (peaks in summer) while NDVI indicates the photosynthetic activity of the vegetation (peaks in spring) [45]. Scipal [158] found that the spatial distribution of the slope maximum generally aligns with climatic conditions. However, in some regions like Africa, the slope peaks in winter. More information about surface roughness and ice formation is needed to quantify such anomalies. Steele-Dunne *et al.* [5] found that ASCAT slope measures the vegetation density. The value of the slope is smaller during the descending pass (9:30 am) than during the ascending pass (9:30 pm), and this diurnal difference is largest in summer. This was attributed to the high transpiration rate in the early morning which reduces plant water content. Pfeil *et al.* [6] identified a spring peak in slope time series in a deciduous broadleaf forest in Austria. The shallowest slope is observed when the plant water content in twigs and branches reaches maximum values. This coincides with the time of the first leaf unfold which enhances the physiological activities.

The curvature (σ'') is thought to be affected by the vertical structure of the vegetation. Studies observed positive curvature in grasslands and wheat fields [5, 66]. At lower incidence angles, scattering is dominated by the ground-bounced term [66, 159], while the direct bounce term is smaller. At higher incidence angles, vertically polarized microwaves interact more with the vertical structure of the wheat plant. As the incidence angle increases, the ground bounce term decreases, and direct scattering from the grain and upper portion of the stalk increases. Smaller curvature in the evening pass was observed [5] which suggests the ground bounce term is more dominant in the evening.

Slope and curvature also contain information about drought. Steele-Dunne *et al.* [5] observed the negative slope anomalies and positive curvature anomalies in the 2011 and 2012 drought events over the north American grasslands. Walraven [62] showed that ASCAT observables correlate well with ISBA simulations during droughts. A study in Australia [160] showed that ASCAT slope and curvature reveal information about wildfires. Negative slope anomalies were observed after fires due to the loss of scatterers like leaves and branches. However, positive curvature anomalies suggest that large vertical tree trunks remain while leaves and branches combust. Therefore, microwaves interact more with the vertical structure of the forests.

These studies confirm the valuable information contained in ASCAT normalized backscatter, slope, and curvature. This information benefits the monitoring of wildfire [160] and drought [5, 62], thus offering promising opportunities for land surface modeling and drought predictions.

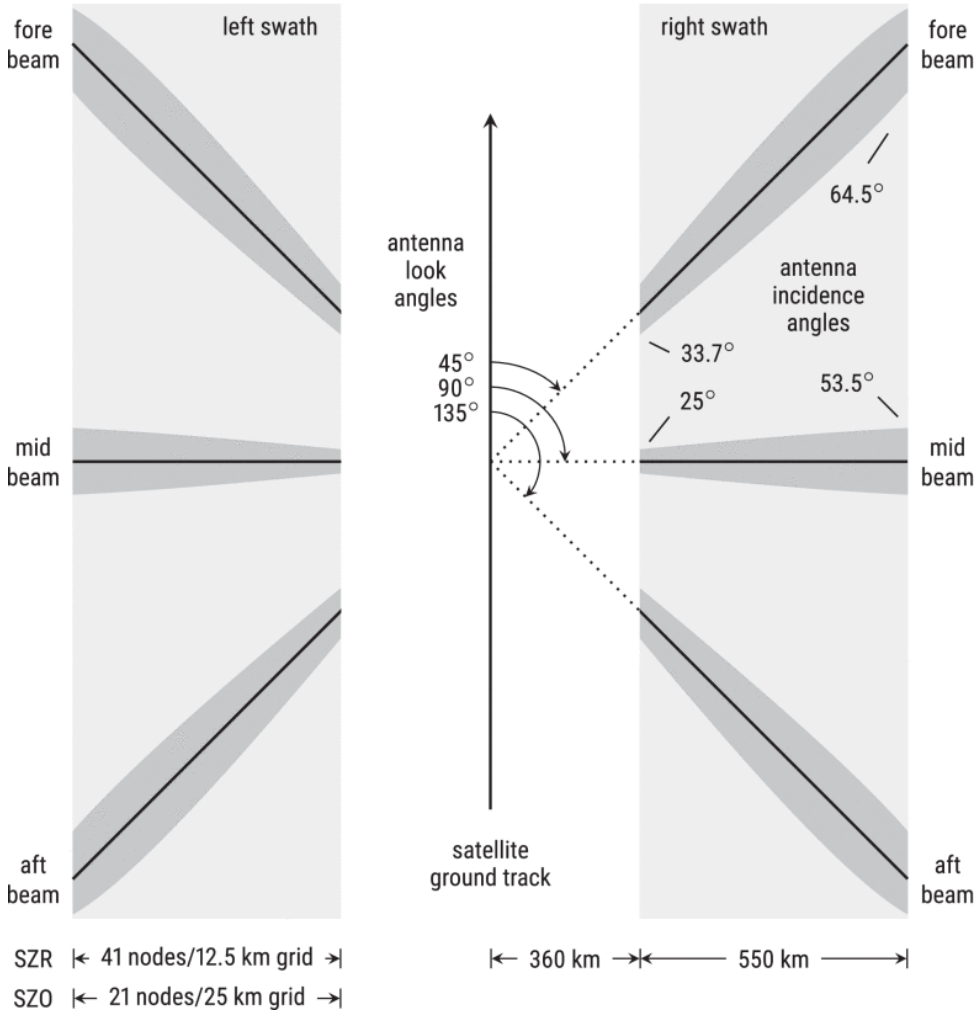

Figure 2.1.: ASCAT swath geometry. The dimensions are symmetric with respect to the satellite ground track. Source is adapted from Hahn *et al.* [4] with permission.

3

MODELING ASCAT BACKSCATTER AND SLOPE WITH A DEEP NEURAL NETWORK

Based on X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. "Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a Deep Neural Network". In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. DOI: 10.1016/j.rse.2022.113116. URL: https://doi.org/10.1016/j.rse.2022.113116

3.1. INTRODUCTION

The previous chapter summarized the importance of directly assimilating ASCAT normalized backscatter, slope, and curvature. To assimilate normalized backscatter, slope, and curvature directly into an LSM, an observation operator is needed to predict these observables based on the outputs of the land surface model. In this chapter, a Deep Neural Network (DNN) is used to directly relate the land surface variables from the ISBA Land Surface Model to the ASCAT backscatter, slope, and curvature. The objectives are two-fold. First, this DNN model could be used as an observation operator to directly assimilate these ASCAT observables into ISBA. Second, the sensitivity of the observables to the ISBA land surface variables can improve our understanding of, and ability to operationally exploit the ASCAT backscatter, slope and curvature.

3.2. METHOD AND MATERIALS

3.2.1. STUDY AREA

The study domain extends from 41.625°N to 50.875N°, and from 4.625°W to 9.375°E as shown in Figure 3.1. The ASCAT data and land surface variables (LSVs) are available almost daily from 2007 to 2019 with a spatial resolution of 0.25° x 0.25°. The study domain comprises 1020 grid points (GPIs).

The albedo and LAI values required by the ISBA model are obtained from the ECOCLIMAP II database, as well as vegetation fractions of 12 plant functional types and other land cover types at each GPI [161]. Data from this global surface parameter database are provided at 1 km resolution. The land cover type assigned to individual grid points is the dominant land cover type within a 25km x 25km window centered on the ASCAT GPI (Figure 3.1 (a)). The corresponding maximum value of the land cover fraction in each GPI is shown in Figure 3.1 (b). Clearly all grid points comprise a mixture of different land cover types. Four representative regions of interest (ROIs) are identified with a comparatively high fraction of the dominant land cover (Figure 3.1 (c)).

Figure A.1 shows the fraction of vegetation cover types of the GPIs in four regions of interest. This mixed composition is accounted for in ISBA as each grid cell is composed of patches of the different cover types.

3.2.2. ASCAT BACKSCATTER-INCIDENCE ANGLE DEPENDENCE

Figure 3.2 summarizes the dynamics in ASCAT normalized backscatter, slope, and curvature during the study period. The interannual variability and seasonal variability are estimated by assuming that the time series is the additive combination of a seasonal cycle (period=365 days), trend, and residual which is decomposed using moving averages [162]. The interannual variability (IAV) and seasonal variability (SV) indicate the standard deviation of the trend and seasonal cycle series respectively. Spatial variability in normalized backscatter, slope, and curvature reflects patterns in dominant land cover. Mean values range from -10.75 to -7.805 dB, from -0.13 to -0.075 dB/deg and from -0.0027 to 0.00050 dB/deg^2 for σ^o_{40}, σ' and σ'', respectively.

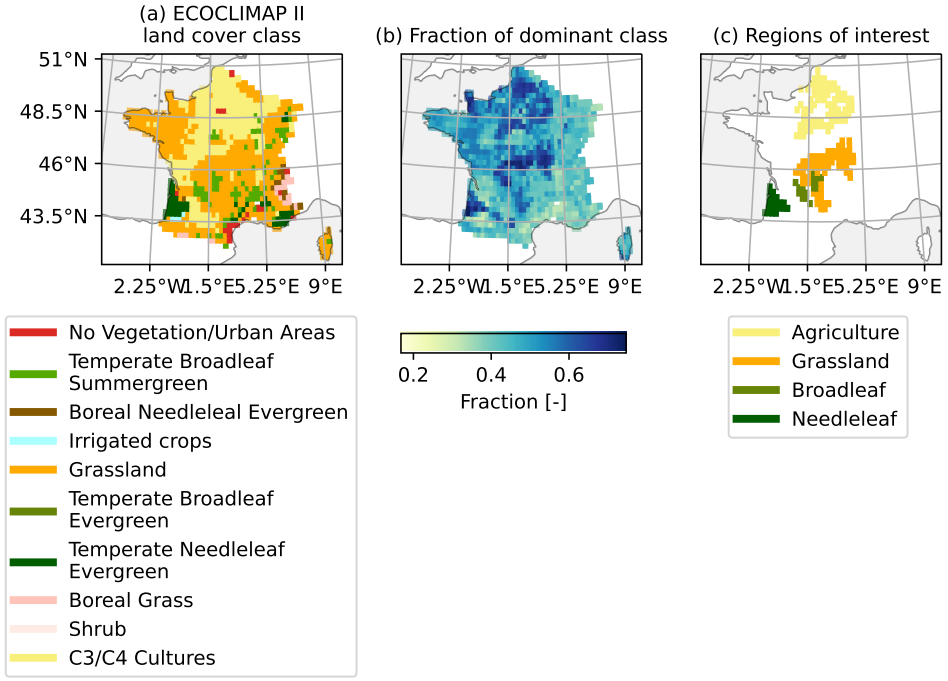

Figure 3.1.: The main land cover types and fraction of France based on ECOCLIMAP II. (a) the dominant vegetation types of France; (b) the vegetation fraction of the dominant vegetation types; (c) 4 regions of interest (ROIs) representing 4 vegetation types: Broadleaf (20 GPIs in lon 3-0, lat 44-46.5 with the fraction of temperate broadleaf cold-deciduous summergreen (TEBD) ≥ 0.35), Agriculture (98 GPIs in lon 0.2-5, lat 47.5-51 with the fraction of C3 cultures (C3) ≥ 0.5), Grassland (76 GPIs in lon 0.4-5, lat 43-48 with the fraction of grass (GRAS) ≥ 0.5) and Needleleaf forests (21 GPIs with the fraction of temperate needleleaf evergreen (TENE) ≥ 0.4).

Aside from urban areas, the highest normalized backscatter values are observed (-7.80 dB) in northern France where the main land cover type is Agricultural C3 crops, and lowest (-10.75 dB) in central France which is dominated by grassland land cover. The highest mean value of σ' (-0.075 dB/deg) is found in grassland, while the lowest value (-0.13 dB/deg) is found in northern France as shown in Figure 3.2 (b). Agricultural crop in northern France has the smallest mean value of σ'. The IAV of normalized backscatter, slope, and curvature range from 0.042 to 0.29 dB, from 0.0013 to 0.0078 dB/deg, and from 7.5e-5 to 0.00054 dB/deg^2, respectively. The corresponding domain-averaged values of IAV are 0.15 dB, 0.0033 dB/deg and 0.00019 dB/deg^2. The IAV of normalized backscatter is largest in northern agricultural land and is generally higher in the north and west parts of France, as

shown in Figure 3.2 (g). Meanwhile, in Figure 3.2 (h - i), it is observed that IAV of σ' and σ'' are the largest in the Les Landes forest area. This is likely due to the impact of Cyclone Klaus in January 2009, after which the forest degraded from 2009 to 2012 and started regeneration during 2013-2017 [127, 163]. As for the seasonal cycle, SV of normalized backscatter, slope and curvature are generally largest in agricultural areas, with values up to 4.2 dB, 0.11 dB/deg and 0.0043 dB/deg^2 for σ^o_{40}, σ' and σ'', respectively. SV is specifically largest in the areas of intensive agricultural land surrounding Paris. The spatial pattern of SV is generally similar to the spatial pattern of range values. The uncertainty of ASCAT σ^o_{40}, σ' and σ'' varies with land cover but is overall small compared to the dynamic range. Uncertainties in σ^o_{40} are highest in agricultural areas in northern France, mountainous areas such as the Alps, and coastal areas.

3.2.3. INTERACTIONS BETWEEN SOIL, BIOSPHERE AND ATMOSPHERE (ISBA) MODEL

Here, the ISBA land surface model [141, 142] is used to simulate water, energy, and carbon exchanges at the land surface. With the "NIT" plant biomass monitoring option, the CO2-responsive version of ISBA simulates plant physiological states and plant growth [22, 143, 152]. Exchanges of water and heat fluxes in the soil are simulated based on a multi-layer diffusion scheme [145, 149].

Parameters of the photosynthesis model and plant growth model are patch-dependent [154]. LSV values for the GPI are weighted combinations of the values obtained for each patch. The nominal values of LSVs from an open loop run of ISBA are used as input for DNN. The model was forced by the latest ERA-5 atmospheric reanalysis from ECWMF [164] from 1996 to 2019. The meteorological forcing data are available on a 0.25° x 0.25° grid, and include rainfall rate, 2m air temperature, 2m specific humidity, wind speed, wind direction, surface pressure, downward direct shortwave radiation, downward diffuse radiation, downward longwave radiation, snowfall rate, and CO2 concentration. All ERA-5 atmospheric variables were interpolated using bi-linear interpolation to match the grid points of ISBA. The model was initiated by spinning up with 20 repetitions using the 1996 forcing data. The open-loop simulation was obtained using the ERA5 forcing data from 1997 to 2019.

3.2.4. DEEP NEURAL NETWORK

Here, a DNN is used to relate the ISBA LSVs to the ASCAT normalized backscatter, slope, and curvature. DNNs are based on artificial neural networks (ANN), and are increasingly used for machine learning applications involving large datasets with high dimensionality [132, 165] including in the geosciences (e.g. [166–168]). A DNN contains an input layer, multiple hidden layers, and an output layer. Each layer contains multiple neurons connected to neurons in adjacent layers using the corresponding weight and bias matrix [169]. During back-propagation, the weights and biases are updated by a stochastic optimization algorithm, Adam [170], that minimizes (maximizes) the loss (accuracy) between the DNN outputs and the

training targets. In order to prevent over-fitting, early-stopping was used during the training process. DNNs contain more parameters (e.g. number of layers and nodes) than other machine learning algorithms, so they have the flexibility to capture complex features from high dimensional systems [171, 172].

One DNN is trained per GPI to obtain normalized backscatter, slope and curvature simultaneously. The structure of DNN is tuned by Bayesian optimization [173]. The parameters to be optimized include the number of layers, the number of neurons of each layer, batch size, the choice of activation function, and learning rate. The optimal parameter combination is that which minimizes RMSE between observed and estimated σ^o_{40}, $\sigma^{'}$ and $\sigma^{''}$ simultaneously during the testing period.

Table A.1 lists the ISBA LSVs that are used as inputs to the DNN. These are selected to account for the influence of soil and vegetation on the ASCAT observables. It is widely understood that surface soil moisture has a significant influence on C-band backscatter [44, 46], therefore WG2 is included. Note that WG2 (5-10cm) rather than WG1 (0-5cm) is used as surface soil moisture following previous studies [88, 102, 109]. The sensing depth of C-band backscatter is generally assumed to be on the order of 2 cm in non-arid environments, therefore there is unlikely to be a *direct* relationship between backscatter, slope or curvature, and root zone soil moisture. However, root zone soil moisture terms (WG3 - WG10) are included as inputs, due to the relationship between root zone soil moisture, plant hydraulics, and plant water content [30, 38, 174–180]. The vegetation interception reservoir water storage (WR) is included to allow for the fact that intercepted water (from dew/irrigation) can have an effect on backscatter, and hence slope and curvature. This has recently been shown for L-band backscatter [181, 182], but it is reasonable to suspect that it also has an effect on ASCAT DVP in C-band [183]. Plant transpiration (LETR) and stomatal conductance (XRS) are included as they are indicators of the transport of moisture from the root zone to the atmosphere and therefore reflect dynamics in plant water content. Leaf Area Index (LAI) and Gross primary production (GPP) are included to account for canopy growth and biomass accumulation as both backscatter, and its incidence angle dependence are influenced by seasonal changes in both biomass itself and the associated changes in canopy architecture. Net radiation (RN) is included to account for the influence of radiation availability on phenology, particularly structural changes in the vegetation (e.g. leaf/bud formation, leaf loss, etc.)[7], and changes in constituent water content associated with seasonality [6]. The LSVs are obtained from the open loop simulations of the ISBA model. The reader is referred to the studies of Albergel *et al.* [88], Draper *et al.* [102], and Barbu *et al.* [108] for information regarding ISBA model error. Note, however, that the nominal open loop values are used without any consideration of model error.

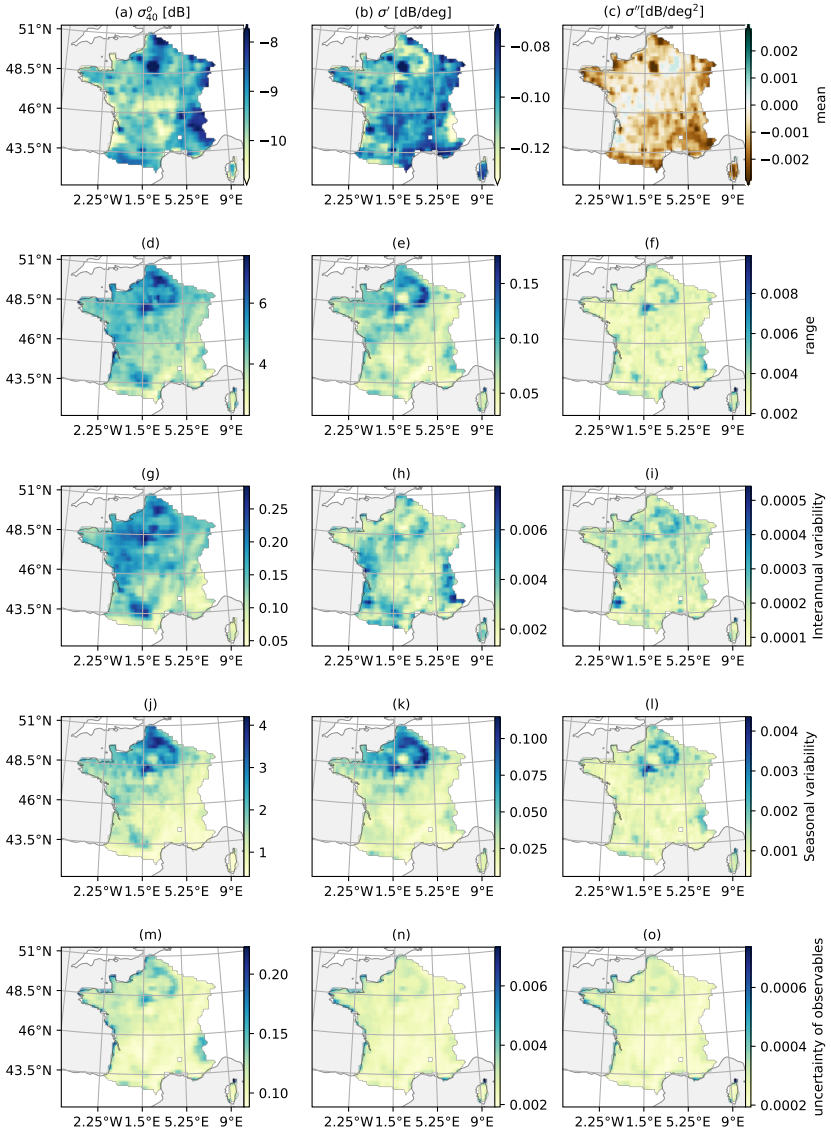

Figure 3.2.: The mean values, range values, and uncertainty of ASCAT backscatter and DVP from 2007 to 2019. Different columns show backscatter, slope, and curvature, respectively. Mean values (a - c), range values (d - f), interannual variability (IAV, g - i), seasonal variability (j - l) and uncertainty (m-o) are shown for the study domain. The mean values greater than 98% and smaller than 2% have been masked as the 98% and 2% values respectively.

The training targets of the DNN are the ASCAT normalized backscatter, slope, and curvature. Rather than training one generic DNN, or one DNN per land cover type, a DNN is generated for each individual grid point. This is motivated by the coarse resolution of the ASCAT observations which means that the ASCAT footprint always contains a combination of several land cover types (Figure 3.1 (b) and Figure A.1). This heterogeneity is highly relevant because (1) the relationship between ISBA LSVs and the quantities affecting microwave backscatter varies per vegetation type (2) microwave interactions with vegetation are different per vegetation type and (3) one needs to consider how contributions from different surface types are "mixed" to yield the observables at footprint scale. Note that the proposed DNN is not simply emulating an RTM. To use most RTMs as a measurement operator, one would need to translate the outputs of ISBA (the LSVs) to the required inputs of the RTM (e.g. roughness, soil moisture, vegetation water contents, vegetation parameters), run an RTM for each of the vegetation types, and then apply a mixing model to obtain the observables at the ASCAT scale. This is particularly challenging as the required inputs to the RTM are generally not simulated by most land surface models (e.g. VWC). The proposed DNN circumvents these challenges by mapping the ISBA LSVs directly to the ASCAT observables at the scale of interest.

The data record was split into a training-testing dataset (2007 - 2016) and an independent validation dataset (2017 - 2019). Before the training process, the inputs and training targets of the DNN are normalized as follows to account for the difference in magnitudes among the variables:

$$x_{i,\text{norm}} = \frac{x_i - \text{mean}(x_i)}{\text{std}(x_i)}, \tag{3.1}$$

where x_i is the i-th input variable or labels and i = 1,...,m for input LSVs or i=1,2,3 for ASCAT data. During the cross-validation process, a "jack-knife" approach was employed where training cycled through the 13-year period excluding the current year from the training-testing set [184, 185]. The best model was chosen as that which minimized RMSE among those submodels according to their performances on the testing dataset. Finally, this best model was validated on the independent validation dataset to evaluate the performance of DNN based on unbiased Root Mean Square Error (ubRMSE), Pearson Correlation Coefficient (ρ) and bias [186].

3.2.5. NORMALIZED SENSITIVITY COEFFICIENT

A sensitivity analysis was conducted to investigate the response of each DNN output to small perturbations in the DNN inputs, i.e. the sensitivity of the ASCAT observables to the ISBA LSVs. As a DNN was generated for each grid point, physical consistency in space and time between adjacent and similar grid cells provides an indication of robustness. Following Xue *et al.* [120], the normalized sensitivity coefficients (NSC) defined by Willis *et al.* [187] were obtained using:

$$\text{NSC}(j,i) = \frac{\partial y_j}{\partial x_i}\frac{x_i}{|y_j|} = \frac{y_j^i - y_j^0}{\delta x_i}\frac{x_i^0}{|y_j^0|}, \tag{3.2}$$

where x_i^0 and y_j^0 are the nominal values of input i and output j, y_j^i is the perturbed output value, δx_i is the amount of perturbation; i = 1, ...,n; j = 1, ..., m; n is the number of input LSVs and m is the number of outputs. The absolute value of y_j^0 is used to normalize the Jacobian term $\frac{\partial y_j}{\partial x_i}$ because the values of σ_{40}^o and σ' are negative. In that case, NSC keeps the same sign as the Jacobian term $\frac{\partial y_j}{\partial x_i}$.

The perturbation is applied to one LSV at a time in order to calculate the NSC. If the perturbation is too small, it will merely amplify model noise. If the perturbation is too large, the marginal function will not be a local estimate of the rate of change in the model output with respect to the change in the input [120]. Hence, a perturbation of size +/-5% of the range of x_i^0 was used.

The NSC is an indicator of the sensitivity of the measurement operator (the DNN model). These are calculated to confirm that the sensitivities of the DNN-based measurement are physically plausible. NSC is normalized Jacobian values, analysis of which is commonly performed in data assimilation studies to understand the sensitivity of the observations to the states [188–191]. This is key to identifying which states will be updated using new observation types [140], and to quantify any seasonal or other variation in this sensitivity. Note that the NSC uses the local linear relationship between the observables and states and therefore allows for a non-linear relationship between the two.

3.3. RESULTS AND DISCUSSION

3.3.1. MODEL PERFORMANCE

In Figure 3.3, the three performance metrics are mapped for the study domain to evaluate the performance of the DNN model. Note that the unbiased RMSE and bias are normalized to account for the variation in dynamic range across grid cells due to the difference in land cover. In Figure A.5 and A.6, the original values of ubRMSE and bias are shown to compare with the uncertainty of ASCAT normalized backscatter, slope, and curvature. ubRMSE is always higher than the uncertainties of the observed variables. Pearson Correlation Coefficient, ρ, is generally higher for slope and normalized backscatter than it is for curvature. The domain-averaged values of ρ are 0.84 and 0.85 for σ_{40}^o and σ', compared to 0.58 for σ''. The spatial patterns in ρ appear to be related to land cover and climate. ρ values for σ_{40}^o and σ' are higher in northern and northwestern France where seasonal variability in both quantities is largest. Comparatively low ρ values for σ_{40}^o are found in the Pyrenees and Alps. In addition to the influence of elevation and complex terrain, this poor performance is likely due to the fact that snow cover was not included in the list of inputs to the DNN in this study. Generally, ρ for σ'' reaches the largest value in central grassland and northern agricultural areas. In a small number of grid cells, ρ is around zero for one or more of the observables, but these are often grid cells with urban areas. For normalized backscatter, ubRMSE is less than 11% of the range for 98% of grid cells, with most exceptions occurring in the Alps or Pyrenees. Bias is also generally low, with 98% of values less than 3.3% of the dynamic range. Generally, the bias is positive in agricultural areas and negative elsewhere with particularly

large negative biases occurring in urban areas. For slope, the lowest ubRMSE values occur in C3/C4 crop areas where ρ is highest. The ubRMSE is less than 19% of the range for 98% of grid cells, with larger values limited to urban areas. Bias in slope is particularly low, with 95% below 3.7% of the dynamic range. Bias is slightly positive in grasslands and negative everywhere else.

The poorest performance is observed in the curvature estimates. In addition to the lower ρ values, the ubRMSE and bias are both higher fractions of dynamic range than in the case of backscatter or slope. The median values of ubRMSE and bias are 18% and -10% of the range, with no clear spatial pattern. The curvature is the second derivative of the relationship between backscatter and incidence angle. It is estimated from the local slope values and is inherently more variable than the first derivative (the slope). From a microwave perspective, it has been argued that the curvature contains information on the relative dominance of various scattering mechanisms, which depends on both the structure and the vertical distribution of moisture within the vegetation. Neither of these quantities is modeled by ISBA or closely related to states simulated by ISBA. For example, Table A.2 shows that the maximum correlation between the LSVs and σ'' is just 0.38. Therefore, it is expected that the curvature is more difficult to predict than the backscatter and curvature.

Additionally, bias shown in Figure 3.3 (g - i) is generally lower than ubRMSE for σ^o_{40} and σ', suggesting that the error is primarily due to variance rather than bias. This suggests that using the DNN as a measurement operator in assimilation could reduce or eliminate the need for bias correction. Figures A.2 and A.3 show that the DNN compresses the range of the ASCAT observables. The median range compression of σ^o_{40}, σ' and σ'' is 13%, 13% and 17% (median), respectively.

Figure 3.4 shows boxplots of performance matrices binned for the four ROIs indicated in Figure 3.1. Performance is best in the agricultural ROI where ρ is around 0.9 and the ubRMSE is around 10% of the range for both σ^o_{40} and σ'. The bias is also lower in the agricultural ROI than in the other cover types. This may be due to the strong seasonal cycles in both σ^o_{40} and σ' in agricultural areas, indicated by a larger range in σ_{40} (Figure 3.4 (m)) and σ' (Figure 3.4 (n)). In contrast, ρ for both σ^o_{40} and σ' are lowest in the broadleaf forest ROI, in which the range of σ^o_{40} and σ' is much lower.

Note that the performance is generally poor and varies considerably within the needleleaf forest ROI despite its limited spatial extent. This spatial heterogeneity may be explained by the damage caused by Cyclone Klaus in January 2009. 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.

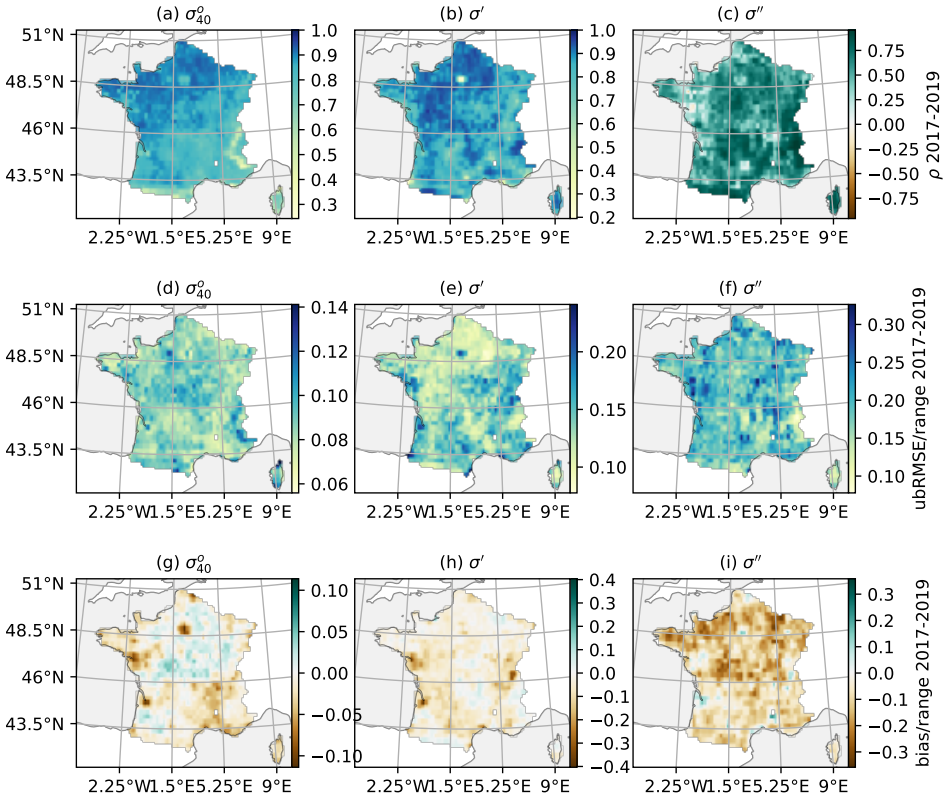

Figure 3.3.: The performance of the DNN during the validation period (2017-2019). The columns show normalized backscatter, slope, and curvature respectively. The rows show the Pearson correlation coefficient (a - c), unbiased RMSE normalized by dynamic range (d - f), and bias normalized by dynamic range (g - i). Note that the colorbars are different on each plot

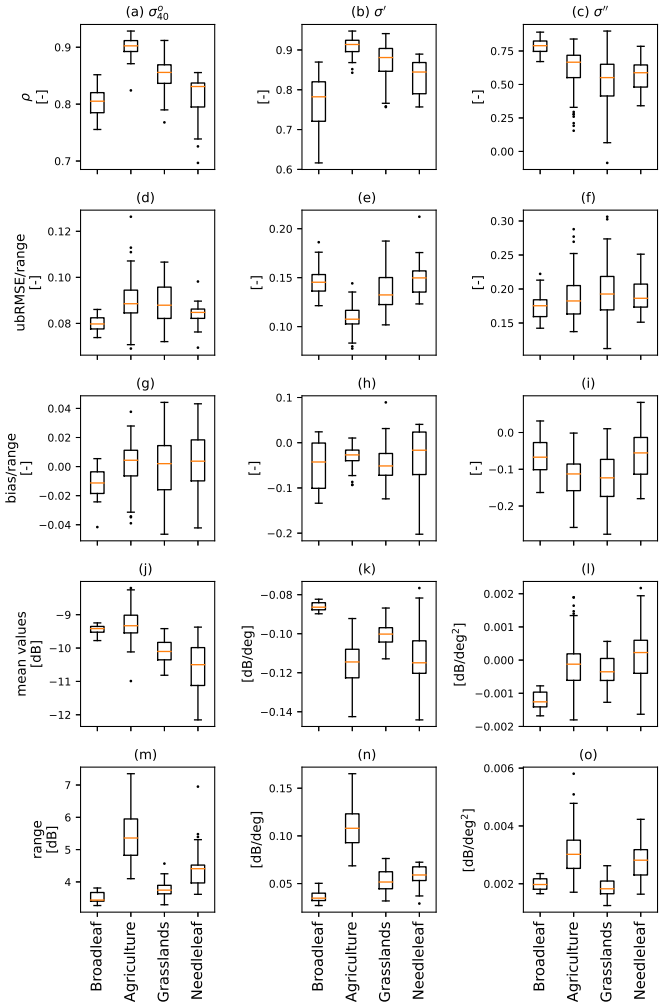

Figure 3.4.: DNN performance during the validation period (2017-2019) for each of the four ROIs. The columns (left to right) correspond to σ^o_{40}, σ' and σ'' respectively. The top three rows correspond to the Pearson correlation coefficient (a-c), ubRMSE normalized by dynamic range (d-f), and bias normalized by dynamic range (g-i). The last two rows show the mean and range of the quantities of interest per ROI. The box extends from the lower to upper quartile values of the data, with a line at the median. The lower whisker is at the lowest datum above Q1 - 1.5*(Q3-Q1), and the upper whisker is at the highest datum below Q3 + 1.5*(Q3-Q1), where Q1 and Q3 are the first and third quartiles. Outliers are indicated as grey dots.

Figure 3.5 shows the time series of predicted and observed normalized backscatter, slope and curvature for two grid points in the Agricultural ROI. Results in the left column are from the grid point within this region that has the best performance in terms of RMSE. The RSMEs in backscatter, slope, and curvature are just 5%, 8% and 13% of the range observed in these quantities. The corresponding values for the grid point presented on the right (worst performance) are almost twice as high.

From a visual comparison, their performance does not seem so different. At both grid points, the DNN captures the seasonal cycle very well. Interannual variability in backscatter and slope is captured reasonably well. This is easiest to see in the slope time series, where the predicted values match those observed in terms of both magnitude and timing. The slope is underestimated in the winters of 2015 and 2016 at the "best" grid point (Figure 3.5). Note that the observed slope is higher in these two years than in the rest of the training period. Their unusually high values may be harder to capture, particularly if they are due to an extreme value of one of the inputs or anomalous variations in a quantity not included in the list of inputs. In general, performance is the poorest for curvature. However, given our limited physical understanding of what the curvature captures in terms of geophysical variables, it is encouraging that the DNN captures the seasonal cycle and interannual variability so well. Clearly, there is scope for improvement through, for example, the inclusion of additional input variables. However, part of the poor performance may be due to the fact that curvature is generally noisier than slope. The predicted slope and curvature are noisier than those observed. This may be attributed to the fact that their values on a given day are based on all local slope values within a 42-day window, effectively smoothing the observations. Applying an Epanechnikov kernel smoother to the predicted slope and curvature *a posteriori* could reduce the random errors.

3.3.2. MODEL SENSITIVITY: A QUALITATIVE ANALYSIS

Figure 3.6 shows the remarkable agreement between the predicted and observed ASCAT observables in all ROIs, despite their diversity in terms of SM and LAI. The strongest seasonal cycles in WG2 and LAI are observed in the Agricultural ROI. Figure 3.6 (b) shows that σ°_{40} closely follows WG2 from day 200 onwards due to the low water content during crop senescence, and the sensitivity to soil moisture in the winter months in the absence of vegetation cover. This sensitivity is lost in spring as LAI rises. Figure 3.6 (b) also shows that the DNN models the strong drop in normalized backscatter from around February to April. This drop occurs long before LAI starts to increase (Figure 3.6 (f)) and starts before surface soil moisture starts to decrease (Figure 3.6 (b)). Such a rapid decrease in the real observed backscatter cannot be explained by the change in WG2 alone. It may be amplified by agricultural management practices e.g. tillage and planting in the spring, though the absence of ground data makes this difficult to prove conclusively. The decrease in σ°_{40} with increasing LAI has been observed in narrow-leaved crops (e.g. wheat and barley) in previous studies [192, 193] and can be attributed to the transition from soil dominating backscatter in the winter, to vegetation dominating during the growing season. Figure 3.6 (j) shows that the slope generally follows the overall seasonal

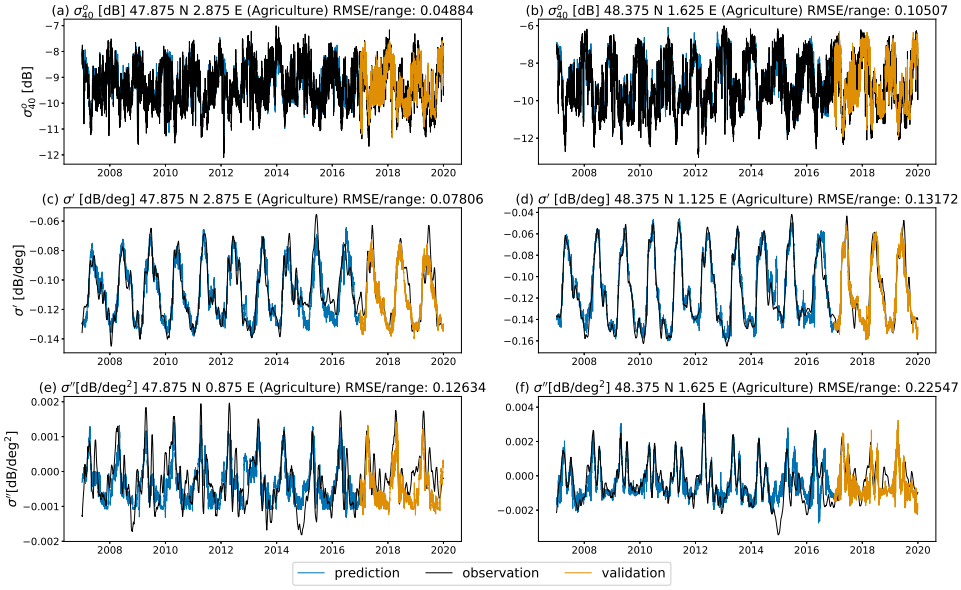

Figure 3.5.: Time series of predicted (blue) vs observed (black) values of ASCAT normalized backscatter, slope, and curvature in the agriculture ROI. The validation period (yellow) is from 2017 to 2019, shown as the brown one. The "best" (left column) and "worst" (right column) grid points were selected based on the maximum and minimum values of RMSE/range among GPIs in the agricultural ROI. The rows show ASCAT normalized backscatter, slope, and curvature respectively.

cycles of LAI closely in the Agricultural ROI. However, they diverge after the LAI peak because ISBA does not account for agricultural management practices (including harvest) and their effect on simulated LAI [9]. In contrast, the limited seasonal variation in WG2 and LAI in the Needleleaf ROI translates to limited variations in both σ°_{40} and σ'. In Figure 3.6 (d), σ°_{40} primarily follows WG2 with some seasonal influence of LAI. The agreement between σ' and LAI in Figure 3.6 (l) is consistent with the slope being an indicator of vegetation density.

The seasonality of WG2 and LAI is similar in Broadleaf and Grassland ROIs due to their geographical proximity and similar climate. More importantly, Figure A.1 shows that the Broadleaf ROI comprised 38% (median value) temperate broadleaf cold-deciduous summer green forest (TEBD) and 30% C4 grass. Similarly, around 20% of the grassland ROI consists of TEBD. Therefore, σ°_{40} and σ' in both ROIs are effectively a weighted mixture of the scattering from both TEBD and C4 grass. Comparing Figure 3.6 (a) and (c), the seasonal cycle of σ°_{40} is larger in grassland because the vegetation is less dense, allowing a stronger response to soil moisture variations. The seasonal cycle in LAI is, nonetheless, sufficient to introduce some

change in the sensitivity to soil moisture. In Figure 3.6 (k) the slope follows LAI reasonably well in Grassland, but there is clearly some influence from the TEBD fraction. The slope is generally higher and more constant in the Broadleaf ROI (Figure 3.6 (i)) due to the presence of higher biomass throughout the year.

Figure 3.6 (i) shows a clear "spring peak" in σ' before the LAI peak is clearly visible in the broadleaf ROI. A similar feature is visible in all cover types but it is on the rising limb of a much stronger seasonal cycle and therefore less significant in the overall cycle. It is also apparent in Figure 3.6 (k) due to the presence of some TEBD in the grassland grid cells. Figure 3.6 shows that the spring peak is not merely an artifact due to the uncertainty of ASCAT processing to obtain σ'. Pfeil *et al.* [6] has identified this spring peak in broadleaf deciduous forests across Europe and attributed it to the elevated water content of twigs and branches prior to leaf out.

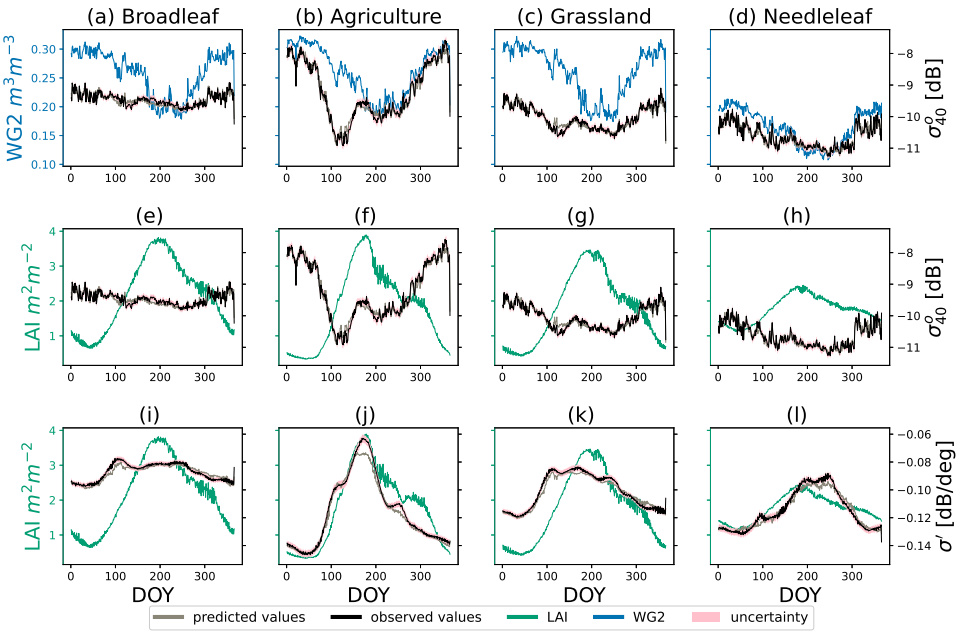

Figure 3.6.: Time series of WG2, LAI w.r.t. normalized backscatter and slope averaged from 2007 to 2019. The predicted value of normalized backscatter (slope) is the grey line and the observed normalized backscatter (slope) is the black line. The error estimate for each ASCAT observable is indicated by pink shading. Blue and green lines are LSVs of interest, e.g. WG2 (a - d) and LAI (e - l), respectively.

3.3.3. MODEL SENSITIVITY: NORMALIZED SENSITIVITY COEFFICIENT

The values of NSC(σ^o_{40}, WG2) in Figure 3.7 (a - f) are consistent with our expectations in terms of the sensitivity of σ^o_{40} to surface soil moisture. NSC(σ^o_{40}, WG2) is generally positive because backscatter increases with soil moisture, and it is highest during the winter months when the reduced vegetation cover leads to an increased sensitivity of σ^o_{40} to WG2. Comparing Figure 3.7 (a - f) to Figure 3.7 (g - l), it is clear that σ^o_{40} is considerably more sensitive to WG2 than to LAI. Values of NSC(σ^o_{40}, LAI) are generally closer to zero. NSC(σ^o_{40}, LAI) is highest, and comparable to the values of NSC(σ^o_{40}, WG2) in May and July in the west indicating that σ^o_{40} during this period is therefore sensitive to both.

Figures 3.7 (m - r) show that σ' is highly sensitive to LAI. NSC(σ', LAI) is generally positive in winter with the exception of areas with needleleaf forests. Striking spatial patterns emerge in May (Figure 3.7 (o)) and July (Figure 3.7 (p)) that correspond to the land cover patterns in Figure 3.1. NSC(σ', LAI) reaches exceptionally high values in May, before becoming equally large in magnitude but negative in July (Figure 3.7 (p).

Figure 3.8 shows seasonal variations in the sensitivity of normalized backscatter to WG2 and LAI in the four ROIs. In all cover types, NSC(σ^o_{40}, WG2) is highest during winter and lowest during the summer months when higher biomass reduces the sensitivity of C-band backscatter to soil moisture. NSC(σ^o_{40}, LAI) is generally lower than NSC(σ^o_{40}, WG2), indicating that σ^o_{40} is more sensitive to WG2 than LAI. The only exception is in the Agriculture ROI during spring when the backscatter dynamics are primarily driven by the increase in biomass during the vegetative stages of the crops, resulting in NSC(σ^o_{40}, LAI) higher than NSC(σ^o_{40}, WG2).

The strongest seasonal cycles in both NSC(σ^o_{40}, WG2) and NSC(σ^o_{40}, LAI) are observed in the Agriculture ROI. This is due to the contrast between bare soil in winter and a vegetated surface during the crop-growing season. σ^o_{40} has a strong seasonal cycle (range of 3dB) which is primarily influenced by WG2 in the bare period (DOY 0 - 100) and DOY (300-365), and a combination of WG2 and LAI in the growing season (DOY 100-300). NSC(σ^o_{40}, WG2) is higher in the Needleleaf ROI than in the Broadleaf ROI because the variations in σ^o_{40} are higher in the Needleleaf ROI and they closely follow WG2 variations (Figure 3.6 (d)). This sensitivity to soil moisture under a needleleaf forest may be due to the multiple scattering between the trunks and soil during low LAI periods. This is similar to results found in L-band backscatter modeling studies in boreal forested areas [194].

In Figures 3.8 (a-d), NSC(σ', LAI) also shows the sensitivity of slope to LAI. Figures 3.8 (a) shows that NSC(σ', LAI) is very low in the Broadleaf ROI. This is because the slope is high throughout the year in the Broadleaf ROI and exhibits little seasonal variation despite the large change in LAI (Figure 3.6 (i)). Conversely, a stronger seasonal cycle is observed in Needleleaf ROI 3.8 (a) because a relatively small change in LAI and stronger seasonal cycle in slope is observed in Needleleaf forest than in Broadleaf forest. The strongest seasonal cycle in NSC(σ', LAI) is observed in the Agriculture ROI where the biomass variations (reflected in LAI) during the crop growing season change the dominant scattering mechanism from surface scattering in winter to volumetric and multiple scattering in summer, resulting in a strong

seasonal change in slope. Note that there are sign changes in (NSC(σ', LAI)) in all ROIs. From Figure 3.6 (i) to (l), these correspond to the difference in timing between the peaks in LAI and those in slope. Combined, Figure 3.6 (i) to (l) and Figure 3.8 suggest that slope is sensitive to more than just LAI alone.

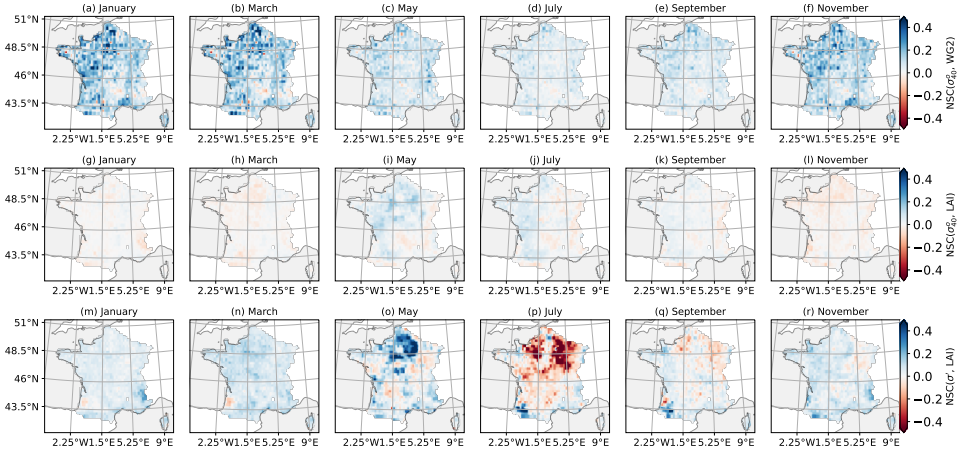

Figure 3.7.: Maps of NSC values of normalized backscatter and slope w.r.t WG2 and LAI. The columns show different months. The rows show different NSC, i.e. NSC(σ^o_{40}, WG2) (a - f), NSC(σ^o_{40}, LAI) (g - l) and NSC(σ', WG2) (g - l).

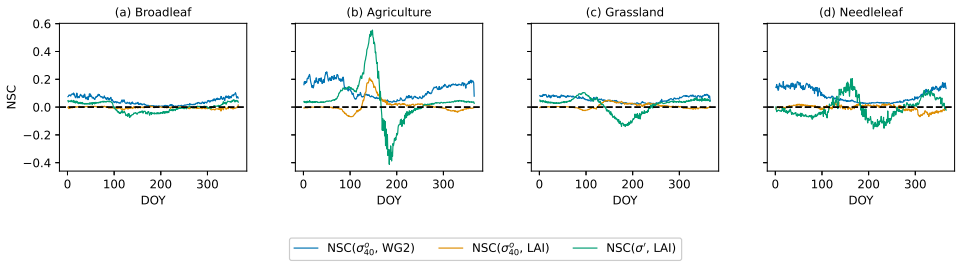

Figure 3.8.: Seasonal cycle of NSC values in Broadleaf (a), Agriculture (b), Grassland (c), and Needleleaf (d) regions of interest. The values correspond to the median value among the GPIs and were calculated using all data from 2007-2019.

Results are presented in Figure 3.9 to explain why the "spring peak" feature in Figure 3.6 (i) is captured by the DNN despite the absence of internal vegetation water content in the DNN inputs. Figure 3.9 (a) shows that the spring peak in slope coincides with a rapid decrease in soil moisture (WG) in all soil layers. The

increase in slope from DOY 160 to 250 also coincides with a decrease in soil moisture throughout the profile. Figure 3.9 (b) shows that the local maximum in slope on DOY 198 coincides with the peak in LAI, but that the spring peak coincides with a local peak in net radiation around DOY 105. Collectively, Figure 3.9 (a) and (b) illustrate that slope variations reflect vegetation growth and are influenced by both water and energy availability in this Broadleaf ROI.

Comparing Figures 3.9 (c) and (d), it is clear that the NSCs related to WG are larger in magnitude than those related to net radiation and LAI. From Figure 3.9 (c), NSC(σ', WG2) is close to zero, but the magnitude of NSC with respect to soil moisture increases with depth reaching values of -0.2 for WG7. Similarity among the NSC values with respect to WG in different layers is due to the strong correlation in soil moisture in the various layers. The magnitude of NSC(σ', LAI) is around half that of NSC with respect to soil moisture. While NSC(σ', RN) is non-zero and positive in spring, the magnitude is an order of magnitude smaller than that of NSC(σ', WG7). This suggests that net radiation may play a part in the dynamics of slope, but that the connection is indirect and weak. This connection between deep soil moisture and slope is compatible with the conclusion of Pfeil *et al.* [6] that the slope is sensing the filling of branch and twigs prior to leaf-out. It is striking that the combination of ISBA and the DNN can predict the observed spring peak by capturing the soil moisture dynamics and mapping it directly to slope change without explicitly modeling the water storage in the vegetation.

3.4. CONCLUSIONS

A DNN was trained per grid point to relate ISBA LSVs to ASCAT normalized backscatter, slope, and curvature for future use as the observation operator within a data assimilation framework. The agreement between the predicted and observed ASCAT observables during the independent validation period was very good, with better estimates obtained for backscatter and slope than for curvature. Given that we have a better understanding of the factors affecting backscatter and slope, and our expectation that curvature is influenced by vegetation geometry, this is perhaps not too surprising. Results also showed that land cover types have a large effect on model performance. Recall that an individual DNN model is obtained for every grid point. Therefore, smooth spatial patterns in the performance metrics and the appearance of features related to dominant land cover and fraction of land cover provide an indication of robustness.

Performance could be improved through the inclusion of additional LSVs to capture controls on ASCAT observables that are currently missing. For example, the relatively poor performance in the Alps is partly because snow states were not included. Snow has a strong influence on backscatter and the seasonal variation in the backscatter-incidence angle relationship. The inclusion of temperature and radiation data provides information that there is a seasonal influence affecting the backscatter, but the explicit inclusion of snow states should lead to an improvement in performance in areas affected by snow cover. Performance could also be improved by adapting the ML implementation or replacing the DNN with an alternative

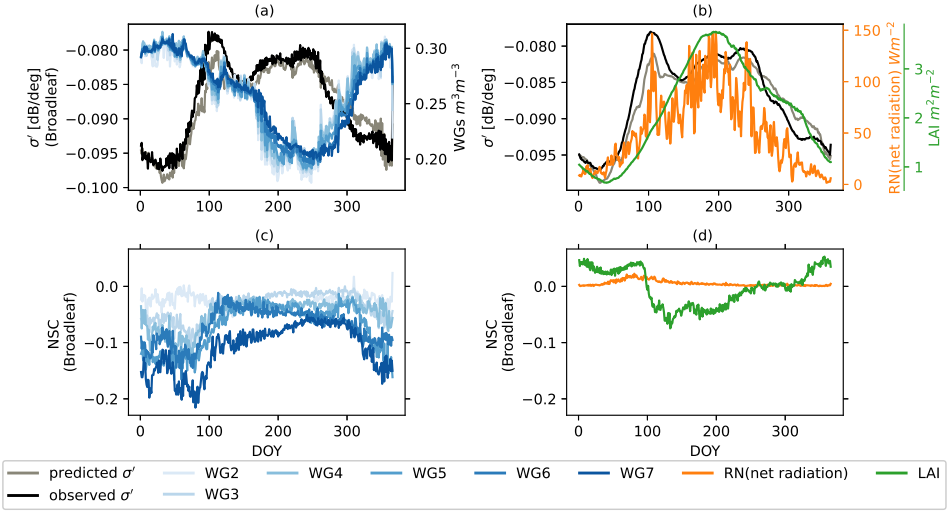

Figure 3.9.: Seasonal cycle of LSVs (including soil moisture from different depths (different WGs) (a) and vegetation related LSVs (LAI and RN) (b)) and their NSC values (c - d) of slope in Broadleaf regions of interest. The values correspond to the median value among the GPIs and were calculated using all data from 2007-2019. The vegetation-related LSVs have been smoothed by a 7-day window in order to compare their peaks with peaks in σ'.

machine-learning approach. The variability of predicted slope and curvature are generally larger than the observations. While this may partly be due to overfitting, it is primarily because the slope and curvature observations are obtained by aggregating local slope values in a 42-day window, i.e. an Epanechnikov kernel applied to the local slopes [4]. The time series of the observations are therefore, by definition, very smooth. In contrast, the DNN is trained with daily input data. One pragmatic solution to reduce the noise in the predicted observables would be to filter them *a posteriori*. Alternatively, one could consider approaches such as Long-Short Term Memory (LSTM) to handle the mapping from a temporal series of inputs to the ASCAT observables. Using LSTM or similar would also provide a means to explicitly handle the strong auto- and cross-correlations in the soil states. Furthermore, these correlations also mean that there is considerable redundancy in the list of input variables. While there is no requirement that the input variables are uncorrelated [195], performing dimensional reduction to provide a reduced set of inputs to the DNN would reduce the computational burden of the training step. One inherent limitation of DNN, and any data-driven approach, is that low probability events receive small weights and can be neglected during training and prediction. This was illustrated by the poor performance of the DNN in the wake of Cyclone

Klaus, for example. In future research, attention must be paid to how low probability events can be flagged and handled in an assimilation system.

The NSC analysis indicates that the backscatter is very sensitive to surface soil moisture and LAI and that these sensitivities vary during the year and among different land covers. The observed spatial and temporal variations in the backscatter NSCs are consistent with our understanding of backscatter from theory, observations, and modeling. This provides an important "sanity check" for our DNN, that it is capable of capturing known sensitivities. NSC values also indicate that the slope is sensitive to LAI but also to other quantities including root zone soil moisture variations or related quantities. Backscatter is sensitive to vegetation water content which is not directly accounted for in ISBA, but varies as water is transported from the root zone to the atmosphere through the vegetation as modeled in plant hydraulics model [174]. Note, however, that the NSC values are influenced by the choice of LSVs included as inputs to the DNN. In particular, our inability to explicitly include surface roughness and vegetation water content in the DNN inputs produces an increase in the sensitivity of the DNN to other ISBA LSVs such as root zone soil moisture. This means that we are effectively using root zone soil moisture to compensate for the absence of vegetation water content data.

The NSC values together with the LSV variations provide new insights into the observables themselves. This was illustrated by the spring peak in the slope of the broadleaf deciduous forest. The DNN model was able to predict the spring peak in slope before the LAI peak. The NSC and LSV time series show that the spring peak is related to radiation availability and root water uptake, consistent with the hypothesis of Pfeil *et al.* [6] that the spring peak could be attributed to increased water content in the branches prior to leaf-out. These results are significant because we can now describe the slope and curvature variations in terms of LSVs rather than scattering mechanisms. This brings us a step closer to being able to use them directly in applications or to constrain models.

One of our primary motivations was to develop a DNN model to use as a measurement operator for assimilation. Results suggest that the DNN model is capable of predicting the observations very well. The NSC analysis provides insight into the information content of the slope and curvature, and which LSVs should be included in the state vector for update. The direct mapping of LSVs to observables circumvents two key challenges related to the use of radiative transfer models in assimilation. First, it circumvents the challenge of translating the LSVs provided by the physical process model to the set of inputs required by the radiative transfer model. Second, the direct mapping between the model and observables ensures that there is a consistent climatology between the predicted and actual observables, obviating the need for CDF-matching or bias correction. Furthermore, this approach lends itself to multi-observation assimilation as the list of LSVs and observables can easily be expanded to produce a measurement operator to ingest data from multiple sources. Future research will investigate the performance of the DNN as a measurement operator in an assimilation context. In particular, additional research is needed to determine the degree to which assimilation using a DNN-based measurement operator can outperform an existing approach (e.g. Water Cloud

Model) in terms of accuracy and practicality.

4

ASSIMILATING ASCAT NORMALIZED BACKSCATTER AND SLOPE INTO ISBA

Based on: X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. "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". In: *Remote Sensing of Environment* 308 (July 2024), p. 114167. ISSN: 0034-4257. DOI: 10.1016/j.rse.2024.114167. URL: http://dx.doi.org/10.1016/j.rse.2024.114167

4.1. INTRODUCTION

In the previous chapter, a Deep Neural Network (DNN) is used to directly relate the land surface variables from the ISBA Land Surface Model to the ASCAT backscatter, slope, and curvature. The DNN-based model achieves a good agreement between the predicted and observed ASCAT observables. Results about normalized sensitivity coefficient (NSC) show that the DNN-based model could reproduce the physical response of ASCAT observables to changes in land surface variables. The DNN could potentially serve as an observation operator in data assimilation to constrain soil and vegetation water dynamics in LSMs.

This chapter builds on Shan *et al.* [124] by integrating the DNN into the Météo-France's modeling platform SURFEX/LDAS Monde data assimilation framework [88]. Following the SURFEX assimilation methodology, ASCAT observables are directly assimilated into the LSM ISBA, using the observation operator developed by Shan *et al.* [124]. This study is conducted in western Europe using ASCAT data from 2017 to 2019, at ASCAT grid points (GPIs) containing International Soil Moisture Network (ISMN) stations. Our hypothesis is that the joint assimilation of normalized backscatter and slope will improve surface and root zone soil moisture as well as vegetation states such as LAI if a) the Jacobians of the observation operator are physically plausible; b) the model and observation uncertainties are reasonable and fit for the data assimilation system. The parameter settings and implementation follow previous SURFEX/LDAS Monde studies such as Albergel *et al.* [88], Barbu *et al.* [109], and Barbu *et al.* [108]. Following the approach of Reichle *et al.* [98], Kolassa *et al.* [197] and Daley [198], diagnostics based on innovations and residuals are used to assess whether the DA system is optimal and the model and observation errors are well-specified. The DA performance is evaluated against in-situ observations of soil moisture and independent LAI data from remote sensing. The DA results are compared to the LSM open loop (OL) to test the hypothesis that joint assimilation leads to improved skill.

4.2. DATA AND METHODOLOGY

4.2.1. DATA

ASCAT NORMALIZED BACKSCATTER AND SLOPE

ASCAT normalized backscatter and slope come from a second order polynomial which describes the ASCAT backscatter dependence on the incidence angle:

$$\sigma^o(\theta) = \sigma^o(\theta_r) + \sigma'(\theta_r)(\theta - \theta_r) + \frac{1}{2}\sigma''(\theta_r)(\theta - \theta_r)^2, \qquad \text{[dB]} \tag{4.1}$$

where the zeroth order term $\sigma^o(\theta_r)$ is the normalised backscatter at a reference angle (40°), and the 1st and 2nd order coefficient $\sigma'(\theta_r)$ and $\sigma''(\theta_r)$ are so-called slope and curvature [4]. Given the values of slope and curvature, measured backscatter at any incidence angle can be extrapolated to the reference angle by

$$\sigma^o(\theta_r) = \sigma^o(\theta) - \sigma'(\theta_r)(\theta - \theta_r) - \frac{1}{2}\sigma''(\theta_r)(\theta - \theta_r)^2. \tag{4.2}$$

Each ASCAT instrument has two sets of three side-looking antennas (oriented at 45° (fore), 90° (mid) and 135° (aft) to the satellite track) each illuminating a 550 km wide swath on either side of the satellite track. Together, these provide so-called backscatter triplets [σ_f, σ_{mid}, σ_a] (Fore, Mid and Aft beam) at the incidence angles [θ_f, θ_{mid}, θ_a] (same angle 34–65° for Fore and Aft beam, 25–55° for Mid beam). Thus two local slope estimates are produced as:

$$\sigma'\left(\frac{\theta_{mid}-\theta_{a/f}}{2}\right)=\frac{\sigma^o_{mid}(\theta_{mid})-\sigma^o_{a/f}(\theta_{a/f})}{\theta_{mid}-\theta_{a/f}}, \qquad \text{[dB/deg]} \qquad (4.3)$$

Then, slope and curvature values on the day d are estimated using all local values distributed over the entire incidence angle range within a 42-day window centered at day d [4]. The value of the length of the window is chosen to balance the bias and variance of estimates. The performance and robustness of the kernel smoother were tested in Hahn *et al.* [4].

VALIDATION DATA

The open loop (OL) and data assimilation (DA) simulations are compared to the in-situ soil moisture measurements of the International Soil Moisture Network (ISMN; [199, 200]) at locations illustrated in Figure 4.1. ISMN stations were excluded if the percentage of effective time series was less than 85%. In addition, following Corchia *et al.* [138], observations were not assimilated if the SURFEX model indicates the presence of snow or frozen soil based on the initial values of the ice in the second layer (WGI2). The dominant vegetation types at ISMN stations are illustrated in Figure 4.1. The ISMN soil moisture measurements are quality controlled, and stations with a data record of less than one year are removed. The relationship between ISMN observations at specific depths and the ISBA simulated soil moisture in various layers is shown in Table A.1. Root zone soil moisture (RZSM) is calculated as the soil moisture in a layer from 0.1 to 0.4m because 20 of the 22 ISMN stations have measurements at 0.2m and 0.4m. In this study, surface soil moisture (SSM) is defined as soil moisture from 0 to 0.1m measured at the ISMN station or simulated by the land surface model. A summary of the ISMN stations is provided in Table A.6.

The LAI 1km version 2 products from the Copernicus Global Land Service (CGLS) [201] are used as the independent evaluation data. Following Albergel *et al.* [88], the 1 km resolution observations are averaged to the 0.25 degree model grid points, as long as 50% of the observation grid points are observed. LAI observations have a sampling interval of 10 days. LAI from ISBA simulations (both OL and DA) are evaluated against the CGLS LAI on days when the CGLS LAI is available.

4.2.2. METHODOLOGY

ISBA-A-GS MODEL

This study uses the CO2-responsive version of ISBA-A-gs (Interactions between Soil, Biosphere and Atmosphere [141, 142], here referred as ISBA) within the SURFEX platform (version 8.1) [69, 88, 151]. For more details please see Section 3.2.3 and Section 2.1.

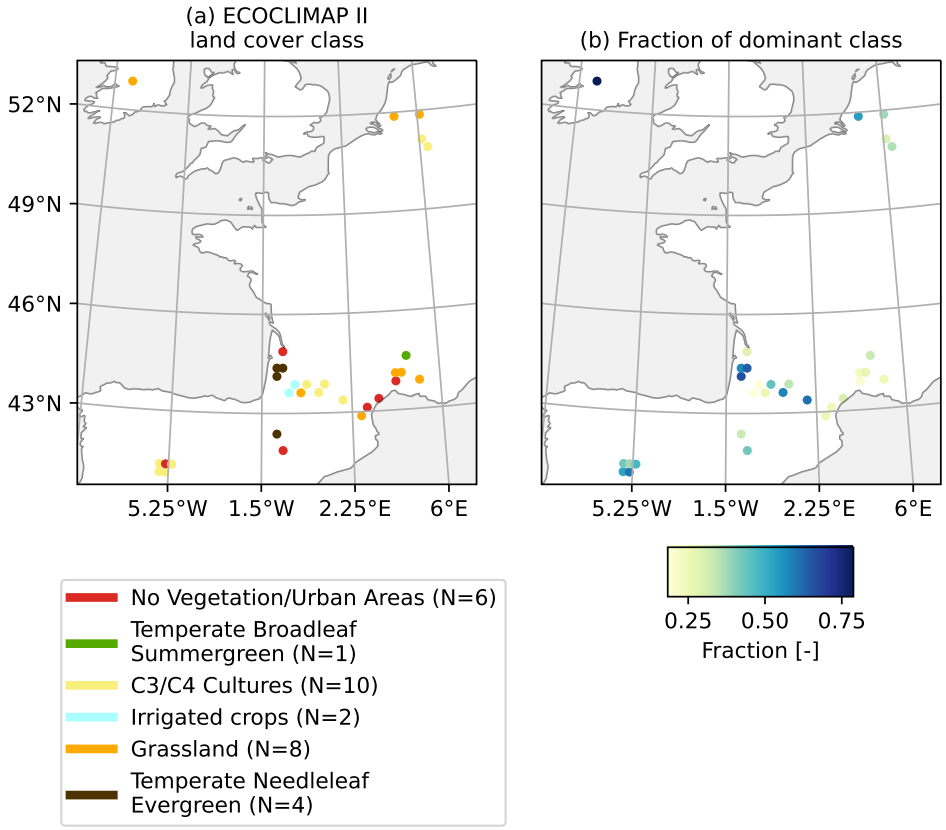

Figure 4.1.: The main land cover types and fraction of France based on ECOCLIMAP II. (a) the dominant vegetation types of the 0.25° x 0.25° GPIs which contain the ISMN stations; (b) the vegetation fraction of the dominant vegetation types of each GPI. The GPI in Ireland is not included in the later calculation of performance metrics due to the short ISMN data record.

The model was forced by the latest ERA5 (ECMWF Reanalysis v5) atmospheric reanalysis from ECWMF [164] from 1996 to 2019. The meteorological forcing data are available on a 0.25° x 0.25° grid, and include rainfall rate, 2m air temperature, 2m specific humidity, wind speed, wind direction, surface pressure, downward direct shortwave radiation, downward diffuse radiation, downward longwave radiation, snowfall rate and CO2 concentration. All ERA5 atmospheric variables were interpolated using bilinear interpolation to match the grid points of ISBA. The model was initiated by spinning up using the 1996 forcing data. 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.

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

$$\tilde{\mathbf{x}}_k^a = \tilde{\mathbf{x}}_k^f + \tilde{\mathbf{K}}_k(\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^f)), \tag{4.4}$$

where superscripts a, f, o indicate the analysis, forecast, and observation, respectively. The term $\tilde{\mathbf{x}}_k$ represents the control vector or state matrix 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 dimensions of the state vector. The term $\mathbf{x}$ represents the state vector aggregated over different patches which contain LAI, WG2, WG3, WG4, WG5, WG6, WG7, WG8 as described in Table A.1. $\mathbf{y}^o$ denotes the observation vector of dimension nbo. $\mathscr{H}$ denotes the non-linear observation operator, i.e. the DNN, which utilizes the input state vector aggregated over different patches within one GPI as well as diagnostic variables. Following Shan *et al.* [124], the relevant diagnostic variables are net radiation (RN), vegetation interception reservoir water storage (WR), plant transpiration (LETR), gross primary production (GPP), and stomatal conductance (XRS) (See Table A.1). The DNN is trained and tested using data from 2007 to 2016. Jackknife cross-validation is used to choose the best model that minimizes RMSE among those submodels according to their performances on the testing dataset. Then the DNN is independently validated from 2017 to 2019 following Shan *et al.* [124]. The input data of DNN comes from the OL experiment of ISBA from 2007 to 2019. Input land surface variables include the soil moisture in different layers and vegetation-related variables (See Table A.1). The DNN is trained independently per GPI to account for the heterogeneity of ASCAT footprints which contain combinations of different patches. The structure of the DNN is tuned by Bayesian optimization [173], including the number of layers, the number of neurons of each layer, batch size, the choice of activation function, and learning rate. More details can be found in Shan *et al.* [124].

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

$$\tilde{\mathbf{K}}_k = \begin{pmatrix} \mathbf{K}_{k,[1]} \\ \mathbf{K}_{k,[2]} \\ \dots \\ \mathbf{K}_{k,[12]} \end{pmatrix} = \tilde{\mathbf{B}}\tilde{\mathbf{J}}_k^T(\tilde{\mathbf{J}}_k\tilde{\mathbf{B}}\tilde{\mathbf{J}}_k^T + \mathbf{R})^{-1}, \tag{4.5}$$

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

$$\mathbf{x}_{k,[tp]}^a = \mathbf{x}_{k,[tp]}^f + \mathbf{K}_{k,[tp]}(\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^f)), \tag{4.6}$$

and $\tilde{\mathbf{J}}$ is the Jacobian values of the model ($\mathbf{M}_{k,[tp]}$) and the linearized observation

operator ($\mathbf{H}_{k,[tp]}$) at time step k for patch p, which follows:

$$\tilde{\mathbf{J}} = \begin{pmatrix} \mathbf{J}_{k,[1]} \\ \mathbf{J}_{k,[2]} \\ \dots \\ \mathbf{J}_{k,[12]} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{k,[1]}\mathbf{M}_{k,[1]} \\ \mathbf{H}_{k,[2]}\mathbf{M}_{k,[2]} \\ \dots \\ \mathbf{H}_{k,[12]}\mathbf{M}_{k,[2]} \end{pmatrix}, \tag{4.7}$$

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

$$\tilde{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_{k,[1]} & 0 & \dots & 0 \\ 0 & \mathbf{B}_{k,[2]} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \mathbf{B}_{k,[12]} \end{pmatrix}, \tag{4.8}$$

Here, 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 $\mathbf{B}_{k,[tp]}$, p=1,...,12 are defined following Mahfouf *et al.* [107], [102], and [88]. 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 [142]. In this study, the perturbation applied to the states of soil moisture in the calculation of the Jacobian matrix were assigned as $1 \times 10^{-4} \times (w_{fc} - w_{wilt})$ following Albergel *et al.* [88], and $0.001 \times LAI$ for LAI. For volumetric surface soil moisture, a mean standard deviation (SD) error of 0.04 m^3m^{-3} is prescribed. For soil moisture in deeper layers, an SD error of 0.02 m^3m^{-3} is used following Mahfouf *et al.* [107], [102], [108], and [109]. The observation error for σ^o_{40} is set as 0.33 dB following Lievens *et al.* [8] and 0.005 dB/deg for σ' following Wagner *et al.* [45] and Hahn *et al.* [4].

The control vector evolution from time k to the end of the 12h assimilation window (k+1) follows:

$$\tilde{\mathbf{x}}^f_{k+1,[tp]} = \mathscr{M}_{k,[tp]}[\tilde{\mathbf{x}}^a_{k,[tp]}] = \tilde{\mathbf{M}}_{k,[tp]}\tilde{\mathbf{x}}^a_{k,[tp]}, \tag{4.9}$$

where $\mathscr{M}_{k,[tp]}$ denotes the LSM ISBA-A-gs at time k at patch p. The term $\tilde{\mathbf{M}}_{k,[tp]}$ is the linearization of $\mathscr{M}_{k,[tp]}$ at patch p. For σ^o_{40} and σ', the assimilation is done every 24 hours at 10 am.

By definition, the Kalman gain in Equations 4.4 and 4.5 would be optimal if the modeling process in Equations 4.9 is linear, and the model and observation noises are zero mean and uncorrelated white noise with correctly specified covariance $\tilde{\mathbf{B}}$ and $\mathbf{R}$ respectively [98]. In this case, the SEKF estimates would provide the best least-squared estimates of the updated states given the value of the OL results, observations, model errors, and observation errors [202]. The following subsection introduces the diagnostics used to determine the degree to which these assumptions are valid, and therefore whether the DA system is optimal.

PERFORMANCE METRICS AND DIAGNOSTICS

The performance of the DA and OL are evaluated using unbiased root mean square error (ubRMSE, [186]), bias, and the Pearson correlation coefficient (ρ) against the ISMN observations and CGLS LAI observations. Additionally, the normalized information contribution (NIC, [69]) is used on ubRMSE and ρ to quantify the relative improvement or degradation of DA compared to OL estimates of states. The NIC is calculated for ubRMSE and ρ as follows:

$$NIC_{ubRMSE} = \frac{ubRMSE_{DA} - ubRMSE_{OL}}{ubRMSE_{OL}} \times 100, \tag{4.10}$$

$$NIC_{\rho} = \frac{\rho_{DA} - \rho_{OL}}{1 - \rho_{OL}} \times 100. \tag{4.11}$$

These metrics provide a measure of the percentage improvement (negative NIC for ubRMSE or positive NIC for ρ) or degradation (positive NIC for ubRMSE or negative NIC for ρ) in ubRMSE and ρ achieved by the DA system compared to the OL estimates.

The evaluation of ISBA simulations against ISMN measurements follows the best practices outlined by Gruber *et al.* [203] and used by e.g. Lievens *et al.* [8], Lannoy *et al.* [95, 96], and Barbu *et al.* [109]. If a GPI contains a single ISMN station, the in-situ soil moisture measurements from the ISMN station are directly compared to the ISBA simulations (open loop and data assimilation experiments). For grid points containing multiple ISMN stations, the soil moisture measurements from all ISMN sites are weighted and averaged, taking into account the land cover at the ISMN site and the fractional cover of that type of ISBA within the GPI.

The impact of data assimilation can be analyzed and diagnosed in the observation space. All diagnostics are summarized in Table 4.1 along with the corresponding hypothesis and references. In an optimally calibrated and linear system that satisfies the assumptions outlined in section 2.2.2, the innovations (observation-minus-forecast, O-F) are zero-mean white noise. This reflects an unbiased analysis that effectively extracts all information from the observations [198]. To test this hypothesis, the time series mean and lagged autocorrelation of the innovations are calculated. Autocorrelations at lags from 1 to 10 days are considered. Specifically, Daley [198] demonstrates that the autocorrelation at a lag of 1 day (C^k_{k+1}) reflects the optimality of the Kalman gain. The optimality of the Kalman gain is achieved when the model and observation operator are linear and there is no misspecification of the forecast (or model) and observation errors [198]. Consider the lagged innovation covariance matrix C^k_{k+1} at time step k defined as equation (2.12) in Daley [198],

$$C^n_{n+1} = \tilde{\mathbf{J}}_k[\tilde{\mathscr{B}}\tilde{\mathbf{J}}^T - \tilde{\mathbf{K}}_k(\tilde{\mathbf{J}}_k\tilde{\mathscr{B}}\tilde{\mathbf{J}}_k^T + \mathscr{R})], \tag{4.12}$$

where $\tilde{\mathbf{K}}$ is the estimated Kalman gain from Equation 4.5, $\tilde{\mathscr{B}}$ and $\mathscr{R}$ are the correctly specified model and observation errors, respectively. If the DA system is optimal, then the estimated Kalman gain is equal to the optimal Kalman gain, i.e. $\tilde{\mathscr{K}}_k = \tilde{\mathscr{B}}\tilde{\mathbf{J}}^T[\tilde{\mathbf{J}}_k\tilde{\mathscr{B}}\tilde{\mathbf{J}}_k^T + \mathscr{R}]^{-1}$.

If the lagged innovation covariance is close to zero, it indicates that the Kalman gain matrix is optimal [198, 202] and the DA system extracts most of the available

information from the observations. Conversely, the observations are not used efficiently if the lagged autocorrelation values are large.

Moreover, if the model errors and observation errors are uncorrelated and normally distributed, the standard deviation of the innovations should be equal to the sum of the covariance of the model forecast and observation errors with correctly specified model and observation error [98, 108, 204, 205]. That is,

$$E[(\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^f))(\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^f))^T] = \mathbf{R} + \tilde{\mathbf{J}}_k \tilde{\mathbf{B}} \tilde{\mathbf{J}}_k^T. \tag{4.13}$$

The term on the left hand side represents the actual errors encountered in the DA systems while the right hand side contains the prescribed, assumed errors. Thus, the time series standard deviation values of the normalized innovations (normalized observation-minus-forecast or normalized O-F) are computed as:

$$\text{std(normalized innovations)} = \text{std}\left(\frac{\mathbf{y}_k^o - \mathscr{H}(\mathbf{x}_k^f)}{\sqrt{\mathbf{R} + \tilde{\mathbf{J}}_k \tilde{\mathbf{B}} \tilde{\mathbf{J}}_k^T}}\right). \tag{4.14}$$

Values of this standard deviation greater than one indicate that the DA system underestimates the actual errors (observation error and forecast error), and values less than one indicate that the actual errors are overestimated. Note that the diagnostic only addresses the total error and does not distinguish between observation and forecast errors.

Another useful diagnostic is the residual (Observation-minus-analysis, or O-A), which is defined as the difference between the observation and the observation equivalent predicted from the updated states. In a well-calibrated system, the time series mean values of the residuals should be zero. The time series standard deviations of residuals should be smaller than the time series standard deviations of innovations, which would reflect a reduction of uncertainty in the estimated observation equivalent in the observation space [98].

The analysis impact is also assessed using the absolute value of the ratio of the residuals to the innovations, i.e. $|(O-A)/(O-F)|$ [197]. The absolute values of residuals should be smaller than the innovations. This means that the observation equivalent($\mathbf{y}^a$) of the analysis ($\mathbf{x}^a$) should be closer to the true observation $\mathbf{y}^o$ compared to the observation equivalent ($\mathbf{y}^f$) of the forecast ($\mathbf{x}^f$) [197]. Thus the mean ratio of $|(O-A)/(O-F)|$ should be smaller than 1. In the scalar case, this ratio is equal to 1 - $\tilde{\mathbf{K}}_k$ where $\tilde{\mathbf{K}}_k$ is the Kalman gain. Values smaller than 1 indicate that the DA system is effective.

4.3. RESULTS

4.3.1. PERFORMANCE

PERFORMANCE ACROSS ALL GPIS

In Figure 4.2, scatter plots of ubRMSE, bias, and ρ are shown to evaluate the performances of the DA and OL experiments against the in-situ ISMN soil moisture and CGLS LAI observations. Summary statistics are reported in Table 4.2.

Table 4.1.: Summary of diagnostics along with the corresponding hypothesis and references.

Diagnostics	Ideal values	Hypothesis	References
mean(innovations)	0	The innovations are zero-mean white noise	[202], [198], [98]
mean(residuals)	0	The system is well-calibrated	[98]
std(innovations) - std(residuals)	> 0	The DA reduces the uncertainty	[98]
Autocorrelation of innovations lagged on different days	0	The innovations are zero-mean white noise and the Kalman gain is optimal	[198], [98]
std(normalized innovations)	1	The actual errors are well prescribed	[204], [205], [108], [98]
mean$\lvert O-A/O-F\rvert$	< 1	The DA system is effective	[197]

Regarding LAI, the median value of ubRMSE from the OL across GPIs is $0.58\,m^2m^{-2}$. The spatially median value of NIC of ubRMSE for LAI is 2.2%. While the positive NIC values indicate that the DA is making the LAI estimate worse, the magnitude of the degradation is so small that the impact is effectively neutral. For bias, the median value across GPIs is $-0.23\,m^2m^{-2}$ for the OL and $-0.30\,m^2m^{-2}$ for the DA. The magnitude of the bias is slightly changed but all GPIs are centered around the 1:1 line. This is because SEKF is a bias-blind data assimilation algorithm so SEKF only reduces the random noise in states rather than correcting the bias [206]. For ρ, most agricultural GPIs fall below the 1:1 line in the scatter plots. Figure 4.2 (c) shows that DA reduces the ρ of LAI for most GPIs. Figure 4.3 (a) provides some insight as to why. The estimated LAI from the OL does not capture the seasonal cycle observed by the CGLS LAI. While DA clearly impacts the estimated LAI, it does not lead to an improvement with respect to the independent observations.

For SSM (WG3), the median values across all GPIs of ubRMSE from the OL is $0.04\,m^3m^{-3}$. In general, the DA updates are neutral on agricultural and grassland GPIs, and the median value across all GPIs of NIC is too small (-0.22%) so that the updates have nearly zero effects. The median values across GPIs of bias from the OL and DA are $-0.081\,m^3m^{-3}$ and $-0.082\,m^3m^{-3}$, respectively. Bias barely changes because SEKF does not correct bias. However, the potential for improvement may be limited by the already excellent performance of the OL simulations, with ubRMSEs values generally below $0.05\,m^3m^{-3}$. For RZSM (WG4&WG5), the median value of ubRMSE OL across GPIs is $0.029\,m^3m^{-3}$. The median values across GPIs of bias from the OL and DA are $-0.084\,m^3m^{-3}$ and $-0.085\,m^3m^{-3}$, respectively. The DA

does not change the bias compared to ISMN measurements. The DA makes limited improvements in terms of ubRMSE and ρ.

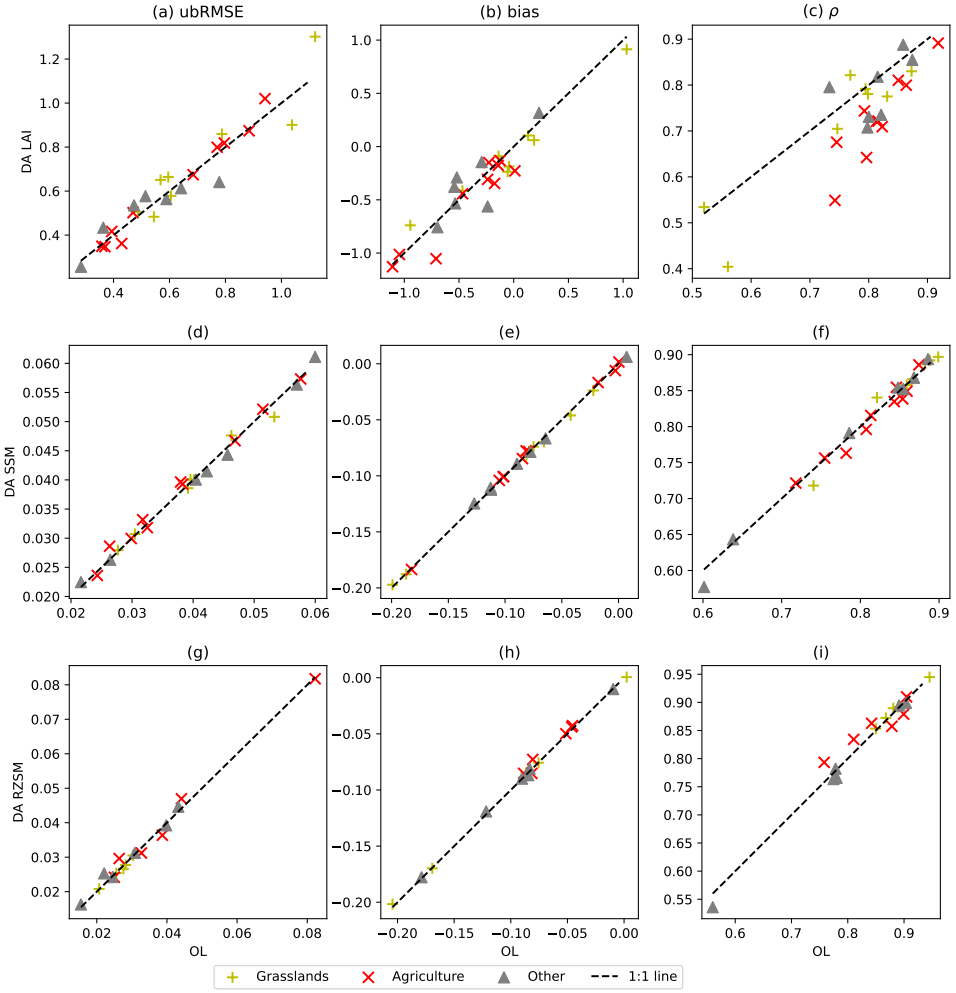

Figure 4.2.: Scatter plot of the ubRMSE, bias and Pearson correlation coefficient for LAI, SSM (WG3, 0.04 m - 0.1 m soil moisture), RZSM (WG4&WG5, 0.1 m - 0.4 m) for OL and DA(σ^o_{40}, σ') evaluated against in-situ ISMN observations and Copernicus LAI observations from 2017 to 2019. Agricultural GPIs are marked as 'x', grasslands GPIs as '+', and other GPIs as triangles. The absence of some stations of RZSM is due to the lack of sensors at some depths in the ISMN at some stations.

Table 4.2.: Statistics of performances of OL and DA averaged across all GPIs, grassland GPIs, and Agriculture GPIs.

	all GPIs	Grassland	Agriculture
LAI			
OL ubRMSE (m^2m^{-2})	0.58	0.6	0.58
DA ubRMSE (m^2m^{-2})	0.58	0.66	0.59
NIC ubRMSE	2.2	6.5	0.81
OL ρ	0.81	0.78	0.81
DA ρ	0.76	0.78	0.72
NIC ρ	-6.5	-3.6	-10
SSM (WG3)			
OL ubRMSE (m^3m^{-3})	0.04	0.04	0.035
DA ubRMSE (m^3m^{-3})	0.04	0.04	0.036
NIC ubRMSE	-0.22	0.8	0.82
OL ρ	0.85	0.87	0.83
DA ρ	0.85	0.87	0.83
NIC ρ	0.027	-0.069	-0.38
RZSM (WG4&WG5)			
OL ubRMSE (m^3m^{-3})	0.029	0.028	0.036
DA ubRMSE (m^3m^{-3})	0.03	0.027	0.034
NIC ubRMSE	-0.51	-0.74	-1.9
OL ρ	0.87	0.88	0.86
DA ρ	0.87	0.89	0.86
NIC ρ	0.3	0.41	1.5

STATES AT AGRICULTURAL GPIS

Figure 4.3 shows the time series of DA, OL, and observations of LAI, SSM (WG3), RZSM (WG4 & WG5), WG6, WG7, and WG8 averaged over the agricultural GPIs in southwestern France (The Saint Felix de Lauragais, Lahas, Condom and Savenes stations from the SMOSMANIA network). Their geographical proximity allows us to average their statistics. Both the OL and DA estimates of LAI are higher than the CGLS LAI observations. The assimilation lowers the LAI values during the spring of 2017 and summer of 2019. The assimilation of σ^o_{40} results in negative increments of LAI from Jan to April in 2017 and from March to July 2019. During these periods, the ratios of residual to the innovation of σ^o_{40} are reduced (Figure 4.4 (a)). In May 2017, the update of LAI is primarily due to the assimilation of σ'. The residual/innovation of σ' is less than unity during this period (Figure 4.4 (b)). The DA draws the LAI towards the CGLS LAI by the negative innovations of σ' (Figure 4.4 (b)) with positive J(σ', LAI) (Figure 4.9 (j)).

From Figure 4.3, DA has little to no effect on SSM (WG3) and RZSM (WG4&WG5) at these stations. The ubRMSE of the OL is already small, so the SSM (WG3) cannot be improved much. The OL cannot be improved much. The difference between

both OL and DA results and the in situ observations could partly be explained by representativeness error due to the mismatch in scales between the in situ, point-scale observations and the OL and DA estimates which are at $0.25° \times 0.25°$. This mismatch results in a large bias between ISBA simulations and ISMN measurements and the SEKF does not correct this bias. Note that the DA creates wetter estimates of the RZSM (WG4&WG5) compared to the OL in Sep 2019. This is attributed to the negative J(σ^o_{40}, RZSM) (Figure 4.9 (c)) and negative innovations of σ^o_{40} shown in Figure 4.4 (a).

It is noticeable that both the OL and DA simulations of SSM and RZSM (WG4 & WG5) have a different dynamic range compared to ISMN in-situ measurements. This systematic bias is due to the fact that soil moisture data from different sources (LSM simulations, satellite retrievals, ground measurements) exhibit different climatologies though they contain consistent information about soil water dynamics [105]. Therefore, ubRMSE is used as the performance metric to reflect the root mean square error of time series anomalies that are computed by removing the mean [186]. This is also used in previous studies when evaluating the DA results against in-situ measurements [8, 69, 88, 92, 95, 96, 109]. Meanwhile, we separated ISMN in-situ measurements into a twin axis in Figure 4.3. This is to emphasize that the dynamics between ISBA and ISMN data are comparable, and the differences in ranges of ISBA simulations and ISMN measurements are primarily related to different climatologies. However, it is essential to pay attention to the representativeness error in the dynamics, because ground-based measurements are sparse and not necessarily representative of large-scale soil moisture simulations [207].

WG6, WG7, and WG8 are pushed towards drier estimates from October 2018 to May 2019. This starts from October 16 2018, after which Figure 4.4 (a) shows that residual/innovation of σ^o_{40} reaches a local minimum. A similar situation occurs on Jan 27 2019, when the updates of WG6, WG7, and WG8 are driven by the assimilation of σ^o_{40}. In this case, DA reduces the ratio of residual to innovations of σ^o_{40} more than σ'. Note that soil moisture in deeper layers have larger increments from October 2018 to May 2019. This is due to the larger sensitivities of the observations to deeper WGs as shown in Figure 4.8 (b), (e), (c) and (f).

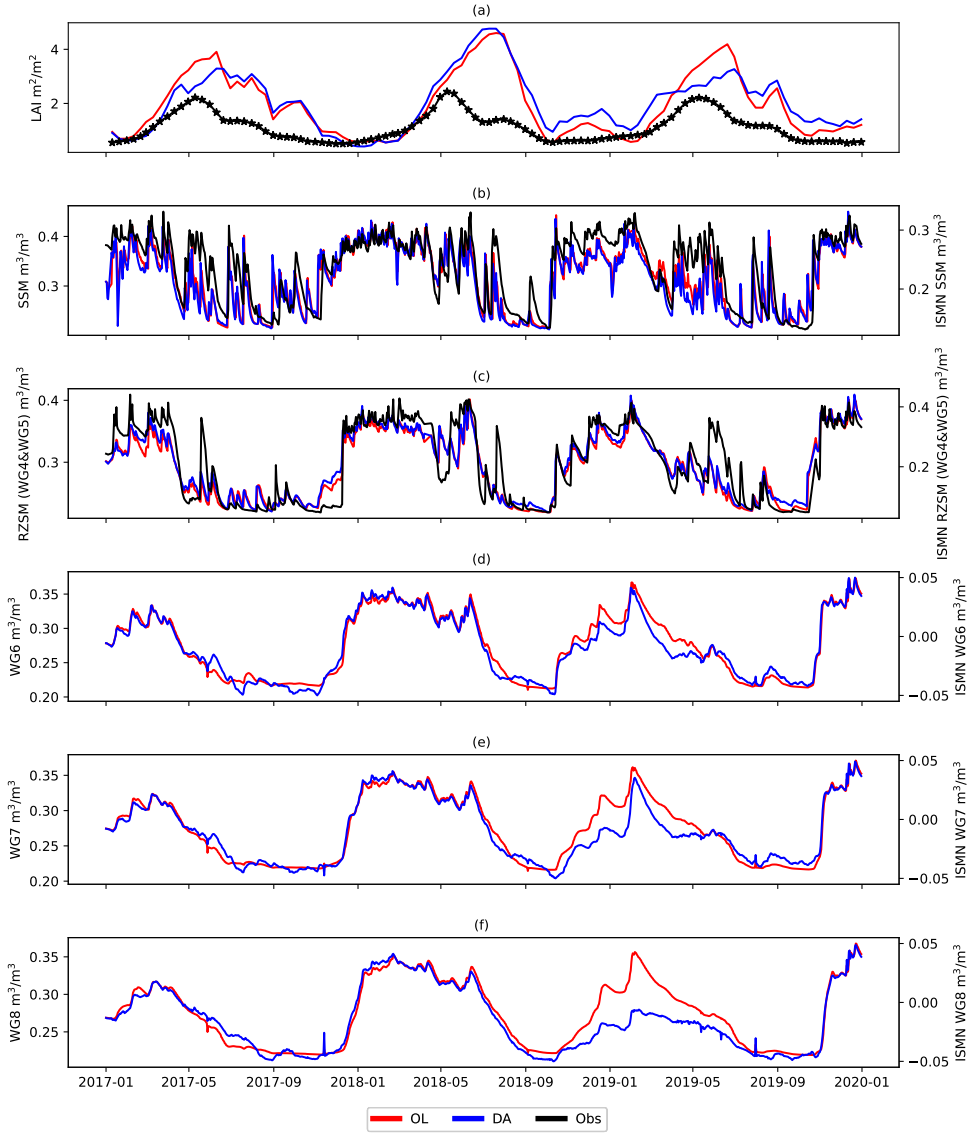

Figure 4.3.: Time series of LAI, WG3 (0.04 m - 0.1 m soil moisture), root zone soil moisture (RZSM, 0.1 m - 0.4 m) for observation, OL and DA(σ^o_{40}, σ') from 2017 to 2019, averaged on agri GPIs in southwestern France (GPIs containing ISMN stations: Saint Felix de Lauragais, Lahas, Condom and Savenes in SMOSMANIA network). Simply averaging all agricultural GPIs might suffer from different weather conditions which shapes different time series behaviors in soil moisture curves.

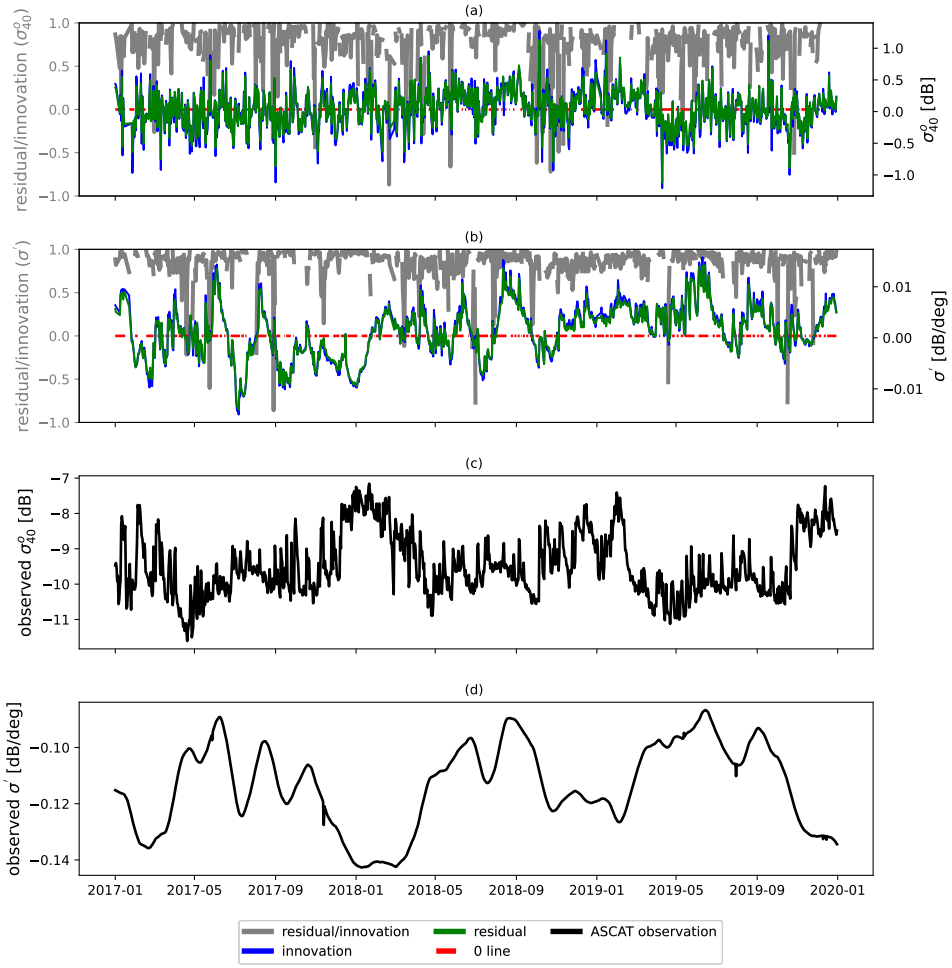

Figure 4.4.: Time series plot innovations (blue line), residuals (green line) and the ration of residuals (grey line) of a) σ^o_{40} and b) σ' from 2017 to 2019, averaged on agri GPIs containing ISMN stations: Saint Felix de Lauragais, Lahas, Condom and Savenes in SMOSMANIA network.

4.3.2. DATA ASSIMILATION DIAGNOSTICS

INNOVATIONS AND RESIDUALS

Figure 4.5 (a-b) and (e-f) show the time series mean and standard deviation of innovations (observation-minus-forecast, O-F). The median and standard deviation across all GPIs are reported in Table 4.3. The time series mean values of innovations are typically small and range from -0.17 dB to 0.25 dB for σ^o_{40} and from

-0.0025 dB/deg to 0.0028 dB/deg for σ'. Overall, the median values across GPIs are 0.018 dB for σ^o_{40}, and 0.00055 dB/deg for σ'. For most GPIs, there is small positive bias of σ^o_{40} and σ', with fewer stations exhibiting negative mean innovations values. Small values of the time series mean of innovations suggest that the assimilation system is nearly bias-free. For residuals, the time mean values of residuals range from -0.15 dB to 0.25 dB for σ^o_{40} and from -0.0023 dB/deg to 0.0026 dB/deg for σ'. Generally there is small positive bias of 0.031 dB for σ^o_{40} and 0.00041 dB/deg for σ'. The time mean values of residuals have smaller median values across GPIs and range compared to the innovations.

The time series standard deviation of innovations range from 0.17 dB to 0.52 dB for σ^o_{40} and from 0.0032 dB/deg to 0.0076 dB/deg for σ'. The median value across GPIs is 0.29 dB for σ^o_{40} and 0.0047 dB/deg for σ'. High values are found in GPIs containing REMEDHUS stations in Spain, where the main land cover type is agricultural. This may also be related to the fact that the subsurface scattering effects may be present in GPIs containing REMEDHUS stations [208]. The standard deviation of innovations of σ' is also higher in GPIs in southwestern France. At these GPIs, σ^o_{40} and σ' typically exhibit strong variability as shown in Figure A.9 (c) and (d). The median value across GPIs of the time series standard deviation of residuals is 0.27 dB and 0.0047 dB/deg for σ^o_{40} and σ', respectively, which is smaller than the innovations. This reduction represents the impact of the ASCAT observations on the DA system, i.e. a decrease in the uncertainties following the assimilation of ASCAT σ^o_{40} and σ'. However, the fact that these reductions are generally small across all GPIs suggests that the DA framework is too conservative in updating the states in observation space.

PRESCRIBED ERRORS AND THE IMPACT OF DA

Figure 4.6 (a-b) shows the standard deviation of the normalized innovations which measures the consistency between prescribed errors and the actual errors in the data assimilation system [98]. Overall, the median values across GPIs of the metric are smaller than 1 with values of 0.83 for σ^o_{40} and 0.87 for σ'). This suggests that, in general, the magnitudes of pre-assumed model and observation errors tend to overestimate the magnitudes of prescribed model errors. This metric varies considerably, particularly with land cover type. Specifically, agricultural GPIs have larger values of the metric compared to grassland GPIs. In grassland GPIs, the median values across GPIs are 0.68 for σ^o_{40} and 0.7 for σ', indicating that the prescribed model and observation errors overestimate one or both of the actual errors. In contrast, on agricultural GPIs, the spatial median value is 1 for σ^o_{40} and 0.78 for σ', suggesting that overestimation is less of an issue in agricultural GPIs.

Additionally, note that the spatial standard deviation of the metric is larger for σ' than for σ^o_{40}. This difference could be partially due to the use of spatially constant observation errors for σ^o_{40} and σ'. This does not capture any spatial variability in observation error due to e.g. heterogeneity, land cover type, etc. As a result, the degree of overestimation varies across different GPIs.

Figure 4.6 (c-d) show the mean values of $|\frac{O-A}{O-F}|$ for σ^o_{40} and σ'. This quantifies the impact of data assimilation by assessing the correction of observation equivalents

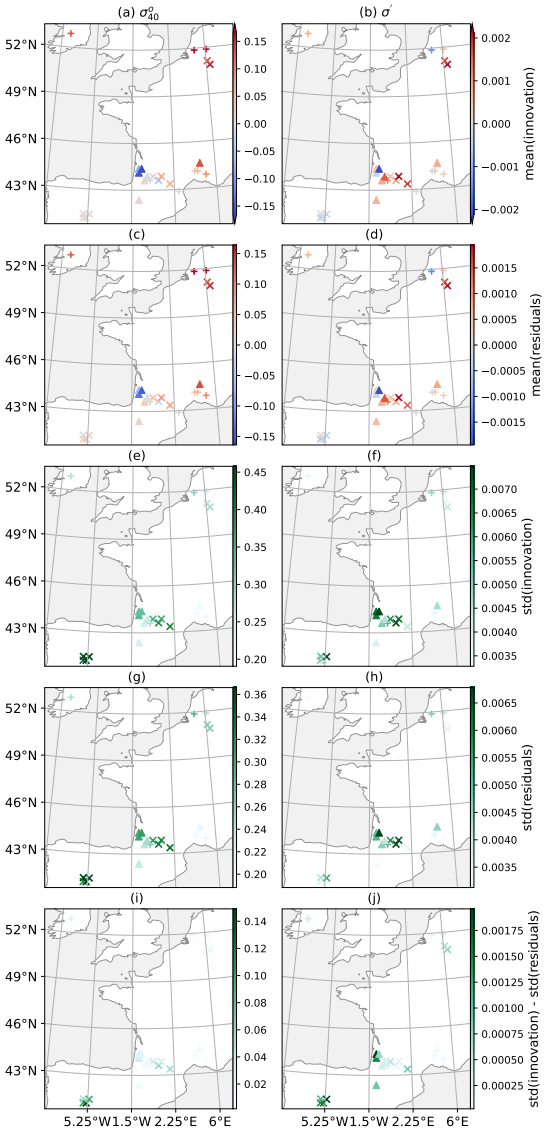

Figure 4.5.: Diagnostics of DA(σ^o_{40}, σ') from 2017 to 2019. Different rows show mean values of innovations and residuals, standard deviations of innovations and residuals, as well as the difference between std of innovations and std of residuals. Different rows show different observations (σ^o_{40} and σ'). Agricultural GPIs are marked as 'x', grasslands GPIs as '+', and other GPIs as triangles.

from the forecast to the true observations. Values range from 0.75 to 0.99 for σ^o_{40} and 0.58 to 0.99 for σ'. Overall, the median values across GPIs of $|\frac{O-A}{O-F}|$ are 0.92 for σ^o_{40} and 0.86 for σ'. Values of the metric are smaller than 1, indicating that the data assimilation is effective in improving the observation equivalents. However, the values are relatively close to unity. This suggests that the analysis assigns similar weights to the observations and ISBA model [197]. The metric varies across the GPIs. In particular, the ratios are generally smaller for agricultural GPIs compared to grassland GPIs for σ^o_{40} and σ' as shown in Table 4.3. This suggests that DA is much more effective in agricultural GPIs. This may be due to the more dynamic soil moisture and LAI cycles of crops compared to grass, reflected by the larger standard deviations of σ^o_{40} and σ' shown in Figure A.9 (c) and (d).

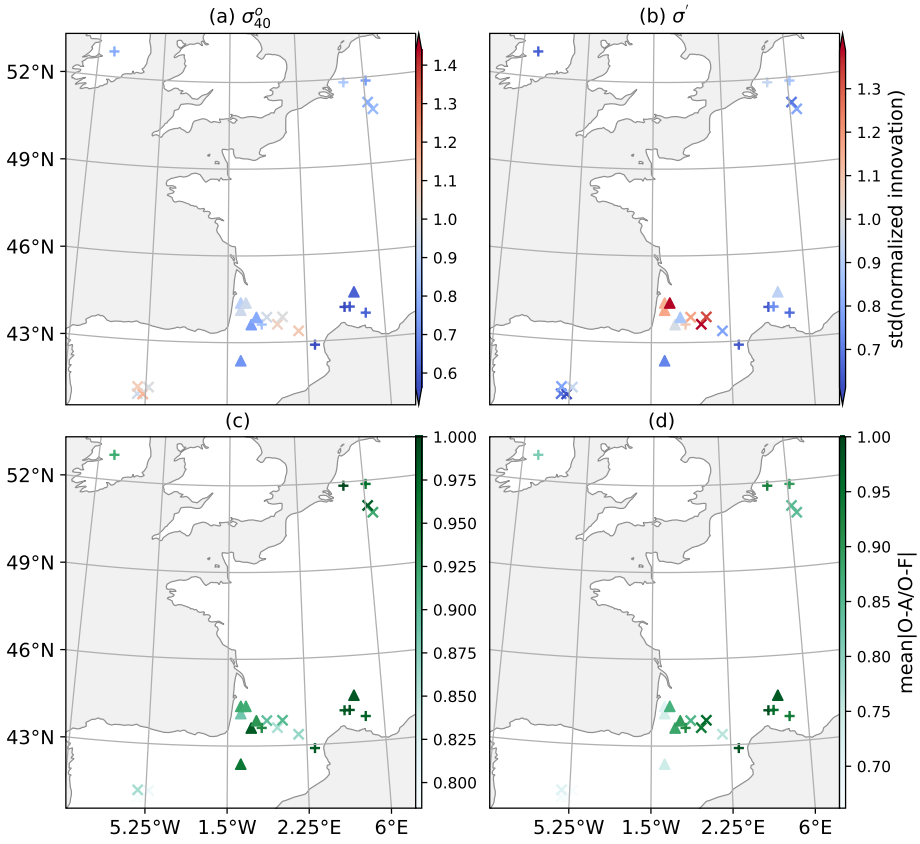

Figure 4.6.: Diagnostics of DA(σ^o_{40}, σ') from 2017 to 2019. Different rows show the standard deviation of normalized innovations and the time series mean values of $|\frac{O-A}{O-F}|$. Different rows show different observations (σ^o_{40} and σ'). Agricultural GPIs are marked as 'x', grasslands GPIs as '+', and other GPIs as triangles.

Table 4.3.: Statistics of diagnostics of DA averaged across all GPIs, grassland GPIs, and Agriculture GPIs.

	All GPIs		Grasslands		Agriculture	
	median	std	median	std	median	std
σ^o_{40} [dB]						
mean(innovations)	0.018	0.097	0.073	0.087	0.02	0.066
mean(residuals)	0.031	0.09	0.076	0.086	0.038	0.059
std(innovations)	0.29	0.092	0.23	0.048	0.39	0.077
std(residuals)	0.27	0.061	0.23	0.043	0.33	0.042
std(norm innovations)	0.83	0.18	0.68	0.13	1	0.12
mean$\lvert O-A/O-F\rvert$	0.92	0.07	0.98	0.026	0.86	0.07
C^k_{k+1} of innovations	0.37	0.067	0.33	0.054	0.36	0.086
σ' [dB/deg]						
mean(innovations)	0.00055	0.0011	0.00039	0.00078	0.00079	0.0012
mean(residuals)	0.00041	0.001	0.00035	0.00074	0.00065	0.0011
std(innovations)	0.0047	0.0014	0.0037	0.0009	0.0053	0.0013
std(residuals)	0.0041	0.0012	0.0035	0.0008	0.004	0.0015
std(norm innovations)	0.86	0.26	0.7	0.17	0.78	0.3
mean$\lvert O-A/O-F\rvert$	0.86	0.11	0.93	0.057	0.8	0.13
C^k_{k+1} of of innovations	0.81	0.17	0.89	0.11	0.64	0.21

OPTIMALITY OF THE DA SYSTEM

Figure 4.7 shows autocorrelations of innovations for σ^o_{40} and σ' at lags from 1 to 10 days. Autocorrelation values are significantly greater than 0 at the 5% significance level, though they decrease as the lag increases. Overall, the large autocorrelation values suggest that the innovations are not white noise. This suggests the suboptimal performance of the DA system [198], so that the DA system is not making optimal use of the ASCAT observations. The lagged autocorrelation values are larger on grassland GPIs than on other GPIs. This suggests, consistent with the previous paragraph, that the DA system is more suboptimal for grasslands than other cover types. The autocorrelation for a lag of 1 day (C^k_{k+1}) depends on the deviations between the optimal weight $\tilde{\mathcal{K}}$ and the estimated weight $\tilde{\mathbf{K}}$ [198]. The statistics of C^k_{k+1} are also reported in Table 4.3. The median value across GPIs of C^k_{k+1} is large with the value of 0.37 for σ^o_{40} and 0.83 for σ'. This shows that the system is more suboptimal for σ' compared to σ^o_{40}. In our DA system, both model and observation errors deviate from the actual errors as shown in Figure 4.6. The overestimation of the term $\tilde{\mathbf{J}}_k\tilde{\mathbf{B}}\tilde{\mathbf{J}}_k^T+\mathbf{R}$ leads to positive C^n_{n+1} values in Figure 4.7. This is one of the reasons why the DA system is suboptimal. However, particular attention needs to be paid to the interpretation of innovation correlation patterns in our DA system. The nonlinearity of the DNN should be considered, as it might contribute to the uncertainty during the assimilation process. The uncertainty associated with DNN's ability to capture the relationship between ASCAT observations and model states

might lead to the suboptimality of the DA system through their impact on equation (25) in [202]. Additionally, σ' is estimated by combining all local slope measurements using an Epanechnikov kernel with a half-width of 42 days, which may introduce autocorrelation of the innovation of σ' [4, 61].

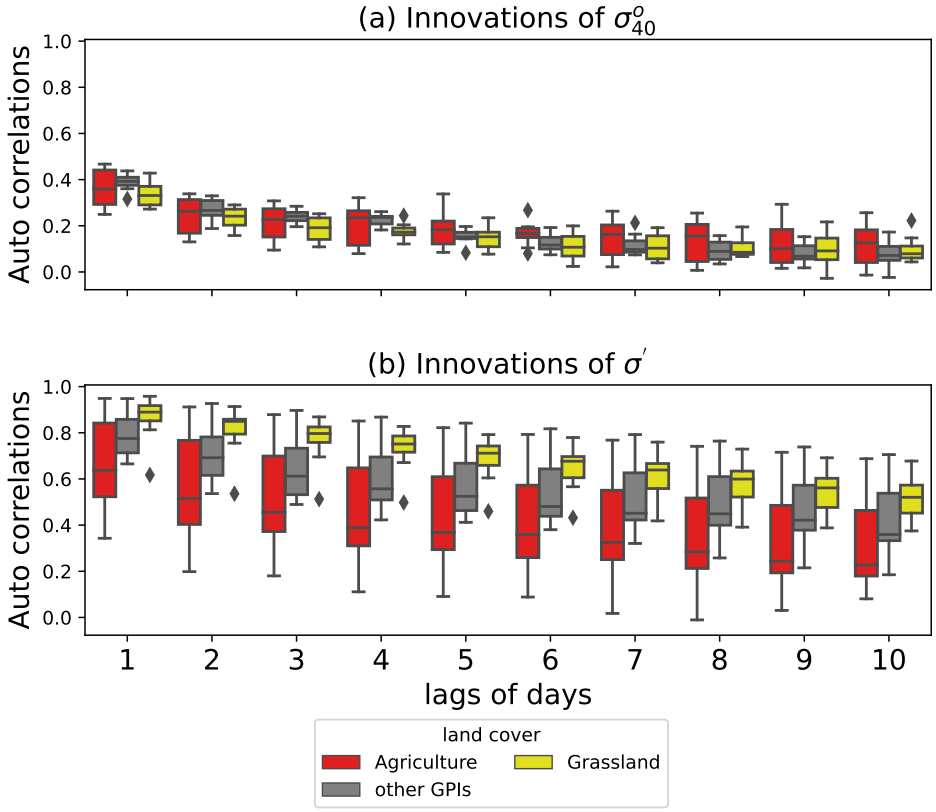

Figure 4.7.: Boxplots of autocorrelation of the innovations of (a) σ^o_{40} and (b) σ' at lags of different days. Agricultural GPIs are marked as red, grasslands GPIs as yellow and other GPIs as grey. The box extends from the lower to upper quartile values of the data, with a line at the median. The lower whisker is at the lowest datum above Q1–1.5*(Q3-Q1), and the upper whisker is at the highest datum below Q3 + 1.5*(Q3-Q1), where Q1 and Q3 are the first and third quartiles. Outliers are indicated as grey dots.

4.4. DISCUSSION

The limited effect of DA may be connected to inconsistency between the Jacobians relating the ASCAT observables to soil moisture in the different layers. The soil moisture in the various layers is highly correlated, so it is reasonable to expect that

the Jacobians (or the NSCs) for the different layers would be similar, at least in sign. Figure 4.8 shows the NSC of ASCAT σ^o_{40} and σ' to RZSM (WG4&WG5) and soil moisture in the other layers. NSC(σ^o_{40}, WG3) exhibits the expected behaviour, i.e. a positive value in winter. In summer, NSC(σ^o_{40}, WG3) fluctuates around zero, suggesting σ^o_{40} loses sensitivity to WG3 because of the presence of vegetation. However, this is not the case for the soil moisture in deeper layers. NSC(σ^o_{40}, WG6) oscillates more than the NSCs for shallower layers during the fall and winter in 2017 and the fall in 2019. This is not plausible because WG6 is damped and delayed with respect to WG3-WG5. This suggests that the DNN may arbitrarily choose among the soil moisture layers during the training process to fit the observed σ^o_{40}.

From Figure 4.3 it is clear that the soil moisture series in adjacent layers (WG6, WG7, and WG8) are highly-correlated. Meanwhile, Figure 4.8 shows that the normalized sensitivity coefficients for each observation (σ^o_{40} or σ') with respect to WG5, WG6 and WG7 vary considerably and often in opposite directions. These layers are sufficiently deep that their soil moisture content should not have any influence on ASCAT σ^o_{40}. However, Figure 4.8 shows that the DNN-estimated σ^o_{40} varies with WG7, and to a lesser degree with WG6. Collectively, these figures suggest that the DNN is overfitting using WG6 and WG7. Therefore, the DNN has been trained to use WG6 and WG7 to adjust σ^o_{40}, resulting in non-zero Jacobian terms for σ^o_{40} and the deep soil moisture states (WG6 and WG7). When these Jacobians are subsequently used in the SEKF, they introduce spurious sensitivity to deep soil moisture and therefore poor DA performance.

The divergent sensitivities have a direct impact on the robustness of the Jacobian matrix (J) in the SEKF update equation 4.4. Figure 4.9 shows the J terms aggregated over patches in equation (7) from 2017 to 2018. Note that here J represents the product of the linearization of ISBA and the Jacobian of the DNN. It is noticeable that the J(σ^o_{40}, WGs) and J(σ', WGs) diverge. The J(σ^o_{40}, WGs) values are positive in spring and become negative in summer and autumn. For σ', J is quite consistent for SSM (WG3) and RZSM (WG4&WG5) but diverges for WG6, WG7 and WG8. The magnitude of the J matrix of WG6, WG7, and WG8 are the largest among the soil moisture. This partly explains why increments on WG6, WG7, and WG8 are so large compared to those in the layers closer to the surface.

Results from Figure 4.6 suggested that the (total) observation and model errors were overestimated, something which would limit the efficacy of assimilation. Rather than reducing the prescribed model errors, which are consistent with those in previous SURFEX studies, we reduced the observation errors assumed for the ASCAT observables to investigate if this rendered the DA system more optimal. For the same four agricultural GPIs considered in Figure 4.3, the DA experiments were repeated assuming observation errors of 0.15 dB for σ^o_{40} and 0.002 dB/deg for σ'. The performance metrics and DA diagnostics are reported in Table A.7 and A.8. The mean values of $|\frac{O-A}{O-F}|$ are reduced, indicating that the DA system tends to follow the observation more than the LSM. The C^k_{k+1} values of innovations are also reduced. So, reducing the observation errors results in a less sub-optimal DA system. However, while the magnitudes of the NIC values increase, the sign of the NIC values does not change in most cases. GPIs at which DA improved state estimates show a

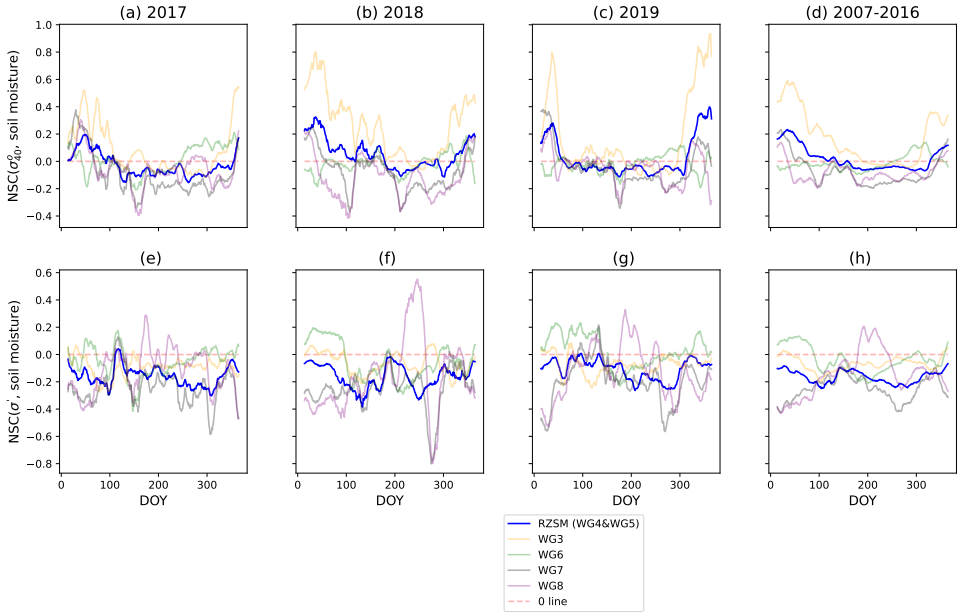

Figure 4.8.: Time series plot of NSC of ASCAT σ^o_{40} and σ' to soil moisture in each layer in independent validation years (2017, 2018 and 2019) and training years (2007-2016) averaged on agri GPIs containing ISMN stations: Saint Felix de Lauragais, Lahas, Condom and Savenes in SMOSMANIA network).

stronger improvement. However, GPIs where DA proved detrimental have a worse performance when the observation error is reduced. Despite the tighter constraints imposed by the observations, the sign of the increments did not change. Instead, it appears that the Jacobians of the observation operator, which establish the link between the land surface variables and the ASCAT observations, had a larger impact on the DA system. While the observation error controls the magnitude of the increments, the Jacobian terms determine the sign of the increments, as shown in Figures 4.8 and 4.9.

Shan *et al.* [124] showed that the DNN could provide excellent predictions of the ASCAT observables. Also, when averaged in time and across many GPIs [124], the NSCs were physically plausible and provided insight into the drivers of dynamics in the ASCAT observables. However, results in Figure 4.9 here show that for a single GPI, at a single time step, the robustness and physical plausibility of the Jacobian may be limited. One key difference with respect to other studies is that we train a DNN for each GPI to account for mixing within the ASCAT footprint, rather than training a generic DNN per cover type. This reduces the amount of training data which, in turn, influences the robustness. Another potential difference is the

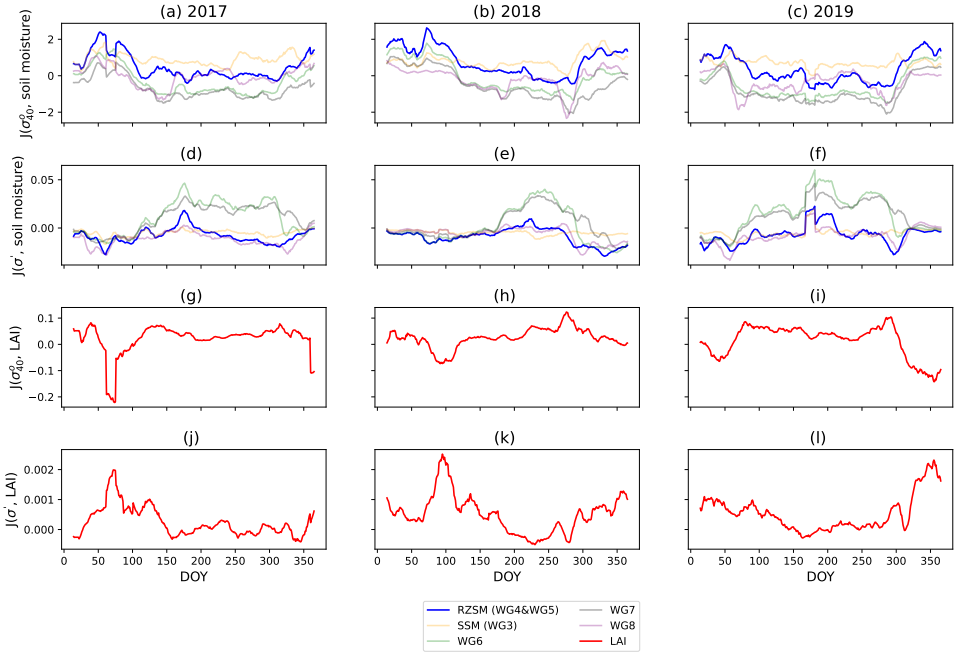

Figure 4.9.: Time series plot of aggregated J matrix in equation (7) of ASCAT σ^o_{40} and σ' to soil moisture in each layer in 2017, 2018 and 2019 averaged on agri GPIs containing ISMN stations: Saint Felix de Lauragais, Lahas, Condom and Savenes in SMOSMANIA network).

complexity of the DNN, specifically the size of the input and output sets. Corchia *et al.* [138] predict backscatter only, and account for the vegetation effects a priori by training the neural network with LAI from satellite observations. In contrast, we train the DNN to predict both backscatter and slope simultaneously and provide a large set of LSVs from the ISBA as input. These methodological differences may increase the complexity of the DNN, influencing the robustness of the Jacobians and hence the efficacy of assimilation. The third difference is the less direct relationship between the land surface model states and the observations, particularly between LSM states and ASCAT slope. While these indirect relationships are not problematic for predicting the observations, the absence of the expected improvement over the OL in our DA results suggests that they may be problematic in the DA framework. This study is therefore not a comparison of DA approaches. It is a study on the implications of including an observation operator based on machine learning in a DA framework. This is important due to the growing interest in using data-driven methods in data assimilation [9, 138, 140, 185, 209].

Results presented here suggest that successful use of ML-derived observation operators in assimilation hinges not just on good agreement between predicted and

true observables, but also on the correct and robust estimation of the Jacobians. In cases like ours, where training data are limited, the robustness of the Jacobian should be included as a criteria for selecting the optimal DNN (or other ML) model to map the LSVs to the observables. However, it is not trivial to achieve physical consistency in the sensitivities of the DNN to land surface variables. Future work is needed to examine the consistency of the DNN under different sets of inputs and reduce the uncertainty of the DNN. Additionally, it is recommended to carefully quantify the observation error, particularly that of the slope, for future assimilation.

4.5. CONCLUSIONS

ASCAT normalized backscatter and slope were assimilated into the ISBA LSM to constrain soil water and vegetation dynamics. The ISBA LSM simulates surface fluxes of water, carbon, and energy at the sub-grid (patch) level and aggregates the results to the grid scale. Here, ASCAT grid level observations were assimilated to produce the updated variables for grid points containing ISMN stations. Following Albergel *et al.* [88], Barbu *et al.* [109] and Barbu *et al.* [108], SEKF was used as the DA algorithm. A DNN trained, tested, and validated by Shan *et al.* [124] was used as the observation operator linking model states with the ASCAT observations. The DA and OL were evaluated against ISMN in-situ station-based soil moisture observations and Copernicus LAI observations to test whether assimilating ASCAT observables helps to constrain the estimates of soil moisture and LAI. Results showed that, in general, the DA has a detrimental effect on the domain median values of ubRMSE and ρ of LAI and neutral effects on SSM (WG3) and RZSM (WG4&WG5). For deeper layers (WG6, WG7, and WG8), large updates are introduced by the assimilation. This is due to the large magnitudes of the Jacobians relating the ASCAT observables to soil moisture in deep layers.

Diagnostics were also analyzed to assess the Gaussianity of innovations and the optimality of the DA system, as well as the prescribed observation and model error. The time series mean of innovations of σ^o_{40} and σ' are within a reasonable range and are small, showing that our DA system reduces the uncertainties of the states in a long time period. The reduction in the standard deviation of residuals compared to the innovations demonstrates the decrease in uncertainties following the assimilation of ASCAT σ^o_{40} and σ' into the DA system. The analysis of normalized innovations and $|O-A/O-F|$ indicated that the model and observation errors are overestimated and this misspecification is more problematic in grassland compared to agricultural areas. In other words, our prescribed observation error is more confident in agricultural GPIs. This could be related to the larger dynamic range of the ASCAT observables in agricultural GPIs shown in Figure A.9. Further, large values of lagged autocorrelation of the innovations indicate that the DA system is suboptimal and the information in ASCAT observables is not effectively used. Potential causes include the misspecification of the forecast and observation error, and the consistency of DNN sensitivities and predictions.

We investigated potential explanations for the limited positive impact of the DA. A second DA experiment was conducted with lower observation errors on agricultural

GPIs. The comparison showed that the DA experiment with reduced observation errors does not perform better than the prescribed observation error, though the reduced observation errors reduce the suboptimality of the DA system. This indicated that the inconsistent sensitivities of the Deep Neural Network (DNN) used in the observation operator may have a greater influence on the updates than the model and observation errors themselves.

With the growing utilization of machine-learning-based observation operators in land data assimilation systems, this study underscores the importance of addressing the robustness of machine learning (ML) methods. A key result of this study is that successful ML training does not ensure successful DA, as the terms of the Jacobian may include spurious effects due to overfitting if too many predictors (inputs) are used in the DNN. Therefore, the robustness of ML-based methods plays a vital role in ensuring the physical plausibility of land-atmosphere mechanisms learned by the data-driven approaches, particularly in terms of the sensitivities of remote sensing observables to land surface variables. Furthermore, this is essential to ensure the transferability of machine learning algorithms as observation operators for assimilating similar remote sensing data into LSMs.

5

EVALUATING THE ML-BASED OBSERVATION OPERATOR IN LAND DATA ASSIMILATION FRAMEWORK

Based on: X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. *Evaluating DNN-based observation operators in land data assimilation: Synthetic experiments with land surface model ISBA-A-gs in Southwestern France.* in preparation. 2024

5.1. INTRODUCTION

The previous chapter assimilated ASCAT normalized backscatter and slope into the ISBA LSM to constrain the soil water and vegetation dynamics. The DNN trained and validated in Chapter 3 is used as the observation operator. Evaluation against ISMN in-situ soil moisture observations and CGLS LAI observations showed that, in general, the DA has a neutral effect on the ubRMSE and ρ of SSM and LAI. The potential reasons for the limited impact of the DA include representation errors and misspecification errors. A second DA experiment with lower observation errors showed that the inconsistency among the terms of the Jacobian matrix of the DNN might have a larger influence on the updates.

This chapter extends the work of Chapter 4 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. The radiative transfer model WCM is assumed to be the perfect observation operator, accurately representing the true scattering mechanism in the synthetic experiment. Different simulations of LSM ISBA are run to generate synthetic truth and open loop (OL) states of land surface variables. DNNs are trained using ISBA OL as input to estimate the synthetic backscatter. Different DNNs are trained with different input choices, representing the inaccurate input choice in the real physical world. Specifically, we hypothesize that assimilating normalized backscatter using DNNs as the observation operator leads to suboptimal performances because a) DNN estimations, designed to fit the training target (synthetic backscatter), are less sensitive to the information about the difference between input OL and synthetic truth, which is important for subsequent DA; b) the presence of redundant input variables or high collinearity makes inaccurate estimation of the DNN-based Jacobian.

5.2. DATA

5.2.1. ASCAT NORMALIZED BACKSCATTER

ASCAT is on board the Metop series of satellites operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) [4]. 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 at $34° \sim 65°$ and $22° \sim 55°$ [4]. 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 [4, 44, 61]. The ASCAT normalized backscatter is available at 0.25 degree resolution. The normalized backscatter dataset consists of data of ascending (09:30 A.M.) and descending orbits (09:30 P.M.).

5.2.2. STUDY AREA

The experiment is conducted over a region in southwestern France as shown in Figure 5.1 (a), extending from 43°N to 46°N latitude and 1°W to 3.5°E longitude. The domain contains 209 grid points (GPIs), with each GPI representing an area of

0.25° by 0.25°. Figure 5.1 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.

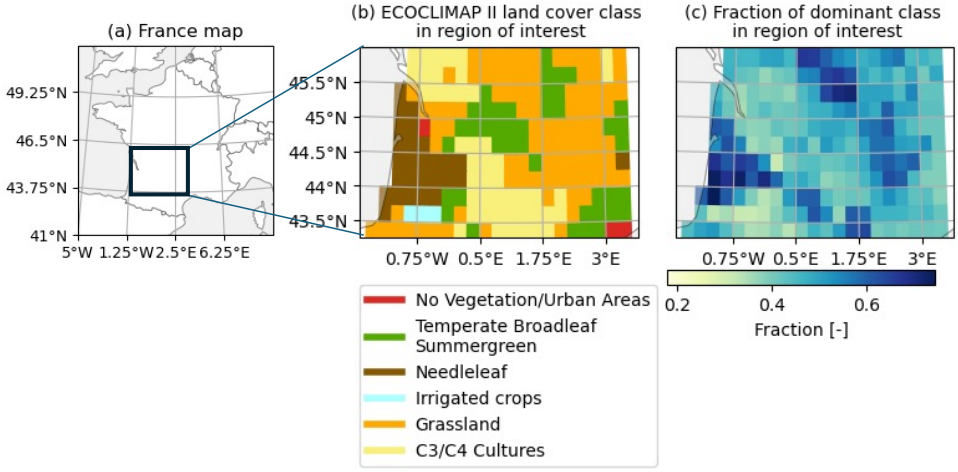

Figure 5.1.: The map of France (a) and the main land cover types and fraction of GPIs in study area based on ECOCLIMAP II. (b) the dominant vegetation types of the 0.25° x 0.25° GPIs in region of interest (southwestern France); (c) the vegetation fraction of the dominant vegetation types of each GPI.

5.3. METHODOLOGY

5.3.1. WATER CLOUD MODEL

In the Water Cloud Model (WCM), the backscatter from a vegetated soil surface is represented by the sum of vegetation backscatter (σ^o_{veg}) and soil backscatter (σ^o_{soil}) which undergoes two-way attenuation due to the vegetation layer:

$$\sigma^o = \sigma^o_{\text{veg}} + \Gamma^2 \sigma^o_{\text{soil}} \tag{5.1}$$

with

$$\sigma^o_{\text{soil(dB)}} = 10\log_{10}\sigma^o_{\text{soil}} = C + D \times \text{SSM}. \tag{5.2}$$

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

$$\Gamma = \exp(-\frac{\text{VOD}}{\cos\theta}). \tag{5.3}$$

The vegetation is represented as a layer with uniformly distributed water droplets [211], characterized by a bulk vegetation descriptor V_1:

$$\sigma^o_{\text{veg}} = A \times V_1 \times (1-\Gamma^2) \times \cos\theta, \tag{5.4}$$

with V_1 being the bulk vegetation descriptor accounting for canopy backscatter and A being a fitting parameter [8] and VOD = $B \times V_2$ [127]. 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.* [128].

The WCM equations contain 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 [127]. Parameter A quantifies the effect of direct vegetation scattering while B is the measure of the vegetation attenuation effect [8]. 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 ($dBm^{-3}m^3$) describes the sensitivity of backscatter to soil moisture. A, B, C and D also depend on the radar configuration.

5.3.2. EXPERIMENT SETUP

The detailed experiment setup is illustrated in Figure 5.2. There are four steps in total: (1) calibration of the WCM model, (2) generation of synthetic truth and observations, (3) DNN training and validation, and (4) 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 gen using the calibrated WCM from step (1). Noise is then added to these WCM outputs to account for observation error, 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 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.

CALIBRATION OF WCM MODEL

The parameters A, B, C, and D of the WCM are determined by calibrating the ISBA OL model simulations against the actual ASCAT normalized backscatter. The model simulations take WG2 and LAI from the ISBA 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) [212, 213]. The objective loss function J minimizes the root mean square error (RMSE) between WCM simulated

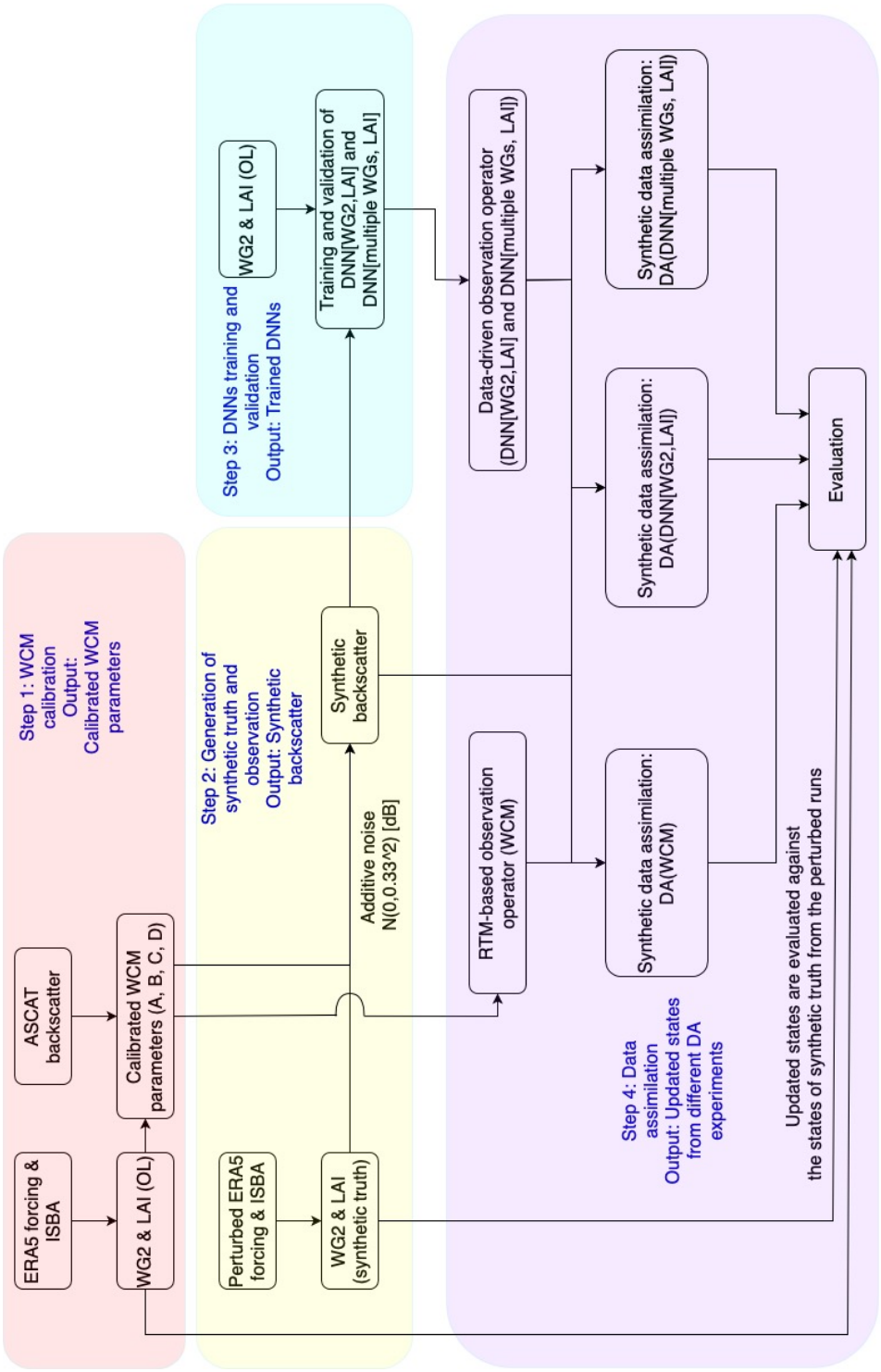

Figure 5.2.: Experiments setup.

backscatter σ^o_{WCM} and ASCAT backscatter σ^o_{ASCAT} at a given GPI. J is defined as:

$$J = \sqrt{\frac{1}{N}\sum_{i}^{N}(\sigma^o_{\mathrm{WCM}} - \sigma^o_{\mathrm{ASCAT}})^2 + W_\alpha \frac{1}{N_\alpha}\sum_{i}^{N_\alpha}\frac{(\alpha_{0,i} - \alpha_i)^2}{\sigma^2_{\alpha_{0,i}}}}, \tag{5.5}$$

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 *et al.* [8], 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 *et al.* [8] and Shamambo *et al.* [127].

GENERATION OF SYNTHETIC TRUTH AND OBSERVATIONS

A perturbed run of the LSM ISBA serves as the “synthetic true" solution and is meant to represent nature [110]. We start from the spin-up of the initial condition on 1 January 1996. Then we run a “perturbed" run from 1996 to 2019. Following Lannoy *et al.* [95] and Reichle *et al.* [110], perturbations are introduced to 1) ERA5 Forcing data to represent our imperfect knowledge about the forcing data [110]; 2) the prognostic state variables like soil moisture in different layers, to mimic the errors in model structure and model parameters [95].

The precipitation (Pcp) and short wave radiation (SW) are perturbed with multiplicative (denoted as “M") lognormal perturbations, and long wave radiation (LW) receives additive (denoted as “A") normal perturbations, as listed in Table 5.1. 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 $\lambda(w_{fc} - w_{wilt})$ 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 [214]. These values are in line with [74, 214].

Synthetic observations of normalized backscatter are generated by taking the WG2 and LAI from the synthetic truth and running the WCM model with the calibrated parameters:

$$\sigma^o_{\mathrm{true}} = \mathrm{WCM}(A, B, C, D, WG2_{\mathrm{true}}, LAI_{\mathrm{true}}) + \epsilon_{\mathrm{true}} \tag{5.6}$$

Observation error is represented through ϵ_{true}, an additive Gaussian noise with zero mean and a standard deviation of 0.33 dB following Lievens *et al.* [8].

DNN TRAINING AND VALIDATION

Two DNNs are trained following the methodology described by Shan *et al.* [124]. The input data of DNN comes from the OL experiment of ISBA from 2007 to 2019.

Table 5.1.: Perturbation applied on the ERA5 forcing data, following [95, 110]

Variables	Type	Std dev	Temporal correlation	Cross correlation with perturbations in white noise		
				Pcp	SW	LW
Pcp	M	0.5	24h	1	-0.8	0.5
Downward SW	M	0.3	24h	-0.8	1	0.5
Downward LW	A	20 Wm^{-2}	24h	0.5	-0.5	1

Similar to the calibration of WCM, the DNN is trained independently per GPI. The structure of the DNN is tuned by Bayesian optimization [173], including the number of layers, the number of neurons of each layer, batch size, the choice of activation function, and learning rate. The DNNs are trained and tested using data from 2007 to 2016. To improve the robustness of the Jacobian as suggested by Shan *et al.* [196], the data from 2007 to 2016 are randomly shuffled and sampled 50 times. 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 ($k = 1,...,50$). 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 [215] because multiple solutions can be found for similar output statistics [216]. The parameters of DNNs are very unstable and may vary drastically from one solution to another [216]. The spread of Jacobians of multiple DNN solutions (trained on same dataset) is referred to the error of Jacobians [217]. Following Krasnopolsky [217], we created an ensemble of DNNs 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 [217] 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 uncertainty of the Jacobian.

In this study, we consider two DNNs, both of which 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 in other layers are included because of the indirect link between root zone soil moisture and vegetation water content [7, 124] which affects radar backscatter [7, 182]. As a result, normalized backscatter would be indirectly influenced by root zone soil moisture.

The calculation of the Jacobians of DNN models follows the methodology of Shan

et al. [124]. 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.

DATA ASSIMILATION EXPERIMENTS

The synthetic backscatter observations generated in Section 5.3.2 are assimilated to the ISBA using the SEKF to constrain the ISBA model from 2017 to 2019.

Table 5.2 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 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 the observation operator. DA(DNN[WG2,LAI]) is the DNN trained on surface soil moisture (WG2) and LAI. Note that these are 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,...,12 are defined following Mahfouf *et al.* [107], Draper *et al.* [102], and Albergel *et al.* [88]. 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 [142]. In this study, the perturbation applied to the states of soil moisture in the calculation of the Jacobian matrix is assigned as $1 \times 10^{-4} \times (w_{fc} - w_{wilt})$ following Albergel *et al.* [88], and $0.001 \times LAI$ for LAI. For volumetric surface soil moisture, a mean standard deviation (SD) error of 0.04 m^3m^{-3} is prescribed. For soil moisture in deeper layers, SD error of 0.02 m^3m^{-3} is used following Mahfouf *et al.* [107], Draper *et al.* [102], Barbu *et al.* [108], and Barbu *et al.* [109]. The observation error for σ^o_{40} is set as 0.33 dB following Lievens *et al.* [8].

5.4. RESULTS

5.4.1. WCM CALIBRATION

The calibrated values of parameters A, B, C, and D are shown in Figure A.11. High values of A and C are obtained in the urban areas of Toulouse and Bordeaux which is consistent with Shamambo *et al.* [127]. The B parameter exhibits higher values in needleleaf forest regions where vegetation attenuation effects are stronger than those in agricultural areas and grasslands [8]. The distribution of parameter D is similar to those reported by Modanesi *et al.* [128], with a peak at 15 $dBm^{-3}m^3$. Additionally, high values for both parameters C and D are found in the Landes forest. Shamambo *et al.* [127] found the same and attributed this to the sandy soil [218], leading to dominant volume scattering, and the flat topography.

5.4.2. SYNTHETIC TRUTH AND OBSERVATIONS

Figure 5.3 (a) and (b) compares 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 m^3m^{-3} with respect to the OL, even though the perturbations applied to forcing and prognostic states are bias-free. Fairbairn *et al.* [214] and Bonan *et al.* [74] also observed a significant negative bias in the perturbed ISBA 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.* [214] linked this negative bias to the nonlinear behavior of evapotranspiration and drainage [214]. 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 [214]. Fairbairn *et al.* [214] 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 differences between the OL and the perturbed run are largest in autumn.

Figure 5.3 (c) compares the (synthetic) true normalized backscatter (σ^o_{true}) to that simulated using WCM given the OL or synthetic true inputs (from Figure 5.3 (a) and (b)). The influence of the dry bias in WG2 between the OL and synthetic truth is clear in Figure 5.3 (c). Recall that σ^o_{true} is generated by WCM simulations using perturbed WG2 and 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 grey lines is the addition of the assumed observation error.

5.4.3. COMPARISON OF DIFFERENT OBSERVATION OPERATORS

Figure 5.4 shows the simulated σ^o based on different observation operators as a function of WG2 and LAI. The ranges for WG2 (0.10 to 0.25 m^3m^{-3}) and LAI (1 to 3 m^2m^{-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 range as shown in Figure 5.3.

The similarity between Figure 5.4 (b) and (d), 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 5.4 (e). The maximum difference between two DNNs is about 0.35 dB when WG2 reaches 0.15 m^3m^{-3}) while the max difference between DNN and WCM is about -0.7 dB) as shown in Figure 5.4 (c).

The gradients of contour lines in Figure 5.4 (a), (b), and (d) 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 Figure 5.4 (b) and (a) that DNN[WG2, LAI] overestimates the sensitivity of σ^o_{40} to WG2 when LAI is low (around 1.5 m^2m^{-2}). The contour line is denser and the directions are different in Figure 5.4 (b) and (d) than they are in Figure 5.4 (a).

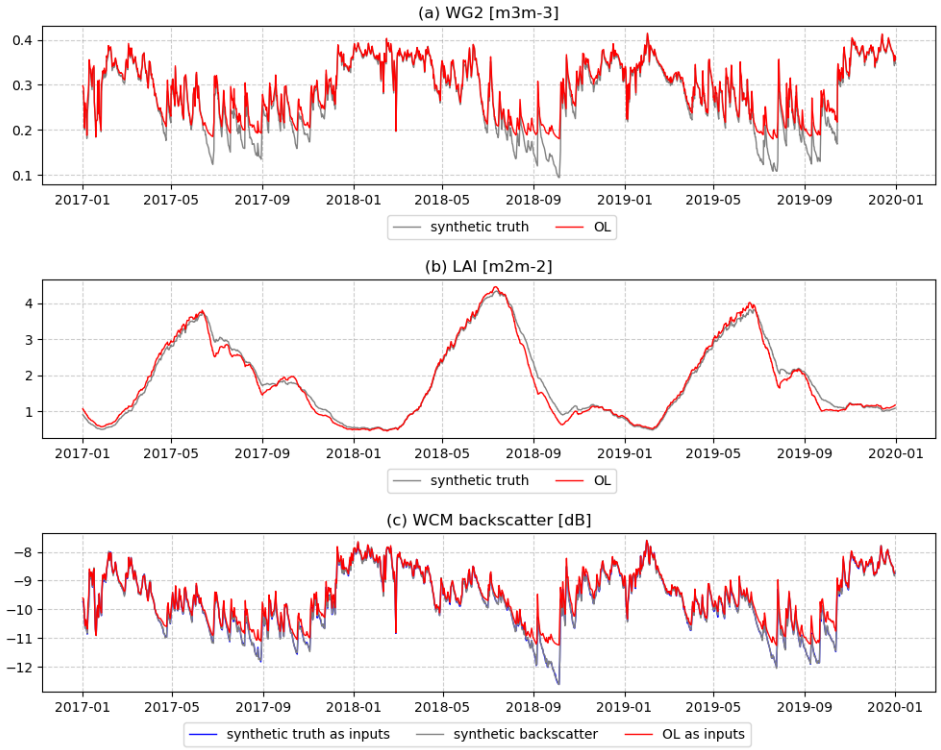

Figure 5.3.: Time series of (a) surface soil moisture (WG2) and (b) LAI from the OL and perturbed run (synthetic truth), (c) observation equivalent (backscatter) estimated using WCM based on inputs from open loop (OL, the red line), synthetic truth of WG2 and LAI (blue line) and synthetic backscatter (σ^o_{true}, dark grey 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.

Further, the gradients of contour lines in Figure 5.4 (c) show the difference in sensitivities in Figure 5.4 (a) and (b). 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.

Fig 5.5 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 performance represents the best estimate of σ^o_{true} with the given WG2 and LAI inputs from the OL.

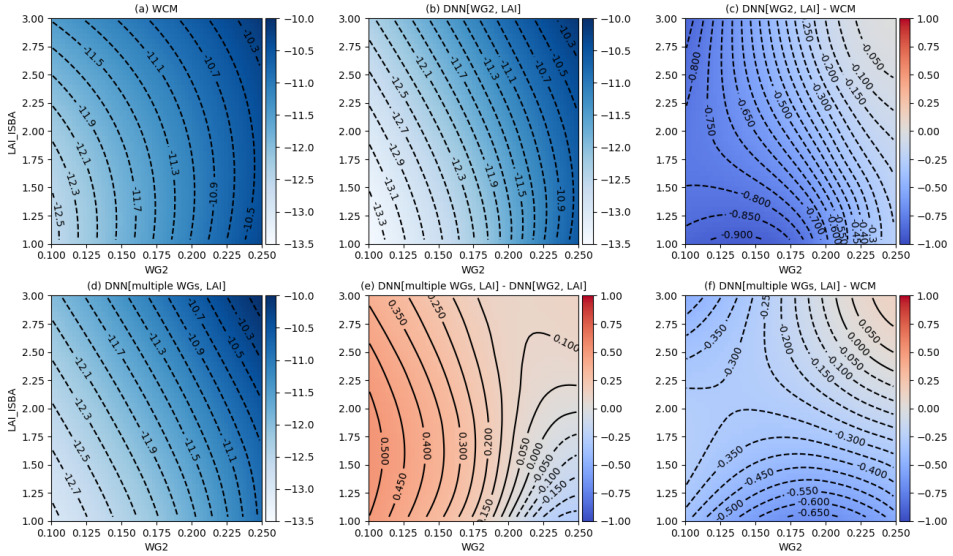

Figure 5.4.: Contour plot of estimated backscattering coefficient (dB) based on different observation operators. The estimate is averaged over all agricultural GPIs. The x and y axes represent the input variables, WG2 and LAI. Results are shown for the three observation operators, in subplots a) WCM, b) DNN[WG2,LAI], and d) DNN[multiple WGs,LAI]. Subplot c), 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.

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 on the forest biomass, as discussed by Corchia *et al.* [138]. Both DNN[WG2,LAI] and DNN[multiple WGs,LAI] have a lower normalized ubRMSE compared to WCM in agricultural GPIs.

Figure 5.6 shows the seasonal variation in the performance of different observation operators taking the OL states as inputs to estimate the synthetic observations. All observation operators exhibit the highest ubRMSE in autumn. Similar to Figure 5.5, 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 5.7 shows that the normalized backscatter simulated with the DNNs with OL as inputs are closer to σ^o_{true} compared to WCM estimation in autumn, using OL as inputs. This is because the DNNs are designed to minimize the loss function between the estimation and σ^o_{true}. 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 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.

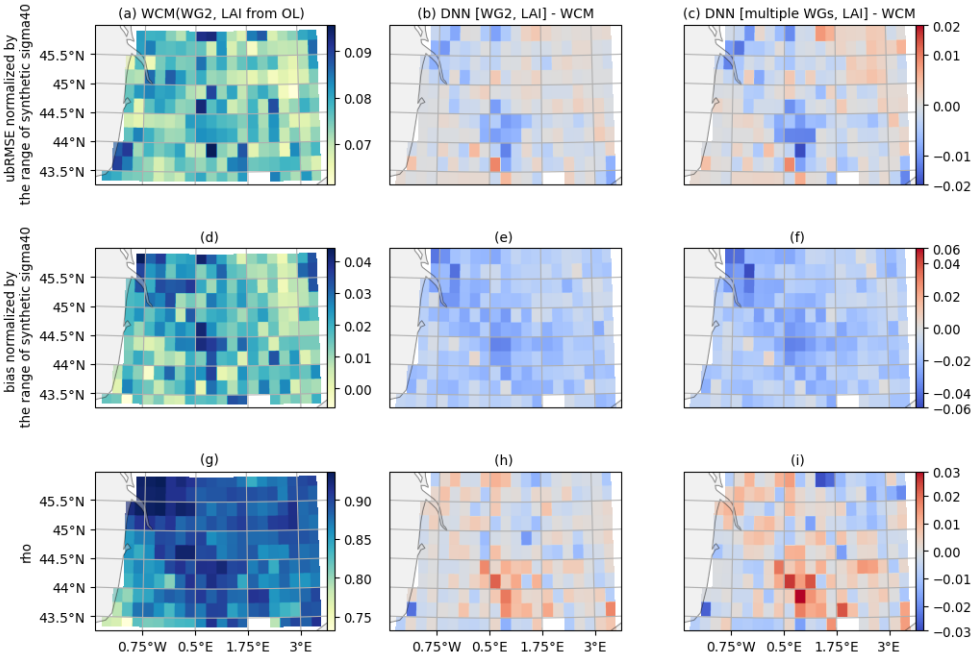

Figure 5.5.: Performance skills of observation equivalents (backscatter) estimated by different observation operators based on WG2 and LAI from open loop (OL) from 2017 to 2019, evaluated against synthetic backscatter. Top, middle, and bottom rows show unbiased RMSE (dB), bias (dB), and correlation coefficient. ubRMSE (dB) and bias (dB) at each individual GPI are normalized by the range of synthetic backscatter (σ^o_{true}). The first column shows the performances of WCM, while the second and third columns show the differences between performances of different DNNs and the WCM.

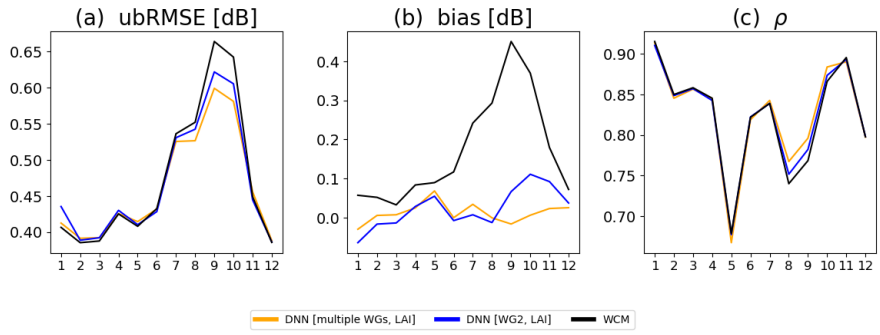

Figure 5.6.: Monthly performance of observation equivalent (backscatter) estimated using different observation operators based on inputs from open loop (OL) from 2017 to 2019, evaluated against synthetic backscatter. The performances are averaged among agricultural GPIs as shown in Figure 5.1.

Table 5.2.: 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

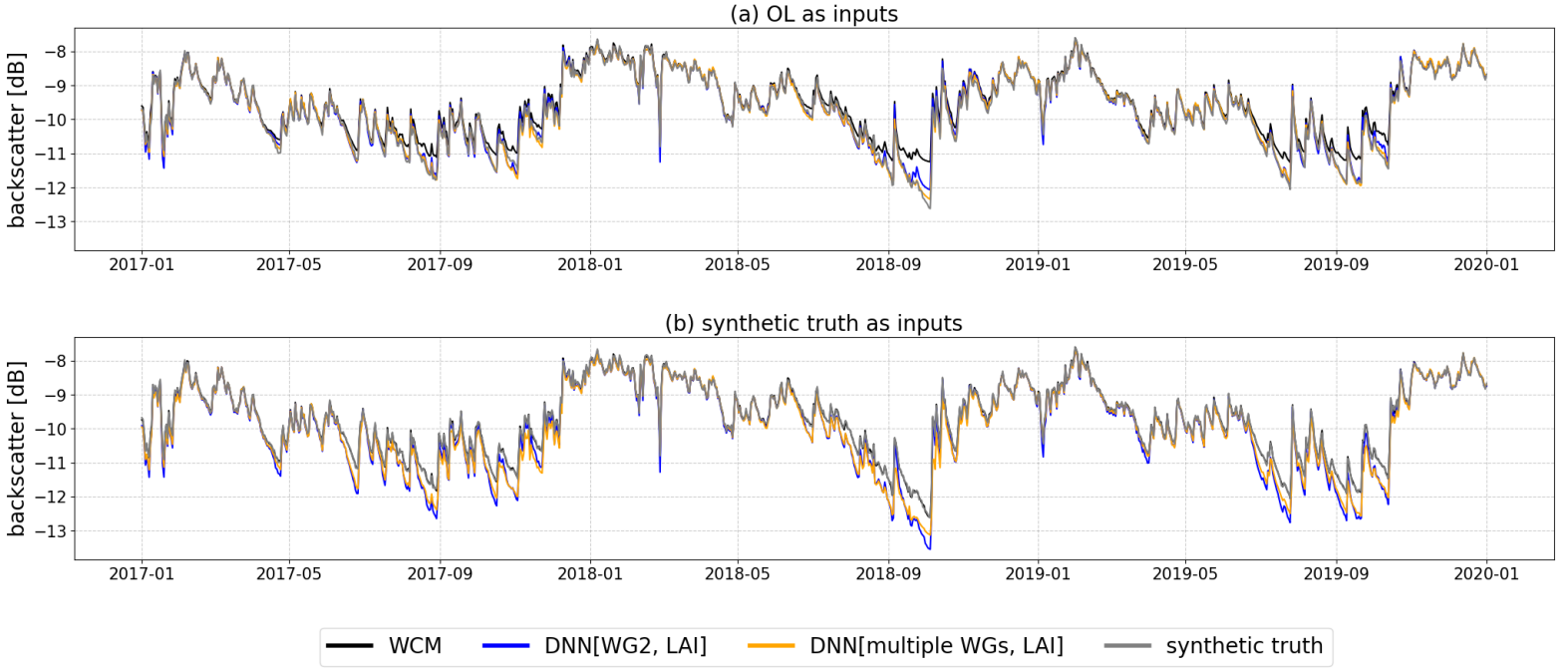

Figure 5.7.: Time series of observation equivalent (backscatter) estimated using different observation operators based on inputs from (a) open loop (OL) (b) synthetic truth of WG2 and LAI (blue line) and synthetic backscatter (σ^o_{true}, dark grey color) from 2017 to 2019. Time series are averaged over agricultural GPIs as shown in Figure 5.1.

5.4.4. JACOBIANS OF DIFFERENT OBSERVATION OPERATORS

Figure 5.8 shows the time series of Jacobians of the different observation operators to WG2 and LAI calculated using the method described by Shan *et al.* [124]. The J(σ^o_{true}, WG2) of WCM represents the "true" sensitivity, i.e. how simulated backscatter is sensitive to WG2 and LAI from the OL. The two DNNs generally reproduce the seasonal cycle of WCM Jacobians. However, there are large biases in J(σ^o_{true}, WG2) of DNN[WG2,LAI] in autumn. DNN[WG2,LAI] overestimates the sensitivity to WG2 compared to WCM which is consistent with Figure 5.4 (c).

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 5.9. The magnitude of the difference between WG2 of the synthetic truth and OL is closely related to the difference in Jacobians between 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 5.8 (a) shows that J(σ^o_{true}, 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(σ^o_{true}, WG2), ... J(σ^o_{true}, WG8) (the green line in the Figure 5.8 (a)) 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(σ^o_{true}, 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(σ^o_{true}, 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 5.3 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 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

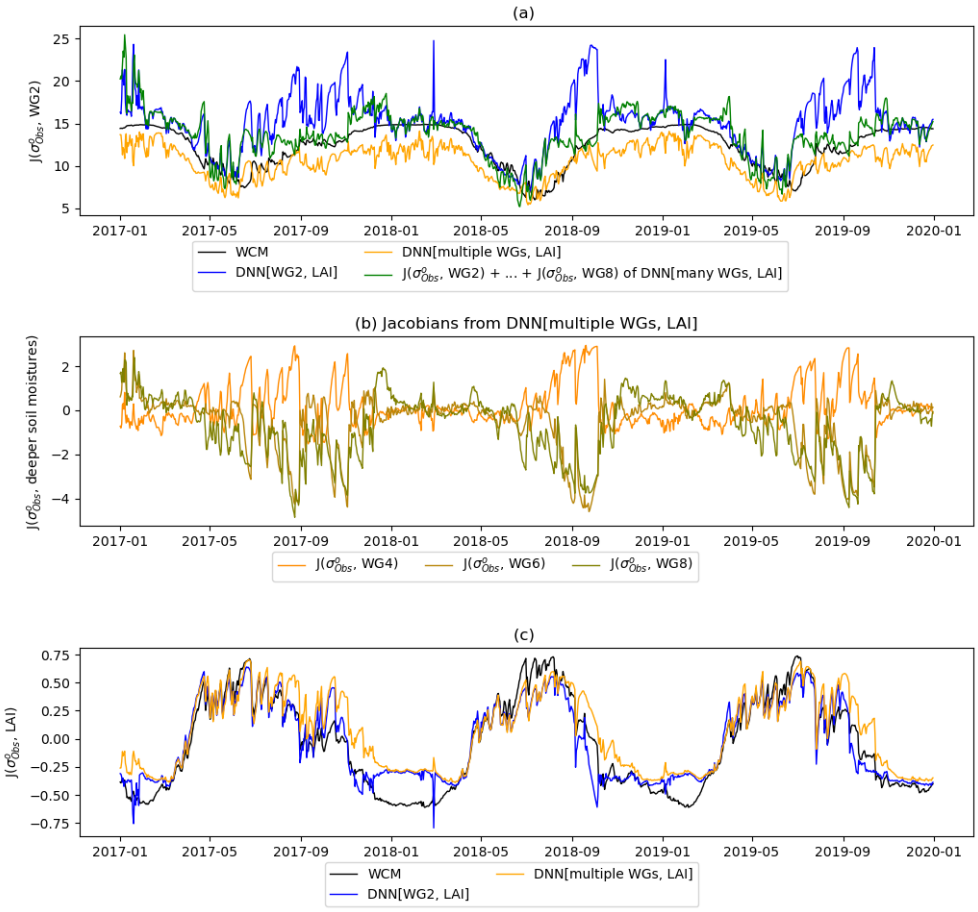

Figure 5.8.: Time series of Jacobians of different observation operators to WG2 and LAI. The Jacobians take the WG2 and LAI from the OL as inputs. The time series are averaged over agricultural GPIs.

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.

5.4.5. DATA ASSIMILATION RESULTS

Figure 5.10 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 A.9. The domain median value of ubRMSE of WG2 from the

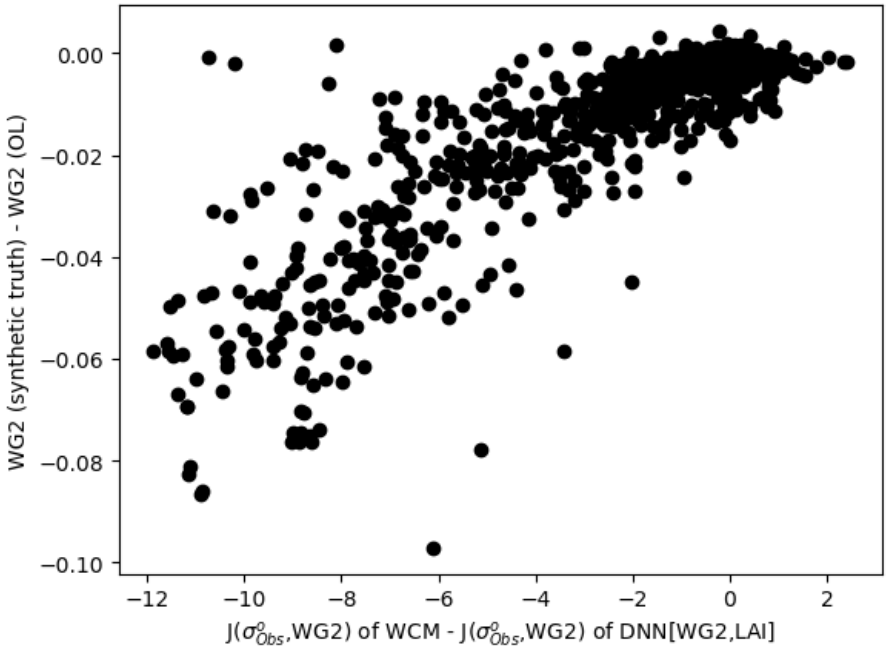

Figure 5.9.: Scatter plot about the differences between the WG2 from synthetic truth and OL (y axis) versus the differences between the J(σ^o_{true}, WG2) of WCM and DNN[WG2,LAI]. The Jacobians take the WG2 and LAI from the OL as inputs. The data are averaged over agricultural GPIs.

Table 5.3.: Details about different model runs, including OL, perturbed run, and different DA experiments

	WCM	DNN[WG2,LAI]	DNN[multiple WGs,LAI]
performances based on OL	-	better than WCM	similar as DNN[WG2,LAI]
J(σ^o_{true}, WG2) based on OL	-	overestimated in autumn	underestimated
J(σ^o_{true}, LAI) based on OL	-	lose sensitivity in winter	lose sensitivity in winter

OL is 0.0265 m^3m^{-3}. WG2 from the OL has a wet bias of 0.0106 m^3m^{-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 [196]. 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 [214]. However, the magnitude reported here is in agreement with the findings presented in the previous synthetic data assimilation experiment Fairbairn *et al.* [214]. 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^3m^{-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^3m^{-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 5.11 (b). 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 5.8.

For monthly performances, Figure 5.11 (d) 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, Figure 5.11 (j) and (m) 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 moisture across various layers—from WG3 to WG8. This allocation leads to large values of components of $\tilde{\mathbf{J}}$ matrix in equation (7) for WG6 and WG8, which results in false updates in WG6 and WG8.

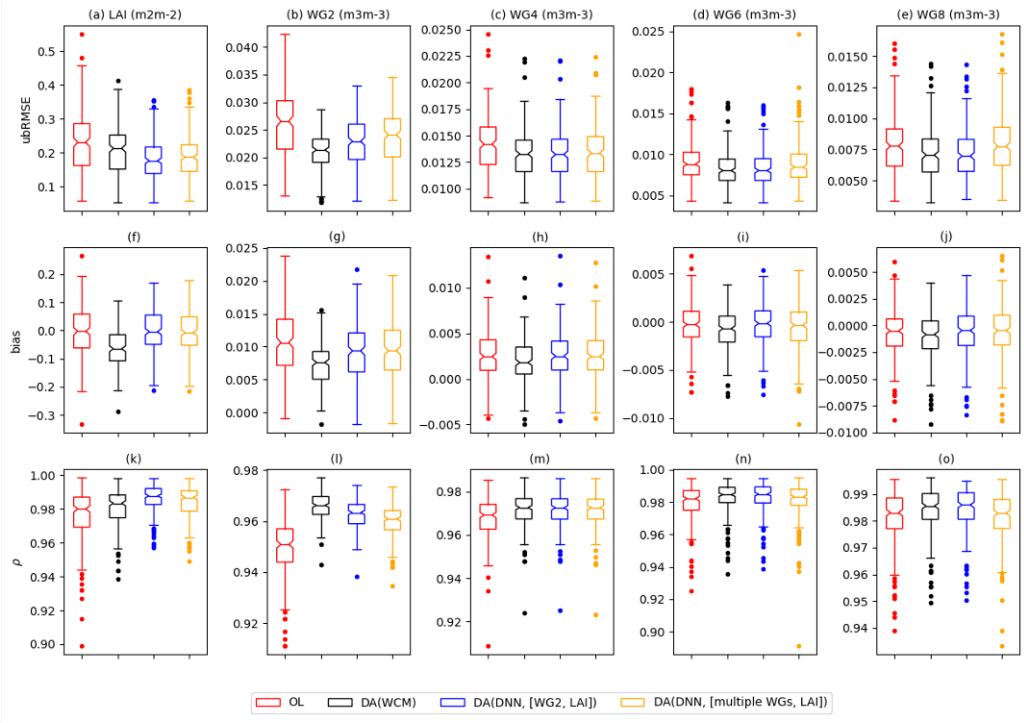

Figure 5.10.: Performances of land surface states (LAI, WG2, WG4, WG6, WG8) from different experiments like OL, DA(WCM), DA(DNN[WG2,LAI]) and DA(DNN[multiple WGs, LAI]) DA(WCM).

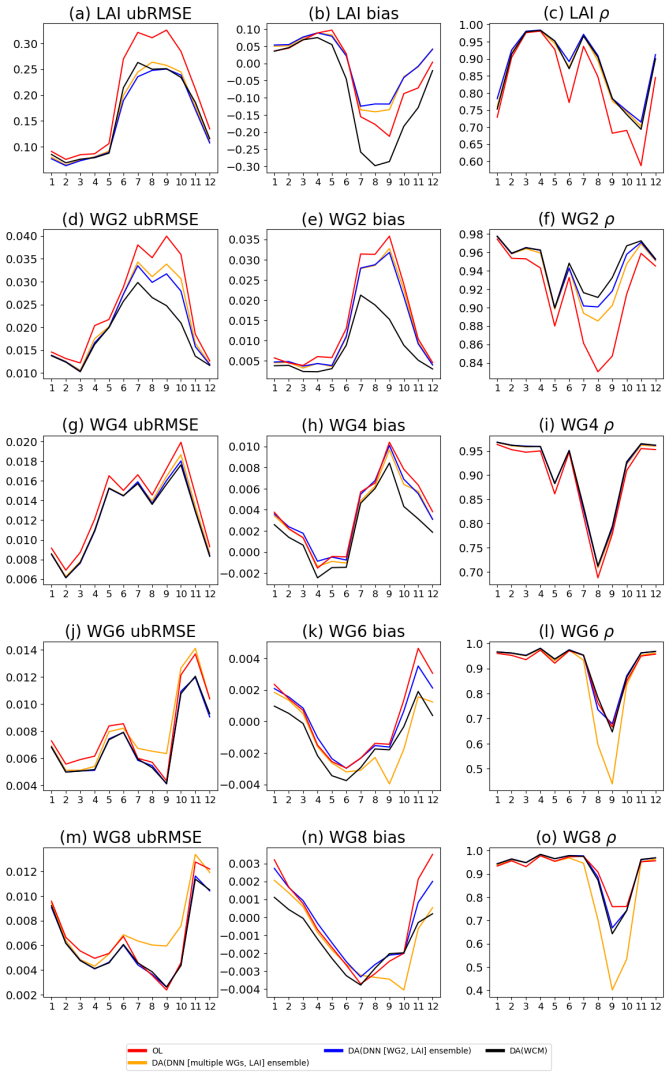

Figure 5.11.: Monthly performances of land surface states (LAI, WG2, WG4, WG6, WG8) from different experiments like OL, DA(WCM), DA(DNN[WG2,LAI]) and DA(DNN[multiple WGs, LAI]) DA(WCM). Data are averaged among agricultural GPIs.

Figure 5.12 shows the time series of WG2 from DA(DNN[WG2,LAI]) and DA(DNN[multiple WGs,LAI]) on 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 5.12 (b). 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 5.12 (d) 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 $\tilde{\mathbf{J}}$ to WG2 in equation (7) (as illustrated in Figure 5.12 (c)). The $\tilde{\mathbf{J}}$ matrix quantifies the sensitivity of observation equivalent to the state in the forecast. Larger absolute values of $\tilde{\mathbf{J}}$ indicate a larger sensitivity. The DA(DNN[multiple WGs,LAI]) has a smaller value of $\tilde{\mathbf{J}}$ because the DNN[multiple WGs,LAI] is overfitting and allocates the sensitivity to soil moisture in multiple layers (suggested in Figure 5.8). This leads to smaller updates of WG2 from DA(DNN[multiple WGs,LAI]).

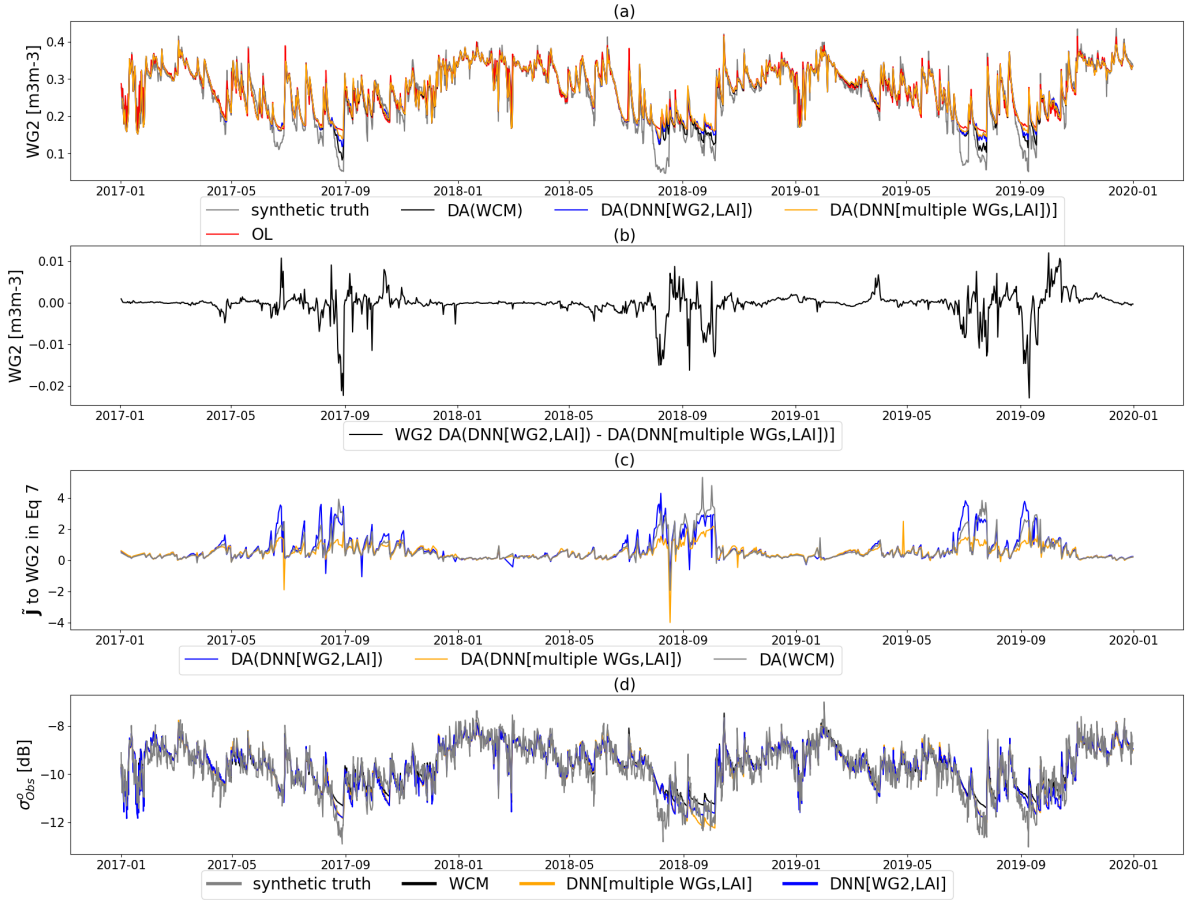

Figure 5.12.: Time series on one GPI (lat 43.875, lon 2.125) of a) WG2 from OL, synthetic truth, DA(DNN[WG2,LAI]), and DA(DNN[multiple WGs,LAI]); b) the difference of WG2 between DA(DNN[WG2,LAI]) and DA(DNN[multiple WGs,LAI]); c) the $\tilde{\mathbf{J}}$ to WG2 in equation (7) aggregated on patches; d) the observation equivalent of different observation operators based on OL.

5.5. DISCUSSION

The comparison between DNN[WG2,LAI] and WCM shows that even with the correct choice of inputs, a data-driven model trained on OL data might be overfitting. 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 data-driven 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 comparison between DNN[WG2,LAI] and WCM highlights the need to train data-driven observation operators with inputs as close to the truth as possible. One unavoidable challenge is that the real truth is unknown.

The motivation to use the states from the OL as training input [140, 185] is to minimize the potential bias in the climatology between the model and the predicted observations.

One solution would be to train the ML-based observation operator on satellite data, similar to [138]'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.

An alternative solution might be to train the DNNs on an ensemble of LSM runs. The ensemble 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").

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 data-driven models; the Jacobian is also important. As discussed in Section 5.4.4, 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 the physical understanding. These challenges may limit the applicability of machine learning algorithms in data assimilation.

5.6. CONCLUSIONS

There is an increasing trend of using machine learning algorithms in earth sciences [132], especially in earth system modeling and data assimilation community [120, 138, 140]. This study compares three synthetic data assimilation experiments using different observation operators. Perturbations are applied to the forcing variables and prognostic states of LSM ISBA to produce synthetic truth (WG2, LAI). Then WCM

takes perturbed WG2 and LAI (synthetic truth) as input to simulate the synthetic observation of normalized backscatter (σ^o_{true}). The DNN[WG2,LAI] uses the WG2 and LAI from the OL as the input data and DNN[multiple WGs, LAI] incorporates soil moisture from multiple layers and LAI from the OL as the input data. Both DNNs are trained to estimate σ^o_{true}. Subsequently, σ^o_{true} is assimilated to improve the OL from ISBA using different observation operators (WCM, DNN[WG2,LAI], and DNN[multiple WGs, LAI]). The study evaluates the performance skills and Jacobians of different observation operators, as well as the performances of the DA experiments.

Results show that the synthetic truth of WG2 has a negative bias compared to the OL with a peak in autumn. This is consistent with previous studies [74, 214].

When using the OL as input to estimate σ^o_{true}, the two DNNs have smaller ubRMSE and bias and larger ρ compared to WCM. Given the assumption that this WCM is perfect in our synthetic experiment, this suggests that two DNNs are overfitting because they are trained on the OL data.

The J(σ^o_{true},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(σ^o_{true},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 data-driven observation operators with inputs as close to the truth as possible.

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(σ^o_{true},WG2). DNN[multiple WGs,LAI] distributes the sensitivity to soil water dynamics to soil moisture in different layers. These underestimates of J(σ^o_{true},WG2) leads to a smaller $\tilde{\mathbf{J}}$ to WG2 in Kalman gain (equation (7)). Therefore, the performance of WG2 from DA(DNN[multiple WGs,LAI]) is worse than DA(DNN[WG2,LAI]). 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, as shown in Figure 5.11 (j) and (m). 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 critical need to carefully examine both the prediction performances and Jacobians of data-driven observation operators, especially given the growing interest in using machine learning algorithms as surrogate models for physical models in data assimilation The synthetic experiment enhances our understanding of the data-driven observation operator.

6

CONCLUSION

6.1. KNOWLEDGE GENERATED AND ORIGINAL CONTRIBUTIONS

6.1.1. ON ESTIMATING RADAR OBSERVABLES USING ISBA STATES

The necessity of constraining the plant water dynamics in the land surface model is presented in Chapter 1. Recent studies highlight that ASCAT normalized backscatter, slope, and curvature contain valuable information about vegetation water dynamics [5–7]. The key motivation of this thesis is to integrate the valuable information from ASCAT observables into the land surface model ISBA. In Chapter 3, a Deep Neural Network was trained as the observation operator to estimate the ASCAT observables using land surface variables from ISBA open loop states as inputs. Model performance shows that the DNN is able to reproduce the observed ASCAT backscatter, slope, and curvature in an independent validation period. The sensitivity of the DNN averaged over region of interest is consistent with our understanding of backscatter from theory, observations, and modeling. The estimated σ^o_{40} is sensitive to the surface soil moisture and LAI. The sensitivity varies with time and is dependent on land cover types. The slope was shown to be sensitive to LAI, but also to root zone soil moisture due to the dependence of vegetation water content on soil moisture.

The DNN directly links land surface variables to the ASCAT observables. It is not a simple emulation of an RTM because an RTM requires parameters (e.g. surface roughness, vegetation structure, vegetation water content) that are not simulated by the ISBA. However, even without vegetation water content or structure, DNN is able to reproduce the dynamics in observed ASCAT normalized backscatter and slope. Estimation performance is good and sensitivity of slope to RZSM echos physiology activities. This demonstrates that the DNN effectively models the indirect relationship between land surface states and ASCAT observables, thereby avoiding the need for explicit parameterization of vegetation water content, which could introduce additional uncertainty into ISBA. Therefore, the information about vegetation water dynamics in ASCAT observables can be directly incorporated into ISBA. Thus, the DNN enables the assimilation of the ASCAT observables to constrain the vegetation water dynamics. This approach is highly transferrable to other radar observables.

6.1.2. ON LAND DATA ASSIMILATION OF ASCAT OBSERVABLES

In Chapter 4, ASCAT normalized backscatter and slope were assimilated into land surface model ISBA using SEKF as the DA algorithm. The Deep Neural Network trained and validated in Chapter 3 was used as the observation operator to link the states to the ASCAT observables. The performances of the DA and OL were evaluated against in-situ soil moisture observations from ISMN, and CGLS LAI observations. Results show that the DA performance is neutral compared to the OL in terms of the ubRMSE and ρ across all validation sites. In addition, an analysis of the DA diagnostics confirms that DA had limited or no impact.

The neutral improvement of the DA is perplexing given the promising training results in Chapter 3. Investigation shows that the major reason might be the inconsistent terms of the Jacobian of the DNN. Compared to the results in Chapter 4 to Chapter 3, the Jacobian terms relating the ASCAT backscatter to the deep soil moisture are not physically plausible anymore, though they make sense when averaged over GPIs (shown in Chapter 3). This means the DNN might be overfitting, which results in poor estimates of the Jacobian terms. While DNN is able to model the indirect link between land surface states and ASCAT observables, it might lead to poor estimates of the Jacobian terms.

Given the growing interest in the use of machine-learning-based observation operators, Chapter 4 shows the importance of understanding the data-driven observation operator within the land data assimilation scheme. While representativeness errors and error specification play some part in the poor DA performance, it is demonstrated that the key factor constraining the efficacy of the SEKF is the correct estimation of the Jacobian that control the degree to which the observations update the states in the SEKF. It is argued that the DNN relating model states to satellite observations must have physically plausible and robust Jacobian for the DNN to be effective in a data assimilation framework.

6.1.3. ON THE EVALUATION OF MACHINE-LEARNING BASED OBSERVATION OPERATOR

Knowing the potential effects of the robustness of the Jacobian of DNN on the DA scheme, synthetic experiments were conducted to evaluate the DNN-based observation operator. The LSM ISBA is perturbed to generate synthetic truths for surface soil moisture (WG2) and leaf area index (LAI), as well as the synthetic backscatter. Two DNNs were trained based on estimating synthetic backscatter using different input sets. The synthetic backscatter was assimilated into ISBA using different observation operators.

Results in Chapter 5 show that in the autumn, both DNNs present better prediction skills compared to WCM. However, the observation equivalents from DNNs fail to capture the bias in autumn between WG2 from the OL and synthetic truth. Additionally, the DNN[WG2,LAI] overestimates the Jacobian in autumn because the DNN is trained using the OL as input data which is biased against the synthetic truth. Therefore, data-driven observation operators should be trained with inputs as close to the truth as possible. The inclusion of soil moisture in other layers slightly improves the estimation performance of the DNN[multiple WGs,LAI]. However, it leads to overfitting, the result of which is that the sensitivity to WG2 is underestimated and the sensitivity to other layers is overesti-

mated. As a result, the updates of WG2 from DA(DNN[multiple WGs,LAI]) are smaller.

Results about DA(DNN[WG2,LAI]) prove that the DA using a data-driven observation operator improves the estimates of ISBA. However, the improvement is suboptimal compared to the best case DA(WCM) due to two main reasons: a) DNN trained on OL data which is biased compared to the synthetic truth and b) the input set of the DNN might lead to overfitting and, thus, underestimate the Jacobian terms relating sigma40 to WG2 and overestimate those relating sigma40 to root zone soil moisture. It is argued that an observation operator with the smallest estimation error might not necessarily guarantee better performance in the data assimilation framework. Careful attention to the Jacobian is crucial when choosing and evaluating the machine-learning-based observation operators. In reality, the true Jacobian is not available for comparison. But, a symptom of overfitting would be poor performance of the DNN or observation operator on test data compared to training data. Finding measures to mitigate overfitting would solve the problems in the Jacobian.

It is also recommended to train the data-driven model on data that contains fewer error compared to “true" data. However, in the physical world, it is impossible to obtain “true" data. Therefore, analyzing the uncertainty or error of the training data beforehand might be useful. An alternative solution is to perturb the ISBA to create an ensemble and train DNN using this ensemble as input. The ensemble runs might include model error information, which is beneficial for the training process. Triple collocation methods can also be used to analyze the uncertainty of the training data [219].

6.1.4. ON CONSTRAINING LAND SURFACE MODEL USING ASCAT OBSERVABLES

The synthetic experiment in Chapter 5 proves that ISBA simulations are improved by assimilating normalized backscatter using a DNN as the observation operator, even without explicitly modeling scattering on the terrestrial surface. The SSM is updated more than LAI because there are more perturbations introduced in soil water dynamics. However, this approach might lead to inconsistencies among the terms of the Jacobian due to the difference between the training data and “true" data, and redundant input variables. In the synthetic experiment, there are no examinations of plant water dynamics because the WCM only accounts for first order scattering from soil and vegetation. Consequently, the synthetic slope was not simulated or assimilated in this experiment. Future work could use higher order RTMs like RT1 [220] to simulate synthetic slope, making it possible to examine the data-driven observation operator for the assimilation of slope. This would benefit the real assimilation of the ASCAT slope into ISBA, thereby better constraining vegetation water dynamics.

6.2. FUTURE RESEARCH

6.2.1. REDUCING UNCERTAINTY OF PLANT HYDRAULIC SCHEMES USING ASCAT OBSERVABLES

Currently, several new schemes related to plant hydraulics have been integrated into LSMs, such as NoahMP-PHS [174] and CLM 5.0 [29]. These models introduce equations

to simulate prognostic states of plant water storage or vegetation water potential at the root, stem, and leaf levels, providing a better representation of plant water stress during dry soil conditions.

Therefore, it is strongly recommended to use ASCAT slope and curvature to constrain the parameters and states in the plant hydraulics schemes. ASCAT slope and curvature can be modeled using the different pools of plant water storage provided by these new schemes. The methodology presented in this thesis is highly transferable to NoahMP-PHS or CLM 5.0. Similar data-driven observation operators could be trained to link the plant water storage with ASCAT observables. The Jacobian of the observation operator would highlight which state the ASCAT slope and curvature are most sensitive to. By assimilating ASCAT observables into these new schemes, the existing uncertainty in plant hydraulics models can be better reduced.

6.2.2. ESTIMATING ASCAT SLOPE AND CURVATURE USING RADIATIVE TRANSFER MODELS

Previous studies have shown that ASCAT slope represents the wet biomass of vegetation [45]. The seasonal cycle of slope reveals vegetation dynamics and differs from the seasonal cycles of other vegetation-related variables like LAI or biomass. To better assimilate ASCAT slope into LSMs, it is important to understand ASCAT slope from the perspective of microwave remote sensing. However, there are few theoretical studies modeling ASCAT slope using RTMs.

It still remains open as to which parts of the plants the ASCAT slope and curvature are most sensitive. This sensitivity varies with plant types and may change from sub-daily to interannual scales. Some studies [5, 66] identified positive curvature in grassland. Steele-Dunne *et al.* [5] found diurnal differences in ASCAT slope and curvature. Blanken [160] observed positive curvature anomalies after wildfires in eucalyptus forests in Australia and explained this might be due to vertical tree trunks being less affected than leaves and branches. Sensitivity analysis is needed to quantify the contribution of plant organs (leaves, branches, and trunks) to slope and curvature. Future studies could use radiative transfer models like MIMICS [129] to simulate slope using different values of moisture content in leaf and trunks.

An improved understanding of ASCAT slope and curvature helps us to better find the land surface state which can be updated by the assimilation of ASCAT slope and curvature. According to Chapter 5, this benefits the data-driven observation operator by avoiding potential overfitting problems.

6.2.3. IMPROVING THE ROBUSTNESS OF THE JACOBIAN OF THE MACHINE-LEARNING BASED OBSERVATION OPERATOR

The use of machine-learning-based observation operators in land data assimilation frameworks becomes increasingly prevalent. The data-driven method avoids explicitly modeling the dielectric properties of soil and vegetation and directly links the model states to satellite observables. Therefore, it avoids the potential cross-correlated errors between retrievals and model simulations. This thesis highlights that there is potential in ML-based measurement operators to get the Jacobian wrong even if the simulated and ob-

served target variables agree well. Future research in the application of machine learning to earth system science should focus on reducing the overfitting and enhancing the robustness of the Jacobian. Chapter 5 suggests that training models using data with fewer errors could be beneficial. Therefore, it might be helpful to better characterize the uncertainty or error of the models and datasets. Tripple collocation analysis might help to characterize the errors in model simulations in the absence of truth [219].

Additionally, the selection of appropriate inputs is also important. Using more input variables could improve the validation performance of the data-driven model. However, this approach may compromise the accuracy of the Jacobian estimates, which are essential in the following DA or earth system model applications. This issue might arise due to the scarcity of training data in our current approach, where the DNN is trained per grid point (GPI). Therefore, it is recommended to compare the current training approach with a more generic method, such as training the DNN per land cover type. Although the estimate performance might be worse, the Jacobian could be more robust.

Including the Jacobian term in the ML training target might be a potential solution to improve the robustness of the data-driven observation operator. Some studies [221] proposed to add a penalty term of the Jacobian during the ML training process. Other studies [222, 223] tried to calculate the Jacobian based on the nearest sampling points of the training data. Future work is needed to examine these new training methods on ASCAT dynamic vegetation parameters or other microwave remote sensing data. These new training methods might lead to more robust estimates of the Jacobian.

6.2.4. CORRECTING THE BIAS AND ERRORS IN LAND SURFACE MODEL SIMULATIONS

Chapter 5 showed that the DNN[WG2,LAI] overestimates the Jacobian of normalized backscatter to WG2 in autumn due to the bias between the open loop and the synthetic truth (perturbed run). This suggests that the bias between the LSM and the truth may play a critical role during the DNN training process, as the bias between the LSM OL and “true" data may lead to overfitting in the Jacobian, even though the "true" data are never known in reality. Therefore, it is recommended to examine the LSM parameterizations prior to the DA. Several studies highlight the importance of enhancing the model regarding parameterizations or parameter tuning [108, 156], or using a bias-aware assimilation method to explicitly correct the model bias [102]. Future work is suggested to explore how the quantification of bias and errors in LSM simulations benefits ML training and the estimation of the Jacobian.

Additionally, it might be helpful to use bias correction techniques before the training of the machine-learning-based observation operator. Although normalization has been applied to make the input variables unitless, results presented in Chapters 4 and 5 show that this technique does not eliminate the errors between the input data and the synthetic truth. Using the uncertainty propogation techniques, like Bayesian Neural Network, may help to quantify the effects of errors in input data on training process.

6.2.5. UNDERSTANDING THE VALUE OF ASCAT SLOPE AND CURVATURE FOR CARBON CYCLE SCIENCE

ASCAT slope and curvature contain valuable information about the vertical biomass distribution of plants [5]. Some preliminary work in Vreugdenhil *et al.* [224] shows that ASCAT slope is correlated with slow and fast carbon processes simulated by terrestrial biosphere model (TBM) DALEC-BETHY in boreal forest site in Finland. Therefore, more work could be done to analyze the relationship between slope and different vegetation carbon pools. The relationship may vary with time and vegetation cover type. It might also depend on whether the physiological activities of vegetation are limited by water conditions or energy conditions. Future research could also focus on monitoring vegetation conditions by analyzing the time series anomalies of ASCAT slope and curvature. This might benefit our understanding of vegetation carbon dynamics. Further, ASCAT slope and curvature could be assimilated directly into TBMs to constrain states and parameters related to the fast and slow water and carbon processes.

6.2.6. ASSIMILATING MULTI-SOURCE EARTH OBSERVATIONS INTO EARTH SYSTEM MODELS

The methodology provided in this thesis offers a framework for assimilating multi-source earth observations, like sun-induced fluorescence, vegetation optical depth, soil moisture, and surface temperature, into earth system models simultaneously. The machine-learning-based observation operator presented in Chapter 3 maps from the space of the model states to the observation space. However, since different earth observations from independent sources are sensitive to different model states, throwing every state and observation into one observation operator may cause overfitting, which results in poor estimates of the Jacobian terms, as suggested in Chapter 5. An alternative way is to train different observation operators for each type of observation, but this may increase the computational costs. Future work could investigate ways to train the observation operators together, such as masking the states during the training process or choosing a different deep learning algorithm, like transformer architectures, to develop a surrogate model for different observations. Transformer architectures are trained using large numbers of learnable parameters to learn complex patterns from the training data and improve weather forecasting [225].

6.2.7. USING ASCAT DYNAMIC VEGETATION PARAMETERS TO MONITOR CLIMATE EXTREMES

As summarized in Chapter 1, LSMs often fail to predict drought accurately due to the deficiencies in estimating water and carbon cycles. However, recent studies have shown that the anomalies of the ASCAT slope and curvature are beneficial to the monitoring and prediction of drought events [5, 62] and wildfires [160]. Therefore, incorporating ASCAT information into LSM is expected to improve the estimations of water states and fluxes during drought conditions. Since ASCAT normalized backscatter, slope, and curvature provide different information about soil and vegetation dynamics, comparing the simultaneous assimilation of ASCAT observables with individual assimilation could help

identify which parts of the model parameterization contain larger errors. This, in turn, would significantly benefit the credibility of land surface model projections and contribute to a better understanding of land-atmosphere interactions.

REFERENCES

[1] Y. Liu, M. Kumar, G. G. Katul, X. Feng, and A. G. Konings. "Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration". In: *Nature Climate Change* 10.7 (June 2020), pp. 691–695. DOI: 10.1038/s41558-020-0781-5.

[2] W. R. L. Anderegg, A. Wolf, A. Arango-Velez, B. Choat, D. J. Chmura, S. Jansen, T. Kolb, S. Li, F. Meinzer, P. Pita, V. R. de Dios, J. S. Sperry, B. T. Wolfe, and S. Pacala. "Plant water potential improves prediction of empirical stomatal models". In: *PLOS ONE* 12.10 (Oct. 2017). Ed. by R. Aroca, e0185481. DOI: 10.1371/journal.pone.0185481.

[3] T. L. Powell, D. R. Galbraith, B. O. Christoffersen, A. Harper, H. M. A. Imbuzeiro, L. Rowland, S. Almeida, P. M. Brando, A. C. L. da Costa, M. H. Costa, N. M. Levine, Y. Malhi, S. R. Saleska, E. Sotta, M. Williams, P. Meir, and P. R. Moorcroft. "Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought". In: *New Phytologist* 200.2 (July 2013), pp. 350–365. DOI: 10.1111/nph.12390.

[4] S. Hahn, C. Reimer, M. Vreugdenhil, T. Melzer, and W. Wagner. "Dynamic Characterization of the Incidence Angle Dependence of Backscatter Using Metop ASCAT". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2348–2359. DOI: 10.1109/jstars.2016.2628523.

[5] S. C. Steele-Dunne, S. Hahn, W. Wagner, and M. Vreugdenhil. "Investigating vegetation water dynamics and drought using Metop ASCAT over the North American Grasslands". In: *Remote Sensing of Environment* 224 (Apr. 2019), pp. 219–235. DOI: 10.1016/j.rse.2019.01.004.

[6] I. Pfeil, W. Wagner, M. Forkel, W. Dorigo, and M. Vreugdenhil. "Does ASCAT observe the spring reactivation in temperate deciduous broadleaf forests?" In: *Remote Sensing of Environment* 250 (Dec. 2020), p. 112042. DOI: 10.1016/j.rse.2020.112042.

[7] A. Petchiappan, S. C. Steele-Dunne, M. Vreugdenhil, S. Hahn, W. Wagner, and R. Oliveira. "The influence of vegetation water dynamics on the ASCAT backscatter-incidence angle relationship in the Amazon". In: (Aug. 2021). DOI: 10.5194/hess-2021-406. URL: https://doi.org/10.5194/hess-2021-406.

[8] H. Lievens, B. Martens, N. Verhoest, S. Hahn, R. Reichle, and D. Miralles. “Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates”. In: *Remote Sensing of Environment* 189 (Feb. 2017), pp. 194–210. DOI: 10.1016/j.rse.2016.11.022.

[9] D. C. Shamambo. “Assimilation de données satellitaires pour le suivi des ressources en eau dans la zone Euro-Méditerranée”. PhD thesis. Toulouse 3, 2020.

[10] A. M. Ukkola, M. G. De Kauwe, A. J. Pitman, M. J. Best, G. Abramowitz, V. Haverd, M. Decker, and N. Haughton. “Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts”. In: *Environmental Research Letters* 11.10 (Oct. 2016), p. 104012. ISSN: 1748-9326. DOI: 10.1088/1748-9326/11/10/104012. URL: http://dx.doi.org/10.1088/1748-9326/11/10/104012.

[11] R. Whitley, J. Beringer, L. B. Hutley, G. Abramowitz, M. G. De Kauwe, R. Duursma, B. Evans, V. Haverd, L. Li, Y. Ryu, B. Smith, Y.-P. Wang, M. Williams, and Q. Yu. “A model inter-comparison study to examine limiting factors in modelling Australian tropical savannas”. In: *Biogeosciences* 13.11 (June 2016), pp. 3245–3265. ISSN: 1726-4189. DOI: 10.5194/bg-13-3245-2016. URL: http://dx.doi.org/10.5194/bg-13-3245-2016.

[12] I. C. Prentice, X. Liang, B. E. Medlyn, and Y.-P. Wang. “Reliable, robust and realistic: the three R’s of next-generation land-surface modelling”. In: *Atmospheric Chemistry and Physics* 15.10 (May 2015), pp. 5987–6005. ISSN: 1680-7324. DOI: 10.5194/acp-15-5987-2015. URL: http://dx.doi.org/10.5194/acp-15-5987-2015.

[13] J. Dong, F. Lei, and W. T. Crow. “Land transpiration-evaporation partitioning errors responsible for modeled summertime warm bias in the central United States”. In: *Nature Communications* 13.1 (Jan. 2022). ISSN: 2041-1723. DOI: 10.1038/s41467-021-27938-6. URL: http://dx.doi.org/10.1038/s41467-021-27938-6.

[14] S. Xie, X. Mo, S. Liu, and S. Hu. “Plant Hydraulics Improves Predictions of ET and GPP Responses to Drought”. In: *Water Resources Research* 59.5 (May 2023). ISSN: 1944-7973. DOI: 10.1029/2022wr033402. URL: http://dx.doi.org/10.1029/2022WR033402.

[15] “The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field”. In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 273.927 (Feb. 1976), pp. 593–610. DOI: 10.1098/rstb.1976.0035.

[16] C. Jacobs, B. van den Hurk, and H. de Bruin. “Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions”. In: *Agricultural and Forest Meteorology* 80.2-4 (July 1996), pp. 111–134. DOI: 10.1016/0168-1923(95)02295-3.

[17] C. M. J. Jacobs. *Direct impact of atmospheric CO2 enrichment on regional transpiration*. Jacobs, 1994.

[18] J. T. Ball, I. E. Woodrow, and J. A. Berry. “A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions”. In: *Progress in Photosynthesis Research*. Springer Netherlands, 1987, pp. 221–224. DOI: 10.1007/978-94-017-0519-6_48.

[19] R. Leuning. “Modelling Stomatal Behaviour and and Photosynthesis of Eucalyptus grandis”. In: *Functional Plant Biology* 17.2 (1990), p. 159. DOI: 10.1071/pp9900159.

[20] R. LEUNING. “A critical appraisal of a combined stomatal-photosynthesis model for C3 plants”. In: *Plant, Cell and Environment* 18.4 (Apr. 1995), pp. 339–355. DOI: 10.1111/j.1365-3040.1995.tb00370.x.

[21] V. K. Arora. “Simulating energy and carbon fluxes over winter wheat using coupled land surface and terrestrial ecosystem models”. In: *Agricultural and Forest Meteorology* 118.1-2 (Aug. 2003), pp. 21–47. DOI: 10.1016/s0168-1923(03)00073-x.

[22] J.-C. Calvet, V. Rivalland, C. Picon-Cochard, and J.-M. Guehl. “Modelling forest transpiration and CO2 fluxes—response to soil moisture stress”. In: *Agricultural and Forest Meteorology* 124.3-4 (Aug. 2004), pp. 143–156. DOI: 10.1016/j.agrformet.2004.01.007. URL: https://doi.org/10.1016/j.agrformet.2004.01.007.

[23] T. Keenan, R. García, A. D. Friend, S. Zaehle, C. Gracia, and S. Sabate. “Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO2 and water fluxes through combined in situ measurements and ecosystem modelling”. In: *Biogeosciences* 6.8 (Aug. 2009), pp. 1423–1444. DOI: 10.5194/bg-6-1423-2009.

[24] M. Reichstein, J. D. Tenhunen, O. Roupsard, J.-m. Ourcival, S. Rambal, F. Miglietta, A. Peressotti, M. Pecchiari, G. Tirone, and R. Valentini. “Severe drought effects on ecosystem CO2 and H2 O fluxes at three Mediterranean evergreen sites: revision of current hypotheses?” In: *Global Change Biology* 8.10 (Sept. 2002), pp. 999–1017. DOI: 10.1046/j.1365-2486.2002.00530.x.

[25] A. Verhoef and S. Allen. “A SVAT scheme describing energy and CO2 fluxes for multi-component vegetation: calibration and test for a Sahelian savannah”. In: *Ecological Modelling* 127.2-3 (Mar. 2000), pp. 245–267. DOI: 10.1016/s0304-3800(99)00213-6.

[26] G. B. Bonan, M. Williams, R. A. Fisher, and K. W. Oleson. “Modeling stomatal conductance in the Earth system: linking leaf water-use efficiency and water transport along the soil-plant-atmosphere continuum”. In: *Geoscientific Model Development Discussions* 7.3 (May 2014), pp. 3085–3159. DOI: 10.5194/gmdd-7-3085-2014.

[27] T. Klein. “The variability of stomatal sensitivity to leaf water potential across tree species indicates a continuum between isohydric and anisohydric behaviours”. In: *Functional Ecology* 28.6 (May 2014). Ed. by S. Niu, pp. 1313–1320. DOI: 10.1111/1365-2435.12289.

[28] H. Cochard, F. Pimont, J. Ruffault, and N. Martin-StPaul. “SurEau: a mechanistic model of plant water relations under extreme drought”. In: *Annals of Forest Science* 78.2 (June 2021). ISSN: 1297-966X. DOI: 10.1007/s13595-021-01067-y. URL: http://dx.doi.org/10.1007/s13595-021-01067-y.

[29] D. Kennedy, S. Swenson, K. W. Oleson, D. M. Lawrence, R. Fisher, A. C. L. da Costa, and P. Gentine. “Implementing Plant Hydraulics in the Community Land Model, Version 5”. In: *Journal of Advances in Modeling Earth Systems* 11.2 (Feb. 2019), pp. 485–513. DOI: 10.1029/2018ms001500.

[30] G. Bohrer, H. Mourad, T. A. Laursen, D. Drewry, R. Avissar, D. Poggi, R. Oren, and G. G. Katul. “Finite element tree crown hydrodynamics model (FETCH) using porous media flow within branching elements: A new representation of tree hydrodynamics”. In: *Water Resources Research* 41.11 (Nov. 2005). DOI: 10.1029/2005wr004181.

[31] A. G. Konings, K. Rao, and S. C. Steele-Dunne. “Macro to micro: microwave remote sensing of plant water content for physiology and ecology”. In: *New Phytologist* 223.3 (Apr. 2019), pp. 1166–1172. DOI: 10.1111/nph.15808.

[32] Y. Y. Liu, R. A. M. de Jeu, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk. “Global long-term passive microwave satellite-based retrievals of vegetation optical depth: LONG-TERM VEGETATION OPTICAL DEPTH”. In: *Geophysical Research Letters* 38.18 (Sept. 2011). ISSN: 0094-8276. DOI: 10.1029/2011gl048684. URL: http://dx.doi.org/10.1029/2011gl048684.

[33] J. Friesen, S. C. Steele-Dunne, and N. van de Giesen. “Diurnal Differences in Global ERS Scatterometer Backscatter Observations of the Land Surface”. In: *IEEE Transactions on Geoscience and Remote Sensing* 50.7 (July 2012), pp. 2595–2602. DOI: 10.1109/tgrs.2012.2193889.

[34] S. C. Steele-Dunne, J. Friesen, and N. van de Giesen. “Using Diurnal Variation in Backscatter to Detect Vegetation Water Stress”. In: *IEEE Transactions on Geoscience and Remote Sensing* 50.7 (July 2012), pp. 2618–2629. DOI: 10.1109/tgrs.2012.2194156.

[35] T. van Emmerik, S. C. Steele-Dunne, J. Judge, and N. van de Giesen. “Impact of Diurnal Variation in Vegetation Water Content on Radar Backscatter From Maize During Water Stress”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.7 (July 2015), pp. 3855–3869. DOI: 10.1109/tgrs.2014.2386142.

[36] T. van Emmerik, S. Steele-Dunne, A. Paget, R. S. Oliveira, P. R. L. Bittencourt, F. de V. Barros, and N. van de Giesen. “Water stress detection in the Amazon using radar”. In: *Geophysical Research Letters* 44.13 (July 2017), pp. 6841–6849. DOI: 10.1002/2017gl073747.

[37] A. G. Konings, M. Piles, N. Das, and D. Entekhabi. “L-band vegetation optical depth and effective scattering albedo estimation from SMAP”. In: *Remote Sensing of Environment* 198 (Sept. 2017), pp. 460–470. DOI: 10.1016/j.rse.2017.06.037.

[38] A. G. Konings, S. S. Saatchi, C. Frankenberg, M. Keller, V. Leshyk, W. R. L. Anderegg, V. Humphrey, A. M. Matheny, A. Trugman, L. Sack, E. Agee, M. L. Barnes, O. Binks, K. Cawse-Nicholson, B. O. Christoffersen, D. Entekhabi, P. Gentine, N. M. Holtzman, G. G. Katul, Y. Liu, M. Longo, J. Martinez-Vilalta, N. McDowell, P. Meir, M. Mencuccini, A. Mrad, K. A. Novick, R. S. Oliveira, P. Siqueira, S. C. Steele-Dunne, D. R. Thompson, Y. Wang, R. Wehr, J. D. Wood, X. Xu, and P. A. Zuidema. "Detecting forest response to droughts with global observations of vegetation water content". In: 27.23 (Sept. 2021), pp. 6005–6024. DOI: 10.1111/gcb.15872. URL: https://doi.org/10.1111/gcb.15872.

[39] I. E. Teubner, M. Forkel, M. Jung, Y. Y. Liu, D. G. Miralles, R. Parinussa, R. van der Schalie, M. Vreugdenhil, C. R. Schwalm, G. Tramontana, G. Camps-Valls, and W. A. Dorigo. "Assessing the relationship between microwave vegetation optical depth and gross primary production". In: *International Journal of Applied Earth Observation and Geoinformation* 65 (Mar. 2018), pp. 79–91. DOI: 10.1016/j.jag.2017.10.006.

[40] L. Moesinger, W. Dorigo, R. de Jeu, R. van der Schalie, T. Scanlon, I. Teubner, and M. Forkel. "The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA)". In: *Earth System Science Data* 12.1 (Jan. 2020), pp. 177–196. DOI: 10.5194/essd-12-177-2020.

[41] A. M. Matheny, G. Mirfenderesgi, and G. Bohrer. "Trait-based representation of hydrological functional properties of plants in weather and ecosystem models". In: *Plant Diversity* 39.1 (Feb. 2017), pp. 1–12. DOI: 10.1016/j.pld.2016.10.001.

[42] A. G. Konings, Y. Yu, L. Xu, Y. Yang, D. S. Schimel, and S. S. Saatchi. "Active microwave observations of diurnal and seasonal variations of canopy water content across the humid African tropical forests". In: *Geophysical Research Letters* 44.5 (Mar. 2017), pp. 2290–2299. DOI: 10.1002/2016gl072388.

[43] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P.-W. Liu, and K. Papathanassiou. "Radar Remote Sensing of Agricultural Canopies: A Review". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2249–2273. DOI: 10.1109/jstars.2016.2639043.

[44] W. Wagner, G. Lemoine, and H. Rott. "A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data". In: *Remote Sensing of Environment* 70.2 (Nov. 1999), pp. 191–207. DOI: 10.1016/s0034-4257(99)00036-x.

[45] W. Wagner, G. Lemoine, M. Borgeaud, and H. Rott. "A study of vegetation cover effects on ERS scatterometer data". In: *IEEE Transactions on Geoscience and Remote Sensing* 37.2 (Mar. 1999), pp. 938–948. DOI: 10.1109/36.752212. URL: https://doi.org/10.1109/36.752212.

[46] W. Wagner, J. Noll, M. Borgeaud, and H. Rott. "Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer". In: *IEEE Transactions on Geoscience and Remote Sensing* 37.1 (1999), pp. 206–216. DOI: 10.1109/36.739155.

[47] R. Schroeder, K. C. McDonald, M. Azarderakhsh, and R. Zimmermann. “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”. In: *Remote Sensing of Environment* 177 (May 2016), pp. 153–159. DOI: 10.1016/j.rse.2016.01.008.

[48] M. Vreugdenhil, S. Hahn, T. Melzer, B. BauerMarschallinger, C. Reimer, W. A. Dorigo, and W. Wagner. “Assessing Vegetation Dynamics Over Mainland Australia With Metop ASCAT”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.5 (May 2017), pp. 2240–2248. DOI: 10.1109/jstars.2016.2618838.

[49] B. Brisco, R. Brown, J. Koehler, G. Sofko, and M. Mckibben. “The diurnal pattern of microwave backscattering by wheat”. In: *Remote Sensing of Environment* 34.1 (Oct. 1990), pp. 37–47. DOI: 10.1016/0034-4257(90)90082-w.

[50] F. Ulaby and P. Batlivala. “Diurnal variations of radar backscatter from a vegetation canopy”. In: *IEEE Transactions on Antennas and Propagation* 24.1 (Jan. 1976), pp. 11–17. DOI: 10.1109/tap.1976.1141298.

[51] H. H. Dixon and J. Joly. “On the ascent of sap.” In: *Proceedings of the Royal Society of London* 57 (1894), pp. 3–5.

[52] S. Saatchi, W. Buermann, H. ter Steege, S. Mori, and T. B. Smith. “Modeling distribution of Amazonian tree species and diversity using remote sensing measurements”. In: *Remote Sensing of Environment* 112.5 (May 2008), pp. 2000–2017. ISSN: 0034-4257. DOI: 10.1016/j.rse.2008.01.008. URL: http://dx.doi.org/10.1016/j.rse.2008.01.008.

[53] S. Frolking, T. Milliman, M. Palace, D. Wisser, R. Lammers, and M. Fahnestock. “Tropical forest backscatter anomaly evident in SeaWinds scatterometer morning overpass data during 2005 drought in Amazonia”. In: *Remote Sensing of Environment* 115.3 (Mar. 2011), pp. 897–907. DOI: 10.1016/j.rse.2010.11.017.

[54] A. P. O’Grady, D. Eamus, and L. B. Hutley. “Transpiration increases during the dry season: patterns of tree water use in eucalypt open-forests of northern Australia”. In: *Tree Physiology* 19.9 (July 1999), pp. 591–597. DOI: 10.1093/treephys/19.9.591.

[55] L. B. Hutley, A. P. O’Grady, and D. Eamus. “Evapotranspiration from Eucalypt open-forest savanna of Northern Australia”. In: *Functional Ecology* 14.2 (Apr. 2000), pp. 183–194. DOI: 10.1046/j.1365-2435.2000.00416.x.

[56] K. Guan, A. Wolf, D. Medvigy, K. K. Caylor, M. Pan, and E. F. Wood. “Seasonal coupling of canopy structure and function in African tropical forests and its environmental controls”. In: *Ecosphere* 4.3 (Mar. 2013), art35. DOI: 10.1890/es12-00232.1.

[57] M. Vreugdenhil, W. Wagner, B. Bauer-Marschallinger, I. Pfeil, I. Teubner, C. Rüdiger, and P. Strauss. “Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study”. In: *Remote Sensing* 10.9 (Sept. 2018), p. 1396. DOI: 10.3390/rs10091396.

[58] K. Scipal, W. Wagner, M. Trommler, and K. Naumann. “The global soil moisture archive 1992-2000 from ERS scatterometer data: first results”. In: *IEEE International Geoscience and Remote Sensing Symposium*. IGARSS-02. IEEE. DOI: 10.1109/igarss.2002.1026129. URL: http://dx.doi.org/10.1109/IGARSS.2002.1026129.

[59] Z. Bartalis, W. Wagner, V. Naeimi, S. Hasenauer, K. Scipal, H. Bonekamp, J. Figa, and C. Anderson. “Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT)”. In: *Geophysical Research Letters* 34.20 (Oct. 2007). DOI: 10.1029/2007gl031088.

[60] V. Naeimi, K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner. “An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations”. In: *IEEE Transactions on Geoscience and Remote Sensing* 47.7 (July 2009), pp. 1999–2013. DOI: 10.1109/tgrs.2008.2011617.

[61] T. Melzer. “Vegetation modelling in WARP 6.0”. In: *Proceedings of the EUMETSAT Meteorological Satellite Conference, Vienna, Austria*. 2013, pp. 16–20.

[62] B. J. Walraven. “Monitoring the impact of droughts on vegetation in Australia using MetOp ASCAT Dynamic Vegetation Parameters”. MA thesis. TU Delft, 2020.

[63] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave remote sensing: Active and passive. Volume 3-From theory to applications*. 1986.

[64] S. Bakhtiari and R. Zoughi. “A model for backscattering characteristics of tall prairie grass canopies at microwave frequencies”. In: *Remote Sensing of Environment* 36.2 (May 1991), pp. 137–147. DOI: 10.1016/0034-4257(91)90036-6.

[65] F. Mattia, T. L. Toan, G. Picard, F. Posa, A. D’Alessio, C. Notarnicola, A. Gatti, M. Rinaldi, G. Satalino, and G. Pasquariello. “Multitemporal c-band radar measurements on wheat fields”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.7 (July 2003), pp. 1551–1560. DOI: 10.1109/tgrs.2003.813531.

[66] J. Stiles, K. Sarabandi, and F. Ulaby. “Electromagnetic scattering from grassland. II. Measurement and modeling results”. In: *IEEE Transactions on Geoscience and Remote Sensing* 38.1 (2000), pp. 349–356. DOI: 10.1109/36.823930.

[67] T. F. Bush and F. T. Ulaby. “Remotely sensing wheat maturation with radar”. In: (1975).

[68] U. Wegmüller. “Signature research for crop classification by active and passive microwaves”. In: *International Journal of Remote Sensing* 14.5 (Mar. 1993), pp. 871–883. DOI: 10.1080/01431169308904383.

[69] C. Albergel, S. Munier, A. Bocher, B. Bonan, Y. Zheng, C. Draper, D. Leroux, and J.-C. Calvet. “LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the Contiguous US: An ERA-5 Driven Reanalysis of the Land Surface Variables”. In: *Remote Sensing* 10.10 (Oct. 2018), p. 1627. DOI: 10.3390/rs10101627.

[70] S. Kumar, J. Kolassa, R. Reichle, W. Crow, G. de Lannoy, P. de Rosnay, N. MacBean, M. Girotto, A. Fox, T. Quaife, C. Draper, B. Forman, G. Balsamo, S. Steele-Dunne, C. Albergel, B. Bonan, J.-C. Calvet, J. Dong, H. Liddy, and B. Ruston. "An Agenda for Land Data Assimilation Priorities: Realizing the Promise of Terrestrial Water, Energy, and Vegetation Observations From Space". In: *Journal of Advances in Modeling Earth Systems* 14.11 (Oct. 2022). DOI: 10.1029/2022ms003259. URL: https://doi.org/10.1029/2022ms003259.

[71] P. A. Dirmeyer, S. Halder, and R. Bombardi. "On the Harvest of Predictability From Land States in a Global Forecast Model". In: *Journal of Geophysical Research: Atmospheres* 123.23 (Dec. 2018). ISSN: 2169-8996. DOI: 10.1029/2018jd029103. URL: http://dx.doi.org/10.1029/2018JD029103.

[72] G. J. M. D. Lannoy, M. Bechtold, C. Albergel, L. Brocca, J.-C. Calvet, A. Carrassi, W. T. Crow, P. de Rosnay, M. Durand, B. Forman, G. Geppert, M. Girotto, H.-J. H. Franssen, T. Jonas, S. Kumar, H. Lievens, Y. Lu, C. Massari, V. R. N. Pauwels, R. H. Reichle, and S. Steele-Dunne. "Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication". In: *Frontiers in Water* 4 (Sept. 2022). DOI: 10.3389/frwa.2022.981745. URL: https://doi.org/10.3389/frwa.2022.981745.

[73] S. Mecklenburg, M. Drusch, L. Kaleschke, N. Rodriguez-Fernandez, N. Reul, Y. Kerr, J. Font, M. Martin-Neira, R. Oliva, E. Daganzo-Eusebio, J. Grant, R. Sabia, G. Macelloni, K. Rautiainen, J. Fauste, P. de Rosnay, J. Munoz-Sabater, N. Verhoest, H. Lievens, S. Delwart, R. Crapolicchio, A. de la Fuente, and M. Kornberg. "ESA's Soil Moisture and Ocean Salinity mission: From science to operational applications". In: *Remote Sensing of Environment* 180 (July 2016), pp. 3–18. DOI: 10.1016/j.rse.2015.12.025. URL: https://doi.org/10.1016/j.rse.2015.12.025.

[74] B. Bonan, C. Albergel, Y. Zheng, A. L. Barbu, D. Fairbairn, S. Munier, and J.-C. Calvet. "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". In: *Hydrology and Earth System Sciences* 24.1 (Jan. 2020), pp. 325–347. DOI: 10.5194/hess-24-325-2020.

[75] K. Imaoka, M. Kachi, M. Kasahara, N. Ito, K. Nakagawa, and T. Oki. "Instrument performance and calibration of AMSR-E and AMSR2". In: *International archives of the photogrammetry, remote sensing and spatial information science* 38.8 (2010), pp. 13–18.

[76] E. Njoku, T. Jackson, V. Lakshmi, T. Chan, and S. Nghiem. "Soil moisture retrieval from AMSR-E". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2 (Feb. 2003), pp. 215–229. DOI: 10.1109/tgrs.2002.808243. URL: https://doi.org/10.1109/tgrs.2002.808243.

[77] R. H. Reichle, R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe. "Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR)". In: *Journal of Geo-*

physical Research 112.D9 (May 2007). DOI: 10.1029/2006jd008033. URL: https://doi.org/10.1029/2006jd008033.

[78] C. S. Draper, J. P. Walker, P. J. Steinle, R. A. de Jeu, and T. R. Holmes. "An evaluation of AMSR–E derived soil moisture over Australia". In: *Remote Sensing of Environment* 113.4 (Apr. 2009), pp. 703–710. DOI: 10.1016/j.rse.2008.11.011.

[79] C. S. Draper, R. H. Reichle, G. J. M. D. Lannoy, and Q. Liu. "Assimilation of passive and active microwave soil moisture retrievals". In: *Geophysical Research Letters* 39.4 (Feb. 2012), n/a–n/a. DOI: 10.1029/2011gl050655. URL: https://doi.org/10.1029/2011gl050655.

[80] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martín-Neira, and S. Mecklenburg. "The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle". In: *Proceedings of the IEEE* 98.5 (May 2010), pp. 666–687. DOI: 10.1109/jproc.2010.2043032. URL: https://doi.org/10.1109/jproc.2010.2043032.

[81] M.-E. Ridler, H. Madsen, S. Stisen, S. Bircher, and R. Fensholt. "Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark". In: *Water Resources Research* 50.11 (Nov. 2014), pp. 8962–8981. DOI: 10.1002/2014wr015392. URL: https://doi.org/10.1002/2014wr015392.

[82] H. Lievens, S. Tomer, A. A. Bitar, G. D. Lannoy, M. Drusch, G. Dumedah, H.-J. H. Franssen, Y. Kerr, B. Martens, M. Pan, J. Roundy, H. Vereecken, J. Walker, E. Wood, N. Verhoest, and V. Pauwels. "SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia". In: *Remote Sensing of Environment* 168 (Oct. 2015), pp. 146–162. DOI: 10.1016/j.rse.2015.06.025.

[83] B. Martens, D. Miralles, H. Lievens, D. Fernández-Prieto, and N. Verhoest. "Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture". In: *International Journal of Applied Earth Observation and Geoinformation* 48 (June 2016), pp. 146–162. DOI: 10.1016/j.jag.2015.09.012. URL: https://doi.org/10.1016/j.jag.2015.09.012.

[84] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. V. Zyl. "The Soil Moisture Active Passive (SMAP) Mission". In: *Proceedings of the IEEE* 98.5 (May 2010), pp. 704–716. DOI: 10.1109/jproc.2010.2043918. URL: https://doi.org/10.1109/jproc.2010.2043918.

[85] I. E. Mladenova, J. D. Bolten, W. Crow, N. Sazib, and C. Reynolds. "Agricultural Drought Monitoring via the Assimilation of SMAP Soil Moisture Retrievals Into a Global Soil Water Balance Model". In: *Frontiers in Big Data* 3 (Apr. 2020). DOI: 10.3389/fdata.2020.00010. URL: https://doi.org/10.3389/fdata.2020.00010.

[86] E. Pinnington, J. Amezcua, E. Cooper, S. Dadson, R. Ellis, J. Peng, E. Robinson, R. Morrison, S. Osborne, and T. Quaife. "Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data". In: *Hydrology and Earth System Sciences* 25.3 (Mar. 2021), pp. 1617–1641. DOI: 10.5194/hess-25-1617-2021. URL: https://doi.org/10.5194/hess-25-1617-2021.

[87] E. Seo, M.-I. Lee, and R. H. Reichle. "Assimilation of SMAP and ASCAT soil moisture retrievals into the JULES land surface model using the Local Ensemble Transform Kalman Filter". In: *Remote Sensing of Environment* 253 (Feb. 2021), p. 112222. DOI: 10.1016/j.rse.2020.112222. URL: https://doi.org/10.1016/j.rse.2020.112222.

[88] C. Albergel, S. Munier, D. J. Leroux, H. Dewaele, D. Fairbairn, A. L. Barbu, E. Gelati, W. Dorigo, S. Faroux, C. Meurey, P. L. Moigne, B. Decharme, J.-F. Mahfouf, and J.-C. Calvet. "Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro- Mediterranean area". In: *Geoscientific Model Development* 10.10 (Oct. 2017), pp. 3889–3912. DOI: 10.5194/gmd-10-3889-2017.

[89] J. M. Sabater, C. Rüdiger, J.-C. Calvet, N. Fritz, L. Jarlan, and Y. Kerr. "Joint assimilation of surface soil moisture and LAI observations into a land surface model". In: *Agricultural and Forest Meteorology* 148.8-9 (July 2008), pp. 1362–1373. DOI: 10.1016/j.agrformet.2008.04.003. URL: https://doi.org/10.1016/j.agrformet.2008.04.003.

[90] C. Albergel, J.-C. Calvet, J.-F. Mahfouf, C. Rüdiger, A. L. Barbu, S. Lafont, J.-L. Roujean, J. P. Walker, M. Crapeau, and J.-P. Wigneron. "Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France". In: *Hydrology and Earth System Sciences* 14.6 (June 2010), pp. 1109–1124. DOI: 10.5194/hess-14-1109-2010. URL: https://doi.org/10.5194/hess-14-1109-2010.

[91] P. Baguis, A. Carrassi, E. Roulin, S. Vannitsem, S. Modanesi, H. Lievens, M. Bechtold, and G. D. Lannoy. "Assimilation of Backscatter Observations into a Hydrological Model: A Case Study in Belgium Using ASCAT Data". In: *Remote Sensing* 14.22 (Nov. 2022), p. 5740. DOI: 10.3390/rs14225740. URL: https://doi.org/10.3390/rs14225740.

[92] H. Lievens, R. H. Reichle, Q. Liu, G. J. M. D. Lannoy, R. S. Dunbar, S. B. Kim, N. N. Das, M. Cosh, J. P. Walker, and W. Wagner. "Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates". In: *Geophysical Research Letters* 44.12 (June 2017), pp. 6145–6153. DOI: 10.1002/2017gl073904. URL: https://doi.org/10.1002/2017gl073904.

[93] D. Rains, H. Lievens, G. J. M. D. Lannoy, M. F. Mccabe, R. A. M. de Jeu, and D. G. Miralles. "Sentinel-1 Backscatter Assimilation Using Support Vector Regression or the Water Cloud Model at European Soil Moisture Sites". In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5. DOI: 10.1109/lgrs.2021.3073484. URL: https://doi.org/10.1109/lgrs.2021.3073484.

[94] S. Modanesi, C. Massari, M. Bechtold, H. Lievens, A. Tarpanelli, L. Brocca, L. Zappa, and G. J. M. D. Lannoy. "Challenges and benefits of quantifying irrigation through the assimilation of Sentinel-1 backscatter observations into Noah-MP". In: *Hydrology and Earth System Sciences* 26.18 (Sept. 2022), pp. 4685–4706. DOI: 10.5194/hess-26-4685-2022. URL: https://doi.org/10.5194/hess-26-4685-2022.

[95] G. J. M. D. Lannoy and R. H. Reichle. "Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model". In: *Hydrology and Earth System Sciences* 20.12 (Dec. 2016), pp. 4895–4911. DOI: 10.5194/hess-20-4895-2016.

[96] G. J. M. D. Lannoy and R. H. Reichle. "Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation". In: *Journal of Hydrometeorology* 17.2 (Feb. 2016), pp. 669–691. DOI: 10.1175/jhm-d-15-0037.1.

[97] R. H. Reichle, G. J. M. D. Lannoy, Q. Liu, J. V. Ardizzone, A. Colliander, A. Conaty, W. Crow, T. J. Jackson, L. A. Jones, J. S. Kimball, R. D. Koster, S. P. Mahanama, E. B. Smith, A. Berg, S. Bircher, D. Bosch, T. G. Caldwell, M. Cosh, Á. González-Zamora, C. D. H. Collins, K. H. Jensen, S. Livingston, E. Lopez-Baeza, J. Martínez-Fernández, H. McNairn, M. Moghaddam, A. Pacheco, T. Pellarin, J. Prueger, T. Rowlandson, M. Seyfried, P. Starks, Z. Su, M. Thibeault, R. van der Velde, J. Walker, X. Wu, and Y. Zeng. "Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using In Situ Measurements". In: *Journal of Hydrometeorology* 18.10 (Sept. 2017), pp. 2621–2645. DOI: 10.1175/jhm-d-17-0063.1. URL: https://doi.org/10.1175/jhm-d-17-0063.1.

[98] R. H. Reichle, G. J. M. D. Lannoy, Q. Liu, R. D. Koster, J. S. Kimball, W. T. Crow, J. V. Ardizzone, P. Chakraborty, D. W. Collins, A. L. Conaty, M. Girotto, L. A. Jones, J. Kolassa, H. Lievens, R. A. Lucchesi, and E. B. Smith. "Global Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using Assimilation Diagnostics". In: *Journal of Hydrometeorology* 18.12 (Dec. 2017), pp. 3217–3237. DOI: 10.1175/jhm-d-17-0130.1. URL: https://doi.org/10.1175/jhm-d-17-0130.1.

[99] Y. Lu, S. C. Steele-Dunne, and G. J. M. D. Lannoy. "Improving Soil Moisture and Surface Turbulent Heat Flux Estimates by Assimilation of SMAP Brightness Temperatures or Soil Moisture Retrievals and GOES Land Surface Temperature Retrievals". In: *Journal of Hydrometeorology* 21.2 (Feb. 2020), pp. 183–203. DOI: 10.1175/jhm-d-19-0130.1. URL: https://doi.org/10.1175/jhm-d-19-0130.1.

[100] C. Albergel, C. Rüdiger, D. Carrer, J.-C. Calvet, N. Fritz, V. Naeimi, Z. Bartalis, and S. Hasenauer. "An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France". In: *Hydrology and Earth System Sciences* 13.2 (Feb. 2009), pp. 115–124. DOI: 10.5194/hess-13-115-2009.

[101] L. Brocca, F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer. “Improving runoff prediction through the assimilation of the ASCAT soil moisture product”. In: *Hydrology and Earth System Sciences* 14.10 (Oct. 2010), pp. 1881–1893. DOI: 10.5194/hess-14-1881-2010. URL: https://doi.org/10.5194/hess-14-1881-2010.

[102] C. Draper, J.-F. Mahfouf, J.-C. Calvet, E. Martin, and W. Wagner. “Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France”. In: *Hydrology and Earth System Sciences* 15.12 (Dec. 2011), pp. 3829–3841. DOI: 10.5194/hess-15-3829-2011.

[103] C. Albergel, P. de Rosnay, C. Gruhier, J. Muñoz-Sabater, S. Hasenauer, L. Isaksen, Y. Kerr, and W. Wagner. “Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations”. In: *Remote Sensing of Environment* 118 (Mar. 2012), pp. 215–226. DOI: 10.1016/j.rse.2011.11.017. URL: https://doi.org/10.1016/j.rse.2011.11.017.

[104] F. Aires, P. Weston, P. Rosnay, and D. Fairbairn. “Statistical approaches to assimilate ASCAT soil moisture information—I. Methodologies and first assessment”. In: *Quarterly Journal of the Royal Meteorological Society* (Mar. 2021). DOI: 10.1002/qj.3997.

[105] R. H. Reichle and R. D. Koster. “Bias reduction in short records of satellite soil moisture”. In: *Geophysical Research Letters* 31.19 (Oct. 2004). DOI: 10.1029/2004gl020938. URL: https://doi.org/10.1029/2004gl020938.

[106] L. Renzullo, A. van Dijk, J.-M. Perraud, D. Collins, B. Henderson, H. Jin, A. Smith, and D. McJannet. “Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment”. In: *Journal of Hydrology* 519 (Nov. 2014), pp. 2747–2762. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2014.08.008. URL: http://dx.doi.org/10.1016/j.jhydrol.2014.08.008.

[107] J.-F. Mahfouf, K. Bergaoui, C. Draper, F. Bouyssel, F. Taillefer, and L. Taseva. “A comparison of two off-line soil analysis schemes for assimilation of screen level observations”. In: *Journal of Geophysical Research* 114.D8 (Apr. 2009). DOI: 10.1029/2008jd011077. URL: https://doi.org/10.1029/2008jd011077.

[108] A. Barbu, J.-C. Calvet, J.-F. Mahfouf, C. Albergel, and S. Lafont. “Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study”. In: *Biogeosciences* 8.7 (2011), pp. 1971–1986.

[109] A. L. Barbu, J.-C. Calvet, J.-F. Mahfouf, and S. Lafont. “Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France”. In: 18.1 (Jan. 2014), pp. 173–192. DOI: 10.5194/hess-18-173-2014. URL: https://doi.org/10.5194/hess-18-173-2014.

[110] R. H. Reichle, J. P. Walker, R. D. Koster, and P. R. Houser. “Extended versus Ensemble Kalman Filtering for Land Data Assimilation”. In: *Journal of Hydrometeorology* 3.6 (Dec. 2002), pp. 728–740. DOI: 10.1175/1525-7541(2002)003<0728:evekff>2.0.co;2.

[111] W. T. Crow and E. F. Wood. “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”. In: *Advances in Water Resources* 26.2 (Feb. 2003), pp. 137–149. DOI: 10.1016/s0309-1708(02)00088-x. URL: https://doi.org/10.1016/s0309-1708(02)00088-x.

[112] G. Balsamo, J.-F. Mahfouf, S. Bélair, and G. Deblonde. “A Global Root-Zone Soil Moisture Analysis Using Simulated L-band Brightness Temperature in Preparation for the Hydros Satellite Mission”. In: *Journal of Hydrometeorology* 7.5 (Oct. 2006), pp. 1126–1146. DOI: 10.1175/jhm525.1. URL: https://doi.org/10.1175/jhm525.1.

[113] X. Han, H.-J. H. Franssen, C. Montzka, and H. Vereecken. “Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations”. In: *Water Resources Research* 50.7 (July 2014), pp. 6081–6105. DOI: 10.1002/2013wr014586. URL: https://doi.org/10.1002/2013wr014586.

[114] D. Fairbairn, A. L. Barbu, A. Napoly, C. Albergel, J.-F. Mahfouf, and J.-C. Calvet. “The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France”. In: *Hydrology and Earth System Sciences* 21.4 (Apr. 2017), pp. 2015–2033. DOI: 10.5194/hess-21-2015-2017.

[115] D. Leroux, J.-C. Calvet, S. Munier, and C. Albergel. “Using Satellite-Derived Vegetation Products to Evaluate LDAS-Monde over the Euro-Mediterranean Area”. In: *Remote Sensing* 10.8 (July 2018), p. 1199. DOI: 10.3390/rs10081199.

[116] E. P. W. Attema and F. T. Ulaby. “Vegetation modeled as a water cloud”. In: *Radio Science* 13.2 (Mar. 1978), pp. 357–364. DOI: 10.1029/rs013i002p00357.

[117] J.-P. Wigneron, Y. Kerr, P. Waldteufel, K. Saleh, M.-J. Escorihuela, P. Richaume, P. Ferrazzoli, P. de Rosnay, R. Gurney, J.-C. Calvet, J. Grant, M. Guglielmetti, B. Hornbuckle, C. Mätzler, T. Pellarin, and M. Schwank. “L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields”. In: *Remote Sensing of Environment* 107.4 (Apr. 2007), pp. 639–655. DOI: 10.1016/j.rse.2006.10.014. URL: https://doi.org/10.1016/j.rse.2006.10.014.

[118] R. H. Reichle, S. Q. Zhang, Q. Liu, C. S. Draper, J. Kolassa, and R. Todling. “Assimilation of SMAP Brightness Temperature Observations in the GEOS Land–Atmosphere Data Assimilation System”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 10628–10643. DOI: 10.1109/jstars.2021.3118595. URL: https://doi.org/10.1109/jstars.2021.3118595.

[119] B. A. Forman and R. H. Reichle. “Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.9 (Sept. 2015), pp. 4431–4441. DOI: 10.1109/jstars.2014.2325780.

[120] Y. Xue and B. A. Forman. “Comparison of passive microwave brightness temperature prediction sensitivities over snow-covered land in North America using machine learning algorithms and the Advanced Microwave Scanning Radiometer”. In: *Remote Sensing of Environment* 170 (Dec. 2015), pp. 153–165. DOI: 10.1016/j.rse.2015.09.009.

[121] J. M. Sabater, L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. D. Rosnay. “From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques”. In: *Journal of Hydrometeorology* 8.2 (Apr. 2007), pp. 194–206. DOI: 10.1175/jhm571.1. URL: https://doi.org/10.1175/jhm571.1.

[122] C. Albergel, E. Dutra, B. Bonan, Y. Zheng, S. Munier, G. Balsamo, P. de Rosnay, J. Muñoz-Sabater, and J.-C. Calvet. “Monitoring and Forecasting the Impact of the 2018 Summer Heatwave on Vegetation”. In: *Remote Sensing* 11.5 (Mar. 2019), p. 520. DOI: 10.3390/rs11050520. URL: https://doi.org/10.3390/rs11050520.

[123] C. Albergel, Y. Zheng, B. Bonan, E. Dutra, N. Rodríguez-Fernández, S. Munier, C. Draper, P. de Rosnay, J. Muñoz-Sabater, G. Balsamo, D. Fairbairn, C. Meurey, and J.-C. Calvet. “Data assimilation for continuous global assessment of severe conditions over terrestrial surfaces”. In: *Hydrology and Earth System Sciences* 24.9 (Sept. 2020), pp. 4291–4316. DOI: 10.5194/hess-24-4291-2020. URL: https://doi.org/10.5194/hess-24-4291-2020.

[124] X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. “Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a Deep Neural Network”. In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. DOI: 10.1016/j.rse.2022.113116. URL: https://doi.org/10.1016/j.rse.2022.113116.

[125] A. Mucia, B. Bonan, C. Albergel, Y. Zheng, and J.-C. Calvet. “Assimilation of passive microwave vegetation optical depth in LDAS-Monde: a case study over the continental USA”. In: *Biogeosciences* 19.10 (May 2022), pp. 2557–2581. ISSN: 1726-4189. DOI: 10.5194/bg-19-2557-2022. URL: http://dx.doi.org/10.5194/bg-19-2557-2022.

[126] S. V. Kumar, T. R. Holmes, R. Bindlish, R. de Jeu, and C. Peters-Lidard. “Assimilation of vegetation optical depth retrievals from passive microwave radiometry”. In: *Hydrology and Earth System Sciences* 24.7 (July 2020), pp. 3431–3450. ISSN: 1607-7938. DOI: 10.5194/hess-24-3431-2020. URL: http://dx.doi.org/10.5194/hess-24-3431-2020.

[127] D. Shamambo, B. Bonan, J.-C. Calvet, C. Albergel, and S. Hahn. “Interpretation of ASCAT Radar Scatterometer Observations Over Land: A Case Study Over Southwestern France”. In: *Remote Sensing* 11.23 (Nov. 2019), p. 2842. DOI: 10.3390/rs11232842.

[128] S. Modanesi, C. Massari, A. Gruber, H. Lievens, A. Tarpanelli, R. Morbidelli, and G. J. M. De Lannoy. “Optimizing a backscatter forward operator using Sentinel-1 data over irrigated land”. In: *Hydrology and Earth System Sciences* 25.12 (Dec.

2021), pp. 6283–6307. ISSN: 1607-7938. DOI: 10.5194/hess-25-6283-2021. URL: http://dx.doi.org/10.5194/hess-25-6283-2021.

[129] F. T. Ulaby, K. Sarabandi, K. Mcdonald, M. Whitt, and M. C. Dobson. "Michigan microwave canopy scattering model". In: *International Journal of Remote Sensing* 11.7 (July 1990), pp. 1223–1253. DOI: 10.1080/01431169008955090.

[130] P. Ferrazzoli, L. Guerriero, and G. Schiavon. "A vegetation classification scheme validated by model simulations". In: *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development.* IEEE, 1997. DOI: 10.1109/igarss.1997.608984.

[131] S.-B. Kim, M. Moghaddam, L. Tsang, M. Burgin, X. Xu, and E. G. Njoku. "Models of L-Band Radar Backscattering Coefficients Over Global Terrain for Soil Moisture Retrieval". In: 52.2 (Feb. 2014), pp. 1381–1396. DOI: 10.1109/tgrs.2013.2250980. URL: https://doi.org/10.1109/tgrs.2013.2250980.

[132] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat. "Deep learning and process understanding for data-driven Earth system science". In: *Nature* 566.7743 (Feb. 2019), pp. 195–204. DOI: 10.1038/s41586-019-0912-1.

[133] F. Chevallier, F. Chéruy, N. A. Scott, and A. Chédin. "A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget". In: *Journal of Applied Meteorology* 37.11 (Nov. 1998), pp. 1385–1397. DOI: 10.1175/1520-0450(1998)037<1385:annafa>2.0.co;2.

[134] S. Castruccio, D. J. McInerney, M. L. Stein, F. L. Crouch, R. L. Jacob, and E. J. Moyer. "Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs". In: *Journal of Climate* 27.5 (Feb. 2014), pp. 1829–1844. DOI: 10.1175/jcli-d-13-00099.1.

[135] J. Verrelst, N. Sabater, J. Rivera, J. Muñoz-Marí, J. Vicent, G. Camps-Valls, and J. Moreno. "Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis". In: *Remote Sensing* 8.8 (Aug. 2016), p. 673. DOI: 10.3390/rs8080673.

[136] I. Fer, R. Kelly, P. R. Moorcroft, A. D. Richardson, E. M. Cowdery, and M. C. Dietze. "Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation". In: *Biogeosciences* 15.19 (Oct. 2018), pp. 5801–5830. DOI: 10.5194/bg-15-5801-2018.

[137] N. Rodríguez-Fernández, P. de Rosnay, C. Albergel, P. Richaume, F. Aires, C. Prigent, and Y. Kerr. "SMOS Neural Network Soil Moisture Data Assimilation in a Land Surface Model and Atmospheric Impact". In: *Remote Sensing* 11.11 (June 2019), p. 1334. DOI: 10.3390/rs11111334.

[138] T. Corchia, B. Bonan, N. Rodríguez-Fernández, G. Colas, and J.-C. Calvet. "Assimilation of ASCAT Radar Backscatter Coefficients over Southwestern France". In: *Remote Sensing* 15.17 (Aug. 2023), p. 4258. DOI: 10.3390/rs15174258. URL: https://doi.org/10.3390/rs15174258.

[139] B. A. Forman, R. H. Reichle, and C. Derksen. “Estimating Passive Microwave Brightness Temperature Over Snow-Covered Land in North America Using a Land Surface Model and an Artificial Neural Network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.1 (Jan. 2014), pp. 235–248. DOI: 10.1109/tgrs.2013.2237913.

[140] Y. Xue, B. A. Forman, and R. H. Reichle. “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”. In: *Water Resources Research* 54.9 (Sept. 2018), pp. 6488–6509. DOI: 10.1029/2017wr022219.

[141] J. Noilhan and S. Planton. “A Simple Parameterization of Land Surface Processes for Meteorological Models”. In: *Monthly Weather Review* 117.3 (Mar. 1989), pp. 536–549. DOI: 10.1175/1520-0493(1989)117<0536:aspols>2.0.co;2.

[142] J. Noilhan and J.-F. Mahfouf. “The ISBA land surface parameterisation scheme”. In: *Global and Planetary Change* 13.1-4 (June 1996), pp. 145–159. DOI: 10.1016/0921-8181(95)00043-7.

[143] J.-C. Calvet, J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron. “An interactive vegetation SVAT model tested against data from six contrasting sites”. In: *Agricultural and Forest Meteorology* 92.2 (July 1998), pp. 73–95. DOI: 10.1016/s0168-1923(98)00091-4.

[144] J.-C. Calvet and J.-F. Soussana. “Modelling CO2-enrichment effects using an interactive vegetation SVAT scheme”. In: *Agricultural and Forest Meteorology* 108.2 (June 2001), pp. 129–152. DOI: 10.1016/s0168-1923(01)00235-0.

[145] B. Decharme, E. Martin, and S. Faroux. “Reconciling soil thermal and hydrological lower boundary conditions in land surface models”. In: *Journal of Geophysical Research: Atmospheres* 118.14 (2013), pp. 7819–7834.

[146] B. Decharme, A. Boone, C. Delire, and J. Noilhan. “Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions”. In: *Journal of Geophysical Research* 116.D20 (Oct. 2011). ISSN: 0148-0227. DOI: 10.1029/2011jd016002. URL: http://dx.doi.org/10.1029/2011JD016002.

[147] J. W. Deardorff. “A Parameterization of Ground-Surface Moisture Content for Use in Atmospheric Prediction Models”. In: *Journal of Applied Meteorology* 16.11 (Nov. 1977), pp. 1182–1185. DOI: 10.1175/1520-0450(1977)016<1182:apogsm>2.0.co;2.

[148] C. Albergel, C. Rüdiger, D. Carrer, J.-C. Calvet, N. Fritz, V. Naeimi, Z. Bartalis, and S. Hasenauer. “An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France.” In: *Hydrology & Earth System Sciences* 13.2 (2009).

[149] A. Boone, V. Masson, T. Meyers, and J. Noilhan. “The influence of the inclusion of soil freezing on simulations by a soil–vegetation–atmosphere transfer scheme”. In: *Journal of Applied Meteorology* 39.9 (2000), pp. 1544–1569.

[150] A. Boone, P. Samuelsson, S. Gollvik, A. Napoly, L. Jarlan, E. Brun, and B. Decharme. "The interactions between soil–biosphere–atmosphere land surface model with a multi-energy balance (ISBA-MEB) option in SURFEXv8 – Part 1: Model description". In: *Geoscientific Model Development* 10.2 (Feb. 2017), pp. 843–872. DOI: 10.5194/gmd-10-843-2017. URL: https://doi.org/10.5194/gmd-10-843-2017.

[151] V. Masson, P. L. Moigne, E. Martin, S. Faroux, A. Alias, R. Alkama, S. Belamari, A. Barbu, A. Boone, F. Bouyssel, P. Brousseau, E. Brun, J.-C. Calvet, D. Carrer, B. Decharme, C. Delire, S. Donier, K. Essaouini, A.-L. Gibelin, H. Giordani, F. Habets, M. Jidane, G. Kerdraon, E. Kourzeneva, M. Lafaysse, S. Lafont, C. L. Brossier, A. Lemonsu, J.-F. Mahfouf, P. Marguinaud, M. Mokhtari, S. Morin, G. Pigeon, R. Salgado, Y. Seity, F. Taillefer, G. Tanguy, P. Tulet, B. Vincendon, V. Vionnet, and A. Voldoire. "The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes". In: *Geoscientific Model Development* 6.4 (July 2013), pp. 929–960. DOI: 10.5194/gmd-6-929-2013. URL: https://doi.org/10.5194/gmd-6-929-2013.

[152] J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, and B. Piguet. "In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network". In: *2007 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2007. DOI: 10.1109/igarss.2007.4423019.

[153] P. Jarvis. "The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field". In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 273.927 (1976), pp. 593–610.

[154] S. Lafont, Y. Zhao, J.-C. Calvet, P. Peylin, P. Ciais, F. Maignan, and M. Weiss. "Modelling LAI, surface water and carbon fluxes at high-resolution over France: comparison of ISBA-A-gs and ORCHIDEE". In: *Biogeosciences* 9.1 (Jan. 2012), pp. 439–456. DOI: 10.5194/bg-9-439-2012. URL: https://doi.org/10.5194/bg-9-439-2012.

[155] A.-L. Gibelin, J.-C. Calvet, J.-L. Roujean, L. Jarlan, and S. O. Los. "Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products". In: *Journal of Geophysical Research* 111.D18 (2006). DOI: 10.1029/2005jd006691.

[156] J.-C. Calvet, S. Lafont, E. Cloppet, F. Souverain, V. Badeau, and C. Le Bas. "Use of agricultural statistics to verify the interannual variability in land surface models: a case study over France with ISBA-A-gs". In: *Geoscientific Model Development* 5.1 (Jan. 2012), pp. 37–54. ISSN: 1991-9603. DOI: 10.5194/gmd-5-37-2012. URL: http://dx.doi.org/10.5194/gmd-5-37-2012.

[157] C. Szczypta, J.-C. Calvet, C. Albergel, G. Balsamo, S. Boussetta, D. Carrer, S. Lafont, and C. Meurey. "Verification of the new ECMWF ERA-Interim reanalysis over France". In: *Hydrology and Earth System Sciences* 15.2 (Feb. 2011), pp. 647–666. ISSN: 1607-7938. DOI: 10.5194/hess-15-647-2011. URL: http://dx.doi.org/10.5194/hess-15-647-2011.

[158] K. Scipal. *Global soil moisture retrieval from ERS scatterometer data*. 2002.

[159] J. M. Stiles and K. Sarabandi. “Electromagnetic scattering from grassland. I. A fully phase-coherent scattering model”. In: *IEEE Transactions on Geoscience and Remote Sensing* 38.1 (2000), pp. 339–348.

[160] E. Blanken. *Novel use of advanced scatterometer (ASCAT) for forest fire monitoring*. 2023.

[161] S. Faroux, A. T. K. Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. L. Moigne. “ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models”. In: *Geoscientific Model Development* 6.2 (Apr. 2013), pp. 563–582. DOI: 10.5194/gmd-6-563-2013.

[162] S. Seabold and J. Perktold. “statsmodels: Econometric and statistical modeling with python”. In: *9th Python in Science Conference*. 2010.

[163] A. J. Teuling, C. M. Taylor, J. F. Meirink, L. A. Melsen, D. G. Miralles, C. C. van Heerwaarden, R. Vautard, A. I. Stegehuis, G.-J. Nabuurs, and J. V.-G. de Arellano. “Observational evidence for cloud cover enhancement over western European forests”. In: *Nature Communications* 8.1 (Jan. 2017). DOI: 10.1038/ncomms14065.

[164] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. “The ERA5 global reanalysis”. In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (June 2020), pp. 1999–2049. DOI: 10.1002/qj.3803. URL: https://doi.org/10.1002/qj.3803.

[165] J. Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural Networks* 61 (Jan. 2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

[166] Y. Hong, K.-L. Hsu, S. Sorooshian, and X. Gao. “Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System”. In: *Journal of Applied Meteorology* 43.12 (Dec. 2004), pp. 1834–1853. DOI: 10.1175/jam2173.1.

[167] Y. Tao, X. Gao, K. Hsu, S. Sorooshian, and A. Ihler. “A Deep Neural Network Modeling Framework to Reduce Bias in Satellite Precipitation Products”. In: *Journal of Hydrometeorology* 17.3 (Mar. 2016), pp. 931–945. DOI: 10.1175/jhm-d-15-0075.1.

[168] G. Tang, D. Long, A. Behrangi, C. Wang, and Y. Hong. “Exploring Deep Neural Networks to Retrieve Rain and Snow in High Latitudes Using Multisensor and Reanalysis Data”. In: *Water Resources Research* 54.10 (Oct. 2018), pp. 8253–8278. DOI: 10.1029/2018wr023830.

[169] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016.

[170] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[171] Y. Bengio. “Learning Deep Architectures for AI”. In: *Foundations and Trends in Machine Learning* 2.1 (2009), pp. 1–127. DOI: 10.1561/2200000006.

[172] M. Bianchini and F. Scarselli. “On the Complexity of Neural Network Classifiers: A Comparison Between Shallow and Deep Architectures”. In: *IEEE Transactions on Neural Networks and Learning Systems* 25.8 (Aug. 2014), pp. 1553–1565. DOI: 10.1109/tnnls.2013.2293637.

[173] J. Snoek, H. Larochelle, and R. P. Adams. “Practical bayesian optimization of machine learning algorithms”. In: *Advances in neural information processing systems*. 2012, pp. 2951–2959.

[174] L. Li, Z.-L. Yang, A. M. Matheny, H. Zheng, S. C. Swenson, D. M. Lawrence, M. Barlage, B. Yan, N. G. McDowell, and L. R. Leung. “Representation of Plant Hydraulics in the Noah-MP Land Surface Model: Model Development and Multiscale Evaluation”. In: *Journal of Advances in Modeling Earth Systems* 13.4 (Apr. 2021). DOI: 10.1029/2020ms002214.

[175] Y. Liu, A. G. Konings, D. Kennedy, and P. Gentine. “Global Coordination in Plant Physiological and Rooting Strategies in Response to Water Stress”. In: 35.7 (July 2021). DOI: 10.1029/2020gb006758. URL: https://doi.org/10.1029/2020gb006758.

[176] S. Bittner, M. Janott, D. Ritter, P. Köcher, F. Beese, and E. Priesack. “Functional–structural water flow model reveals differences between diffuse- and ring-porous tree species”. In: *Agricultural and Forest Meteorology* 158-159 (June 2012), pp. 80–89. DOI: 10.1016/j.agrformet.2012.02.005. URL: https://doi.org/10.1016/j.agrformet.2012.02.005.

[177] M. Janott, S. Gayler, A. Gessler, M. Javaux, C. Klier, and E. Priesack. “A one-dimensional model of water flow in soil-plant systems based on plant architecture”. In: *Plant and Soil* 341.1-2 (Nov. 2010), pp. 233–256. DOI: 10.1007/s11104-010-0639-0. URL: https://doi.org/10.1007/s11104-010-0639-0.

[178] C. A. Scott, W. G. Bastiaanssen, and M.-D. Ahmad. “Mapping root zone soil moisture using remotely sensed optical imagery”. In: *Journal of Irrigation and Drainage Engineering* 129.5 (2003), pp. 326–335.

[179] A. M. Matheny, G. Bohrer, S. R. Garrity, T. H. Morin, C. J. Howard, and C. S. Vogel. “Observations of stem water storage in trees of opposing hydraulic strategies”. In: *Ecosphere* 6.9 (Sept. 2015), art165. DOI: 10.1890/es15-00170.1. URL: https://doi.org/10.1890/es15-00170.1.

[180] A. M. Matheny, R. P. Fiorella, G. Bohrer, C. J. Poulsen, T. H. Morin, A. Wunderlich, C. S. Vogel, and P. S. Curtis. “Contrasting strategies of hydraulic control in two codominant temperate tree species”. In: *Ecohydrology* 10.3 (Dec. 2016). DOI: 10.1002/eco.1815. URL: https://doi.org/10.1002/eco.1815.

[181] P. C. Vermunt, S. Khabbazan, S. C. Steele-Dunne, J. Judge, A. Monsivais-Huertero, L. Guerriero, and P.-W. Liu. "Response of Subdaily L-Band Backscatter to Internal and Surface Canopy Water Dynamics". In: 59.9 (Sept. 2020), pp. 7322–7337. DOI: 10.1109/tgrs.2020.3035881. URL: https://doi.org/10.1109/tgrs.2020.3035881.

[182] S. Khabbazan, S. Steele-Dunne, P. Vermunt, J. Judge, M. Vreugdenhil, and G. Gao. "The influence of surface canopy water on the relationship between L-band backscatter and biophysical variables in agricultural monitoring". In: *Remote Sensing of Environment* 268 (Jan. 2022), p. 112789. DOI: 10.1016/j.rse.2021.112789. URL: https://doi.org/10.1016/j.rse.2021.112789.

[183] S. C. Steele-Dunne, S. Hahn, W. Wagner, and M. Vreugdenhil. "Towards Including Dynamic Vegetation Parameters in the EUMETSAT H SAF ASCAT Soil Moisture Products". In: *Remote Sensing* 13.8 (Apr. 2021), p. 1463. DOI: 10.3390/rs13081463.

[184] R. H. McCuen *et al. Hydrologic analysis and design*. Vol. 888. Pearson prentice hall Upper Saddle River, NJ, 2005.

[185] B. A. Forman and Y. Xue. "Machine learning predictions of passive microwave brightness temperature over snow-covered land using the special sensor microwave imager (SSM/I)". In: *Physical Geography* 38.2 (Sept. 2016), pp. 176–196. DOI: 10.1080/02723646.2016.1236606.

[186] D. Entekhabi, S. Yueh, P. E. O'Neill, K. H. Kellogg, A. Allen, R. Bindlish, M. Brown, S. Chan, A. Colliander, W. T. Crow, *et al.* "SMAP handbook–soil moisture active passive: Mapping soil moisture and freeze/thaw from space". In: (2014).

[187] R. Willis and W. W.-G. Yeh. "Groundwater systems planning and management". In: (1987).

[188] C. Rüdiger, C. Albergel, J.-F. Mahfouf, J.-C. Calvet, and J. P. Walker. "Evaluation of the observation operator Jacobian for leaf area index data assimilation with an extended Kalman filter". In: *Journal of Geophysical Research* 115.D9 (May 2010). DOI: 10.1029/2009jd012912. URL: https://doi.org/10.1029/2009jd012912.

[189] F. Chevallier and J.-F. Mahfouf. "Evaluation of the Jacobians of Infrared Radiation Models for Variational Data Assimilation". In: 40.8 (Aug. 2001), pp. 1445–1461. DOI: 10.1175/1520-0450(2001)040<1445:eotjoi>2.0.co;2. URL: https://doi.org/10.1175/1520-0450(2001)040%3C1445:eotjoi%3E2.0.co;2.

[190] L. Fillion and J.-F. Mahfouf. "Jacobians of an Operational Prognostic Cloud Scheme". In: 131.11 (Nov. 2003), pp. 2838–2856. DOI: 10.1175/1520-0493(2003)131<2838:joaopc>2.0.co;2. URL: https://doi.org/10.1175/1520-0493(2003)131%3C2838:joaopc%3E2.0.co;2.

[191] L. Garand, D. S. Turner, M. Larocque, J. Bates, S. Boukabara, P. Brunel, F. Chevallier, G. Deblonde, R. Engelen, M. Hollingshead, D. Jackson, G. Jedlovec, J. Joiner, T. Kleespies, D. S. McKague, L. McMillin, J.-L. Moncet, J. R. Pardo, P. J. Rayer, E. Salathe, R. Saunders, N. A. Scott, P. V. Delst, and H. Woolf. "Radiance and Jacobian intercomparison of radiative transfer models applied to HIRS and AMSU channels". In: 106.D20 (Oct. 2001), pp. 24017–24031. DOI: 10.1029/2000jd000184. URL: https://doi.org/10.1029/2000jd000184.

[192] G. Macelloni, S. Paloscia, P. Pampaloni, F. Marliani, and M. Gai. “The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops”. In: *IEEE Transactions on Geoscience and Remote Sensing* 39.4 (Apr. 2001), pp. 873–884. DOI: 10.1109/36.917914.

[193] G. Fontanelli, S. Paloscia, M. Zribi, and A. Chahbi. “Sensitivity analysis of X-band SAR to wheat and barley leaf area index in the Merguellil Basin”. In: *Remote Sensing Letters* 4.11 (Oct. 2013), pp. 1107–1116. DOI: 10.1080/2150704x.2013.842285.

[194] A. Tabatabaeenejad, M. Burgin, and M. Moghaddam. “Potential of L-Band Radar for Retrieval of Canopy and Subcanopy Parameters of Boreal Forests”. In: *IEEE Transactions on Geoscience and Remote Sensing* 50.6 (June 2012), pp. 2150–2160. DOI: 10.1109/tgrs.2011.2173349.

[195] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang. “The expressive power of neural networks: A view from the width”. In: *Advances in neural information processing systems* 30 (2017).

[196] X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. “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”. In: *Remote Sensing of Environment* 308 (July 2024), p. 114167. ISSN: 0034-4257. DOI: 10.1016/j.rse.2024.114167. URL: http://dx.doi.org/10.1016/j.rse.2024.114167.

[197] J. Kolassa, R. Reichle, and C. Draper. “Merging active and passive microwave observations in soil moisture data assimilation”. In: *Remote Sensing of Environment* 191 (Mar. 2017), pp. 117–130. DOI: 10.1016/j.rse.2017.01.015. URL: https://doi.org/10.1016/j.rse.2017.01.015.

[198] R. Daley. “The Lagged Innovation Covariance: A Performance Diagnostic for Atmospheric Data Assimilation”. In: *Monthly Weather Review* 120.1 (Jan. 1992), pp. 178–196. DOI: 10.1175/1520-0493(1992)120<0178:tlicap>2.0.co;2. URL: https://doi.org/10.1175/1520-0493(1992)120%3C0178:tlicap%3E2.0.co;2.

[199] W. A. Dorigo, W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver, A. Gruber, M. Drusch, S. Mecklenburg, P. van Oevelen, A. Robock, and T. Jackson. “The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements”. In: *Hydrology and Earth System Sciences* 15.5 (May 2011), pp. 1675–1698. DOI: 10.5194/hess-15-1675-2011. URL: https://doi.org/10.5194/hess-15-1675-2011.

[200] W. Dorigo, A. Xaver, M. Vreugdenhil, A. Gruber, A. Hegyiová, A. Sanchis-Dufau, D. Zamojski, C. Cordes, W. Wagner, and M. Drusch. “Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network”. In: *Vadose Zone Journal* 12.3 (Jan. 2013), vzj2012.0097. DOI: 10.2136/vzj2012.0097. URL: https://doi.org/10.2136/vzj2012.0097.

[201] A. Verger, F. Baret, and M. Weiss. “Near Real-Time Vegetation Monitoring at Global Scale”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.8 (Aug. 2014), pp. 3473–3481. DOI: 10.1109/jstars.2014.2328632. URL: https://doi.org/10.1109/jstars.2014.2328632.

[202] T. Kailath. “An innovations approach to least-squares estimation–Part I: Linear filtering in additive white noise”. In: *IEEE Transactions on Automatic Control* 13.6 (Dec. 1968), pp. 646–655. DOI: 10.1109/tac.1968.1099025. URL: https://doi.org/10.1109/tac.1968.1099025.

[203] A. Gruber, G. D. Lannoy, C. Albergel, A. Al-Yaari, L. Brocca, J.-C. Calvet, A. Colliander, M. Cosh, W. Crow, W. Dorigo, C. Draper, M. Hirschi, Y. Kerr, A. Konings, W. Lahoz, K. McColl, C. Montzka, J. Muñoz-Sabater, J. Peng, R. Reichle, P. Richaume, C. Rüdiger, T. Scanlon, R. van der Schalie, J.-P. Wigneron, and W. Wagner. “Validation practices for satellite soil moisture retrievals: What are (the) errors?” In: *Remote Sensing of Environment* 244 (July 2020), p. 111806. DOI: 10.1016/j.rse.2020.111806. URL: https://doi.org/10.1016/j.rse.2020.111806.

[204] R. H. Reichle, D. B. McLaughlin, and D. Entekhabi. “Hydrologic Data Assimilation with the Ensemble Kalman Filter”. In: *Monthly Weather Review* 130.1 (Jan. 2002), pp. 103–114. DOI: 10.1175/1520-0493(2002)130<0103:hdawte>2.0.co;2. URL: https://doi.org/10.1175/1520-0493(2002)130%3C0103:hdawte%3E2.0.co;2.

[205] G. Desroziers, L. Berre, B. Chapnik, and P. Poli. “Diagnosis of observation, background and analysis-error statistics in observation space”. In: *Quarterly Journal of the Royal Meteorological Society* 131.613 (Oct. 2005), pp. 3385–3396. DOI: 10.1256/qj.05.108. URL: https://doi.org/10.1256/qj.05.108.

[206] D. P. Dee. “Bias and data assimilation”. In: *Quarterly Journal of the Royal Meteorological Society* 131.613 (Oct. 2005), pp. 3323–3343. DOI: 10.1256/qj.05.137. URL: https://doi.org/10.1256/qj.05.137.

[207] R. H. Reichle, R. D. Koster, J. Dong, and A. A. Berg. “Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation”. In: *Journal of Hydrometeorology* 5.3 (June 2004), pp. 430–442. DOI: 10.1175/1525-7541(2004)005<0430:gsmfso>2.0.co;2. URL: https://doi.org/10.1175/1525-7541(2004)005%3C0430:gsmfso%3E2.0.co;2.

[208] W. Wagner, R. Lindorfer, T. Melzer, S. Hahn, B. Bauer-Marschallinger, K. Morrison, J.-C. Calvet, S. Hobbs, R. Quast, I. Greimeister-Pfeil, and M. Vreugdenhil. “Widespread occurrence of anomalous C-band backscatter signals in arid environments caused by subsurface scattering”. In: *Remote Sensing of Environment* 276 (July 2022), p. 113025. DOI: 10.1016/j.rse.2022.113025. URL: https://doi.org/10.1016/j.rse.2022.113025.

[209] S. de Roos, L. Busschaert, H. Lievens, M. Bechtold, and G. J. D. Lannoy. “Optimisation of AquaCrop backscatter simulations using Sentinel-1 observations”. In: *Remote Sensing of Environment* 294 (Aug. 2023), p. 113621. DOI: 10.1016/j.rse.2023.113621. URL: https://doi.org/10.1016/j.rse.2023.113621.

[210] X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. *Evaluating DNN-based observation operators in land data assimilation: Synthetic experiments with land surface model ISBA-A-gs in Southwestern France*. in preparation. 2024.

[211] R. Bindlish and A. P. Barros. "Parameterization of vegetation backscatter in radar-based, soil moisture estimation". In: *Remote Sensing of Environment* 76.1 (Apr. 2001), pp. 130–137. ISSN: 0034-4257. DOI: 10.1016/s0034-4257(00)00200-5. URL: http://dx.doi.org/10.1016/S0034-4257(00)00200-5.

[212] Q. Duan, S. Sorooshian, and V. Gupta. "Effective and efficient global optimization for conceptual rainfall-runoff models". In: *Water Resources Research* 28.4 (Apr. 1992), pp. 1015–1031. ISSN: 1944-7973. DOI: 10.1029/91wr02985. URL: http://dx.doi.org/10.1029/91WR02985.

[213] Q. Duan, S. Sorooshian, and V. K. Gupta. "Optimal use of the SCE-UA global optimization method for calibrating watershed models". In: *Journal of Hydrology* 158.3–4 (June 1994), pp. 265–284. ISSN: 0022-1694. DOI: 10.1016/0022-1694(94)90057-4. URL: http://dx.doi.org/10.1016/0022-1694(94)90057-4.

[214] D. Fairbairn, A. L. Barbu, J.-F. Mahfouf, J.-C. Calvet, and E. Gelati. "Comparing the ensemble and extended Kalman filters for in situ soil moisture assimilation with contrasting conditions". In: *Hydrology and Earth System Sciences* 19.12 (Dec. 2015), pp. 4811–4830. DOI: 10.5194/hess-19-4811-2015.

[215] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[216] F. Aires, C. Prigent, and W. B. Rossow. "Neural Network Uncertainty Assessment Using Bayesian Statistics: A Remote Sensing Application". In: *Neural Computation* 16.11 (Nov. 2004), pp. 2415–2458. DOI: 10.1162/0899766041941925. URL: http://dx.doi.org/10.1162/0899766041941925.

[217] V. M. Krasnopolsky. "Reducing uncertainties in neural network Jacobians and improving accuracy of neural network emulations with NN ensemble approaches". In: *Neural Networks* 20.4 (May 2007), pp. 454–461. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2007.04.008. URL: http://dx.doi.org/10.1016/j.neunet.2007.04.008.

[218] A. Danquechin Dorval, C. Meredieu, and F. Danjon. "Anchorage failure of young trees in sandy soils is prevented by a rigid central part of the root system with various designs". In: *Annals of Botany* 118.4 (July 2016), pp. 747–762. ISSN: 1095-8290. DOI: 10.1093/aob/mcw098. URL: http://dx.doi.org/10.1093/aob/mcw098.

[219] W. A. Dorigo, K. Scipal, R. M. Parinussa, Y. Y. Liu, W. Wagner, R. A. M. de Jeu, and V. Naeimi. "Error characterisation of global active and passive microwave soil moisture datasets". In: *Hydrology and Earth System Sciences* 14.12 (Dec. 2010), pp. 2605–2616. ISSN: 1607-7938. DOI: 10.5194/hess-14-2605-2010. URL: http://dx.doi.org/10.5194/hess-14-2605-2010.

[220] R. Quast, W. Wagner, B. Bauer-Marschallinger, and M. Vreugdenhil. “Soil moisture retrieval from Sentinel-1 using a first-order radiative transfer model—A case-study over the Po-Valley”. In: *Remote Sensing of Environment* 295 (Sept. 2023), p. 113651. ISSN: 0034-4257. DOI: 10.1016/j.rse.2023.113651. URL: http://dx.doi.org/10.1016/j.rse.2023.113651.

[221] M.-C. Su, Y.-Z. Hsieh, C.-H. Wang, and P.-C. Wang. “A Jacobian Matrix-Based Learning Machine and Its Applications in Medical Diagnosis”. In: *IEEE Access* 5 (2017), pp. 20036–20045. ISSN: 2169-3536. DOI: 10.1109/access.2017.2677458. URL: http://dx.doi.org/10.1109/ACCESS.2017.2677458.

[222] F. Latrémolière, S. Narayanappa, and P. Vojtěchovskỳ. “Estimating the Jacobian matrix of an unknown multivariate function from sample values by means of a neural network”. In: *arXiv preprint arXiv:2204.00523* (2022).

[223] Q. Liu and X. Liang. “Physics constraint Deep Learning based radiative transfer model”. In: *Optics Express* 31.17 (Aug. 2023), p. 28596. ISSN: 1094-4087. DOI: 10.1364/oe.493818. URL: http://dx.doi.org/10.1364/OE.493818.

[224] M. Vreugdenhil, S. Steele-Dunne, X. Shan, T. Kaminski, M. Aurela, E. Bueechi, W. Dorigo, W. Knorr, J. Lemmetyinen, N. Rodriguez-Fernandez, *et al.* “Seven Frozen Trees in Sodankyla: Relating ASCAT slope to water and carbon processes over a Boreal forest using in-situ, model and reanalysis data”. In: *EGU General Assembly Conference Abstracts*. 2023, EGU–12005.

[225] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian. “Accurate medium-range global weather forecasting with 3D neural networks”. In: *Nature* 619.7970 (July 2023), pp. 533–538. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06185-3. URL: http://dx.doi.org/10.1038/s41586-023-06185-3.

A

SUPPLEMENTARY MATERIAL

Table A.1.: Land Surface Variables from ISBA that are used as inputs to the DNN.

LSVs	unit	Full names
WG1	$m^3 m^{-3}$	surface soil moisture (0-0.01m)
WG2	$m^3 m^{-3}$	soil moisture (0.01-0.04m)
WG3	$m^3 m^{-3}$	soil moisture (0.04-0.1m)
WG4	$m^3 m^{-3}$	soil moisture (0.1-0.2m)
WG5	$m^3 m^{-3}$	soil moisture (0.2-0.4m)
WG6	$m^3 m^{-3}$	soil moisture (0.4-0.6m)
WG7	$m^3 m^{-3}$	soil moisture (0.6-0.8m)
WG8	$m^3 m^{-3}$	soil moisture (0.8-1.0m)
WG9	$m^3 m^{-3}$	soil moisture (1.0-1.5m)
WG10	$m^3 m^{-3}$	soil moisture (1.5-2.0m)
LAI	$m^2 m^{-2}$	leaf area index
RN	$W m^{-2}$	net radiation
LETR	$kg m^{-2} s^{-1}$	transpiration
WR	$kg m^{-2}$	vegetation interception reservoir water storage
GPP	$kg CO_2 m^{-2} s^{-1}$	gross primary productivity
g_s	$mm s^{-1}$	leaf stomatal conductance

Table A.6.: Information about the GPIs containing ISMN stations.

lat	lon	lat (ISMN)	lon (ISMN)	station	depth	depth (DA)	start	end
41.125	-5.625	41.23923	-5.54427	Zamarron	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.125	-5.375	41.19603	-5.35997	Canizal	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.125	-5.375	41.18264	-5.47708	Las_Bodegas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/03/2017	12/31/2019
41.125	-5.375	41.20548	-5.41558	Las_Eritas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.625	41.37338	-5.54850	Las_Arenas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.28546	-5.30003	La_Cruz_de_Elias	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.44649	-5.35870	Las_Brozas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.42413	-5.37403	Las_Victorias	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.35757	-5.33113	Llanos_de_la_Boveda	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.45586	-5.41100	Paredinas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.26504	-5.38049	Carretoro	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.39392	-5.32146	Casa_Periles	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.375	41.38134	-5.42922	El_Coto	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019

A

41.375	-5.125	41.30010	-5.24704	Concejo_del_Monte	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.375	-5.125	41.34709	-5.22474	Las_Vacas	0.0 ~0.05	[[0.01, 0.04], [0.04, 0.1]]	01/02/2017	12/31/2019
41.875	-0.625	41.78263	-0.72168	Penaflor	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	06/03/2017	12/31/2019
42.375	-0.875	42.30873	-0.80560	Aguero	0.0 ~0.06	[[0.01, 0.04], [0.04, 0.1]]	01/22/2017	12/31/2019
42.875	2.625	42.96000	2.53000	Mouthoumet	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
42.875	2.625	42.96000	2.53000	Mouthoumet	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
42.875	2.625	42.96000	2.53000	Mouthoumet	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.125	2.875	43.15000	2.95670	Narbonne	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.125	2.875	43.15000	2.95670	Narbonne	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.125	2.875	43.15000	2.95670	Narbonne	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	09/04/2019
43.375	1.875	43.44170	1.88000	Saint_Felix_de_Lauragais	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.375	1.875	43.44170	1.88000	Saint_Felix_de_Lauragais	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.375	1.875	43.44170	1.88000	Saint_Felix_de_Lauragais	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	10/16/2017	12/31/2019
43.375	1.875	43.44170	1.88000	Saint_Felix_de_Lauragais	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.375	3.375	43.43833	3.40333	Pezenas	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.375	3.375	43.43833	3.40333	Pezenas	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.375	3.375	43.43833	3.40333	Pezenas	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.375	3.375	43.43833	3.40333	Pezenas	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.625	-0.375	43.63970	-0.43500	Urgons	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.625	-0.375	43.63970	-0.43500	Urgons	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.625	-0.375	43.63970	-0.43500	Urgons	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	11/02/2019
43.625	-0.375	43.63970	-0.43500	Urgons	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.625	0.125	43.66640	0.22170	Peyrusse_Grande	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.625	0.125	43.66640	0.22170	Peyrusse_Grande	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.625	0.125	43.66640	0.22170	Peyrusse_Grande	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.625	0.125	43.66640	0.22170	Peyrusse_Grande	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.625	0.875	43.54720	0.88780	Lahas	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	02/13/2019
43.625	0.875	43.54720	0.88780	Lahas	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.625	0.875	43.54720	0.88780	Lahas	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.875	-0.125	43.99360	-0.04690	Creon_dArmagnac	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.875	-0.125	43.99360	-0.04690	Creon_dArmagnac	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	10/11/2017	12/31/2019
43.875	-0.125	43.99360	-0.04690	Creon_dArmagnac	0.3 ~0.3	[[0.2, 0.4]]	10/11/2017	12/31/2019
43.875	-0.125	43.99360	-0.04690	Creon_dArmagnac	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	10/11/2017	12/31/2019
43.875	0.375	43.97440	0.33610	Condom	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.875	0.375	43.97440	0.33610	Condom	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.875	0.375	43.97440	0.33610	Condom	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.875	0.375	43.97440	0.33610	Condom	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	02/14/2019
43.875	1.125	43.82500	1.17670	Savenes	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	03/28/2019
43.875	1.125	43.82500	1.17670	Savenes	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	01/01/2019
43.875	1.125	43.82500	1.17670	Savenes	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	03/23/2019
44.125	-0.875	44.14750	-0.84560	Sabres	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
44.125	-0.875	44.14750	-0.84560	Sabres	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
44.125	-0.875	44.14750	-0.84560	Sabres	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
44.125	-0.875	44.14750	-0.84560	Sabres	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.375	-0.875	44.49126	-0.75800	hillan2	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
44.375	-0.875	44.49126	-0.75800	hillan2	0.15 ~0.15	[[0.1, 0.2]]	01/01/2017	12/31/2019
44.375	-0.875	44.49126	-0.75800	hillan2	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
44.375	-0.875	44.49126	-0.75800	hillan2	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/12/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/12/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.34 ~0.34	[[0.2, 0.4]]	01/01/2017	12/12/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.21 ~0.21	[[0.2, 0.4]]	01/01/2017	12/12/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.45 ~0.45	[[0.4, 0.6]]	01/01/2017	12/12/2019
44.375	-0.875	44.47200	-0.76750	grandcal	0.56 ~0.56	[[0.4, 0.6]]	01/01/2017	12/12/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	11/23/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.55 ~0.55	[[0.4, 0.6]]	01/01/2017	12/31/2019
44.375	-0.625	44.46700	-0.72690	fraye	0.45 ~0.45	[[0.4, 0.6]]	01/01/2017	12/31/2019
44.875	-0.625	44.78990	-0.57650	parcmeteo	0.01 ~0.01	[[0.01, 0.04]]	01/01/2017	12/31/2019
44.875	-0.625	44.78990	-0.57650	parcmeteo	0.03 ~0.03	[[0.01, 0.04]]	01/01/2017	12/31/2019
44.875	-0.625	44.78990	-0.57650	parcmeteo	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
44.875	-0.625	44.78990	-0.57650	parcmeteo	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.875	4.125	43.79517	4.09083	Villevielle	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.875	4.125	43.79517	4.09083	Villevielle	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/20/2019
43.875	4.125	43.79517	4.09083	Villevielle	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.875	4.125	43.79517	4.09083	Villevielle	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
43.875	5.125	43.88367	5.16483	Cabrieres_dAvignon	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
43.875	5.125	43.88367	5.16483	Cabrieres_dAvignon	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
43.875	5.125	43.88367	5.16483	Cabrieres_dAvignon	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
43.875	5.125	43.88367	5.16483	Cabrieres_dAvignon	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.125	4.125	44.24300	4.01017	La_Grand_Combe	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
44.125	4.125	44.24300	4.01017	La_Grand_Combe	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	01/01/2017	12/31/2019
44.125	4.125	44.24300	4.01017	La_Grand_Combe	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.125	4.125	44.24300	4.01017	La_Grand_Combe	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
44.125	4.375	44.22200	4.34483	Mejannes-le-Clap	0.05 ~0.05	[[0.04, 0.1]]	06/11/2018	12/31/2019
44.125	4.375	44.22200	4.34483	Mejannes-le-Clap	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	06/11/2018	12/31/2019
44.125	4.375	44.22200	4.34483	Mejannes-le-Clap	0.3 ~0.3	[[0.2, 0.4]]	06/11/2018	11/01/2019
44.125	4.375	44.22200	4.34483	Mejannes-le-Clap	0.1 ~0.1	[[0.04, 0.1], [0.1, 0.2]]	06/11/2018	12/31/2019
44.625	4.625	44.62833	4.56667	Berzeme	0.3 ~0.3	[[0.2, 0.4]]	01/01/2017	12/31/2019
44.625	4.625	44.62833	4.56667	Berzeme	0.05 ~0.05	[[0.04, 0.1]]	01/01/2017	12/31/2019
44.625	4.625	44.62833	4.56667	Berzeme	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.05 ~0.05	[[0.04, 0.1]]	01/04/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
50.875	6.375	50.98920	6.32355	Gevenich	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019

50.875	6.375	50.98920	6.32355	Gevenich	0.05 ~0.05	[[0.04, 0.1]]	01/04/2017	12/31/2019
51.125	6.125	51.13274	6.16918	Wildenrath	0.2 ~0.2	[[0.1, 0.2], [0.2, 0.4]]	01/01/2017	12/31/2019
51.125	6.125	51.13274	6.16918	Wildenrath	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
51.125	6.125	51.13274	6.16918	Wildenrath	0.05 ~0.05	[[0.04, 0.1]]	01/04/2017	12/31/2019
51.125	6.125	51.13274	6.16918	Wildenrath	0.5 ~0.5	[[0.4, 0.6]]	01/01/2017	12/31/2019
51.875	4.875	51.93493	4.78968	3sh9gmem	0.0 ~0.1	[[0.01, 0.04], [0.04, 0.1]]	02/17/2017	06/19/2019
51.875	6.125	51.95936	6.14875	4ajxh4nj	0.0 ~0.1	[[0.01, 0.04], [0.04, 0.1]]	03/10/2017	11/30/2018

A

Table A.2.: The Pearson Correlation Coefficient between input LSVs and output ASCAT observables.

LSVs	backscatter	slope	curvature
WG1	0.64	-0.46	0.20
WG2	0.66	-0.47	0.21
WG3	0.63	-0.51	0.23
WG4	0.57	-0.55	0.26
WG5	0.47	-0.57	0.31
WG6	0.40	-0.54	0.34
WG7	0.35	-0.50	0.36
WG8	0.31	-0.46	0.37
WG9	0.23	-0.37	0.38
WG10	0.17	-0.28	0.36
RN	-0.23	0.43	-0.04
GPP	-0.19	0.41	-0.07
LETR	-0.21	0.44	-0.07
WR	0.29	0.07	-0.07
LAI	-0.21	0.62	-0.34
XRS	-0.09	0.38	-0.16

Table A.3.: The p-values corresponding to Table A.2.

LSVs	backscatter	slope	curvature
WG1	0.00	0.00	0.02
WG2	0.00	0.00	0.03
WG3	0.00	0.01	0.02
WG4	0.00	0.00	0.02
WG5	0.01	0.00	0.01
WG6	0.00	0.00	0.02
WG7	0.01	0.00	0.01
WG8	0.01	0.00	0.01
WG9	0.01	0.00	0.01
WG10	0.03	0.01	0.01
RN	0.02	0.00	0.04
GPP	0.02	0.00	0.03
LETR	0.02	0.00	0.03
WR	0.00	0.06	0.07
LAI	0.02	0.00	0.01
XRS	0.04	0.00	0.03

Table A.4.: The Pearson Correlation Coefficient between each pairs of input LSVs.

	WG1	WG2	WG3	WG4	WG5	WG6	WG7	WG8	WG9	WG10	RN	GPP	LETR	WR	LAI	XRS
WG1	1.00	0.99	0.95	0.89	0.80	0.73	0.67	0.62	0.52	0.43	-0.28	-0.19	-0.29	0.27	-0.35	-0.07
WG2	0.99	1.00	0.98	0.91	0.82	0.74	0.69	0.64	0.54	0.45	-0.28	-0.20	-0.29	0.25	-0.37	-0.09
WG3	0.95	0.98	1.00	0.96	0.88	0.80	0.74	0.69	0.59	0.50	-0.29	-0.21	-0.30	0.19	-0.41	-0.12
WG4	0.89	0.91	0.96	1.00	0.95	0.88	0.82	0.77	0.67	0.57	-0.30	-0.22	-0.30	0.12	-0.46	-0.15
WG5	0.80	0.82	0.88	0.95	1.00	0.97	0.92	0.88	0.78	0.68	-0.30	-0.21	-0.27	0.04	-0.49	-0.18
WG6	0.73	0.74	0.80	0.88	0.97	1.00	0.98	0.95	0.87	0.77	-0.28	-0.19	-0.24	-0.00	-0.49	-0.19
WG7	0.67	0.69	0.74	0.82	0.92	0.98	1.00	0.99	0.92	0.84	-0.25	-0.17	-0.21	-0.02	-0.46	-0.18
WG8	0.62	0.64	0.69	0.77	0.88	0.95	0.99	1.00	0.96	0.89	-0.22	-0.15	-0.18	-0.04	-0.44	-0.17
WG9	0.52	0.54	0.59	0.67	0.78	0.87	0.92	0.96	1.00	0.95	-0.16	-0.10	-0.12	-0.06	-0.38	-0.14
WG10	0.43	0.45	0.50	0.57	0.68	0.77	0.84	0.89	0.95	1.00	-0.10	-0.05	-0.06	-0.05	-0.27	-0.09
RN	-0.28	-0.28	-0.29	-0.30	-0.30	-0.28	-0.25	-0.22	-0.16	-0.10	1.00	0.80	0.79	-0.02	0.34	0.58
GPP	-0.19	-0.20	-0.21	-0.22	-0.21	-0.19	-0.17	-0.15	-0.10	-0.05	0.80	1.00	0.95	-0.05	0.50	0.72
LETR	-0.29	-0.29	-0.30	-0.30	-0.27	-0.24	-0.21	-0.18	-0.12	-0.06	0.79	0.95	1.00	-0.12	0.49	0.63
WR	0.27	0.25	0.19	0.12	0.04	-0.00	-0.02	-0.04	-0.06	-0.05	-0.02	-0.05	-0.12	1.00	0.25	0.13
LAI	-0.35	-0.37	-0.41	-0.46	-0.49	-0.49	-0.46	-0.44	-0.38	-0.27	0.34	0.50	0.49	0.25	1.00	0.60
XRS	-0.07	-0.09	-0.12	-0.15	-0.18	-0.19	-0.18	-0.17	-0.14	-0.09	0.58	0.72	0.63	0.13	0.60	1.00

Table A.5.: The corresponding p-values of Table A.4.

	WG1	WG2	WG3	WG4	WG5	WG6	WG7	WG8	WG9	WG10	RN	GPP	LETR	WR	LAI	XRS
WG1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05
WG2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03
WG3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02
WG4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.01
WG5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.09	0.00	0.01
WG6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.08	0.00	0.01
WG7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.01	0.00	0.07	0.00	0.02
WG8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.03	0.01	0.06	0.01	0.02
WG9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.01	0.06	0.03	0.06	0.01	0.03
WG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.04	0.04	0.05	0.06	0.01	0.04
RN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	1.00	0.00	0.00	0.22	0.00	0.00
GPP	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.06	0.04	0.00	1.00	0.00	0.11	0.00	0.00
LETR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.05	0.00	0.00	1.00	0.01	0.00	0.00
WR	0.00	0.00	0.00	0.04	0.09	0.08	0.07	0.06	0.06	0.06	0.22	0.11	0.01	1.00	0.00	0.02
LAI	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	1.00	0.00
XRS	0.05	0.03	0.02	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.00	0.00	0.00	0.02	0.00	1.00

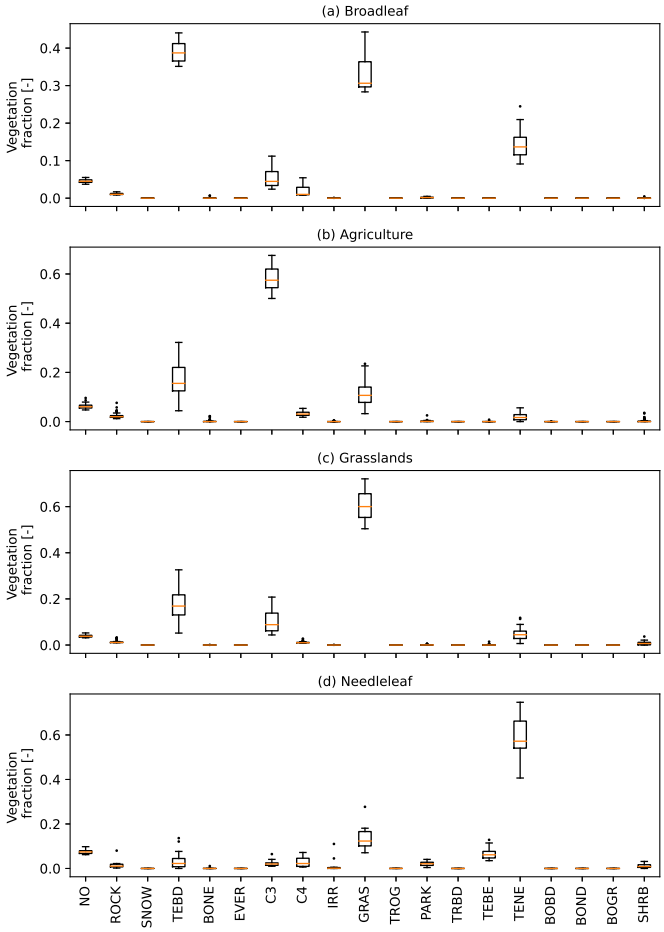

Figure A.1.: Boxplot illustrating the composition of the representative ROIs shown in Figure 1 (c) in terms of the ECOCLIMAP II land cover types. The box is drawn from Q1 to Q3 with a horizontal red line drawn in the middle to denote the median. The lower whisker is at the lowest datum above Q1-1.5*(Q3-Q1) and the upper whisker at the highest datum below Q3+1.5*(Q3-Q1), where Q1 and Q3 are the first and third quartiles. Outliers are indicated as black dots.

A

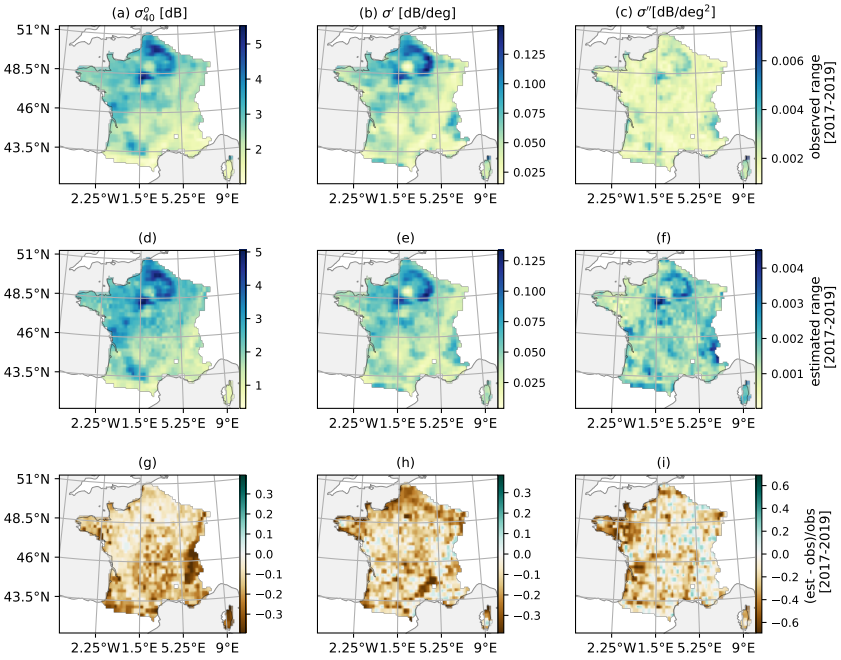

Figure A.2.: Maps illustrating range compression due to DNN in the validation period (2017-2019). The top row shows the observed range of the normalized backscatter (a), slope (b) and curvature (c). The corresponding ranges of the values estimated by the DNN are mapped in (d - f). The range compression for each variable is shown in (g - i).

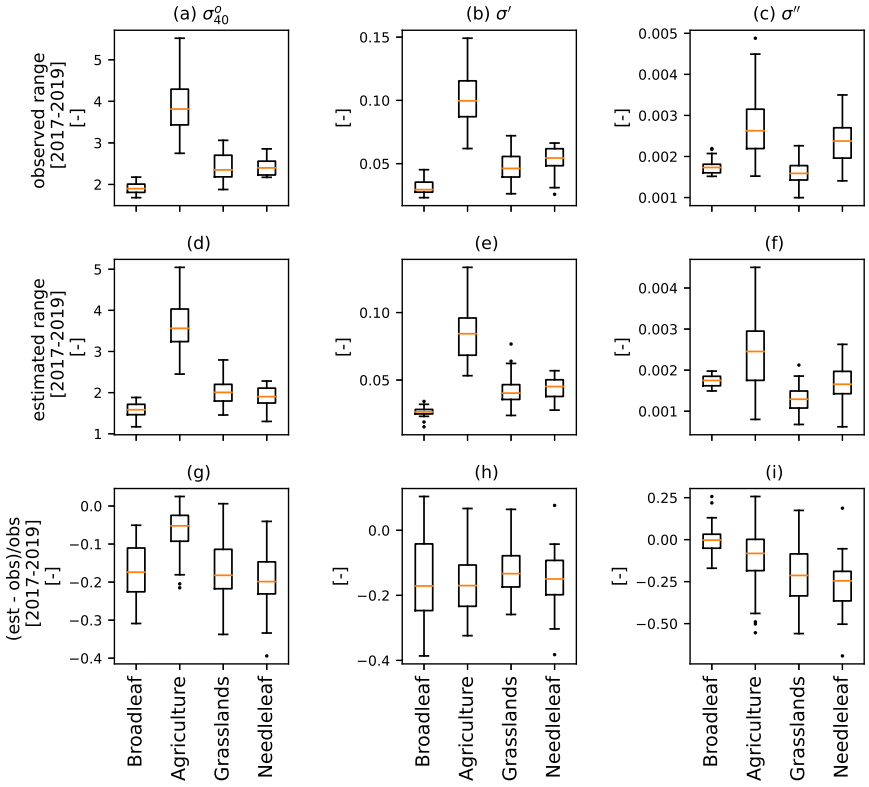

Figure A.3.: Observed range (a - c), range estimated (d - f), range compression (g - i) of normalized backscatter, slope and curvature averaged from 2017 to 2019.

A

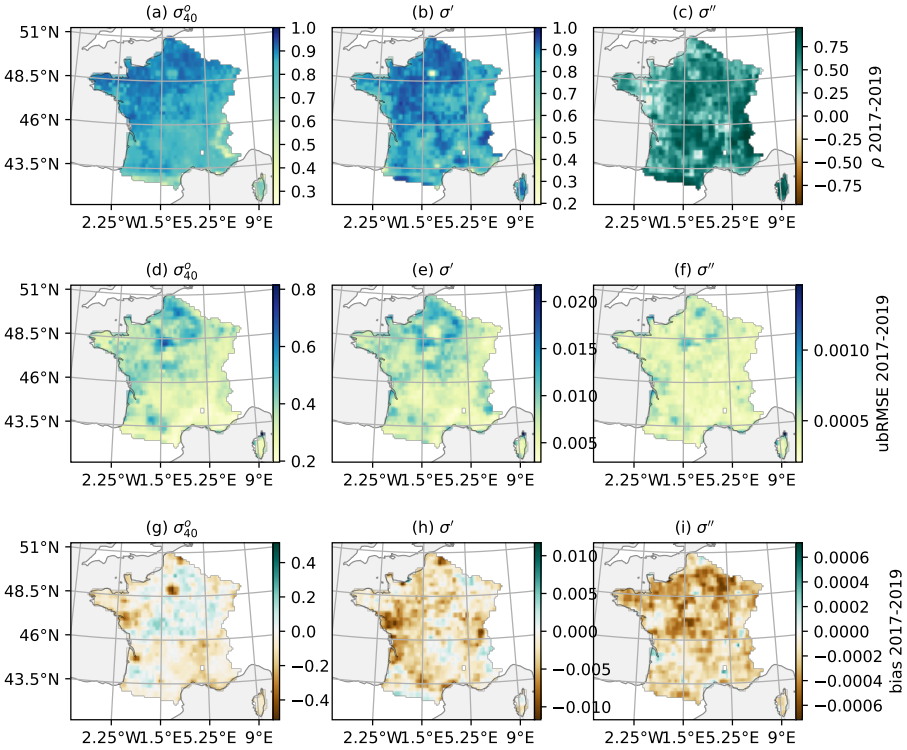

Figure A.4.: The performance of the DNN during the validation period (2017-2019). The columns show normalized backscatter, slope and curvature respectively. The rows show the Pearson correlation coefficient (a -c), unbiased RMSE (d - f) and bias (g - i). Note that the colorbars are different on each plot.

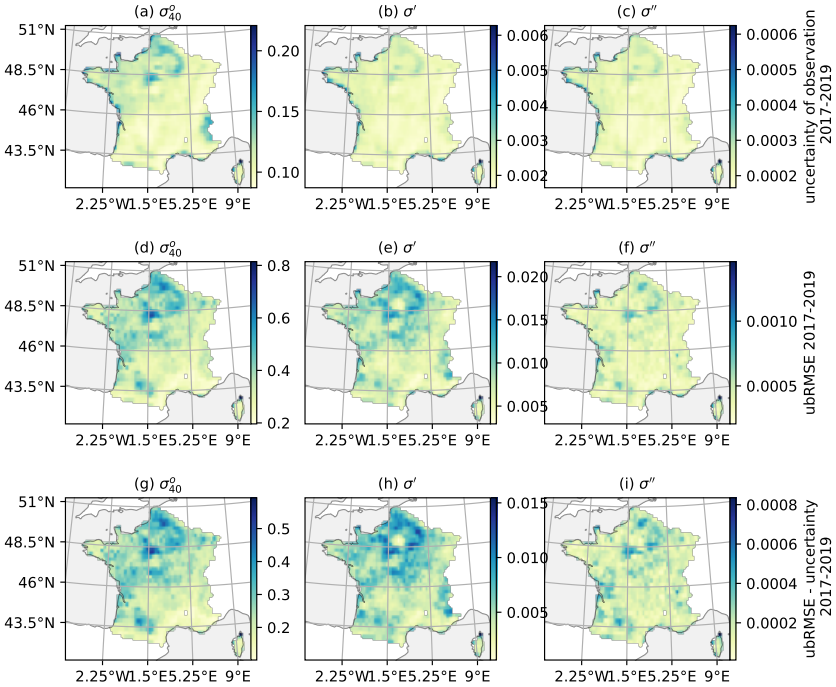

Figure A.5.: Comparison between the uncertainty in observed ASCAT normalized backscatter, slope and curvature and the ubRMSE of the DNN model. The uncertainty is averaged from 2017 to 2019.

A

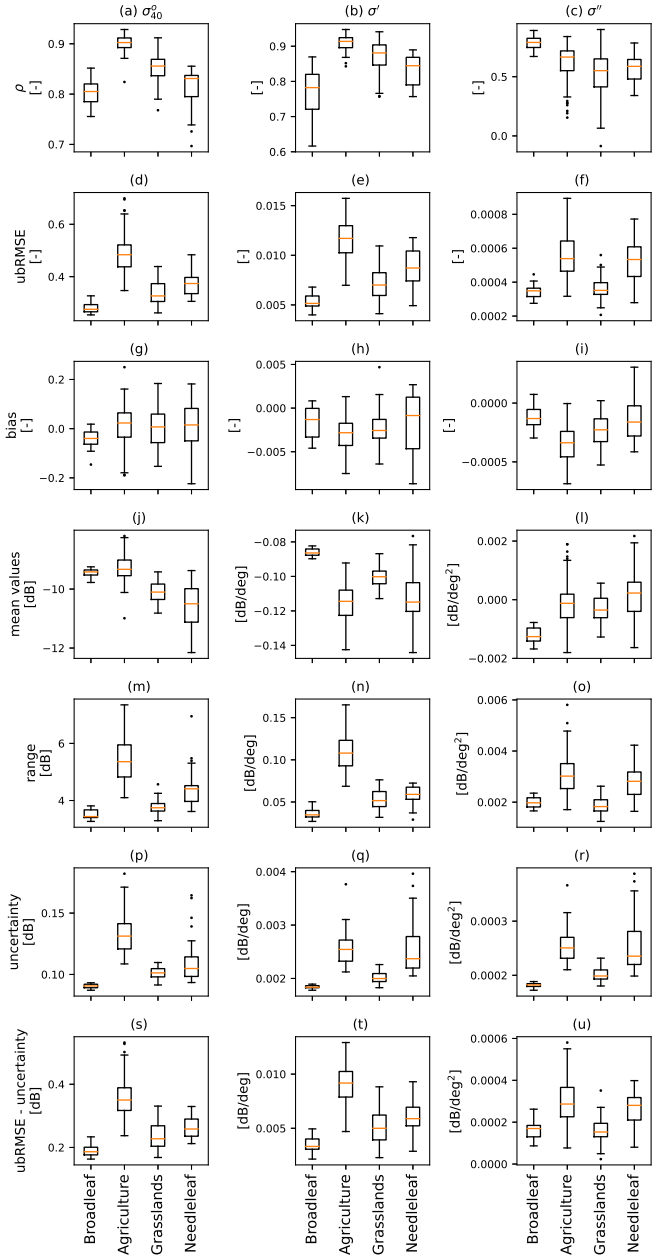

Figure A.6.: DNN performance during the validation period (2017-2019) for each of the four ROIs. The columns (left to right) show normalized backscatter, slope, and curvature respectively. The first five rows correspond to the Pearson correlation coefficient (a - c), unbiased RMSE (d - f), bias (g - i), uncertainty (j - l) and the difference between ubRMSE and uncertainty (m - o). The last two rows show the mean and range of the quantity of interest per vegetation type. The box extends from the lower to upper quartile values of the data, with a line at the median. The lower whisker is at the lowest datum above Q1 - 1.5*(Q3-Q1), and the upper whisker at the highest datum below Q3 + 1.5*(Q3-Q1), where Q1 and Q3 are the first and third quartiles. Outliers are indicated as grey dots.

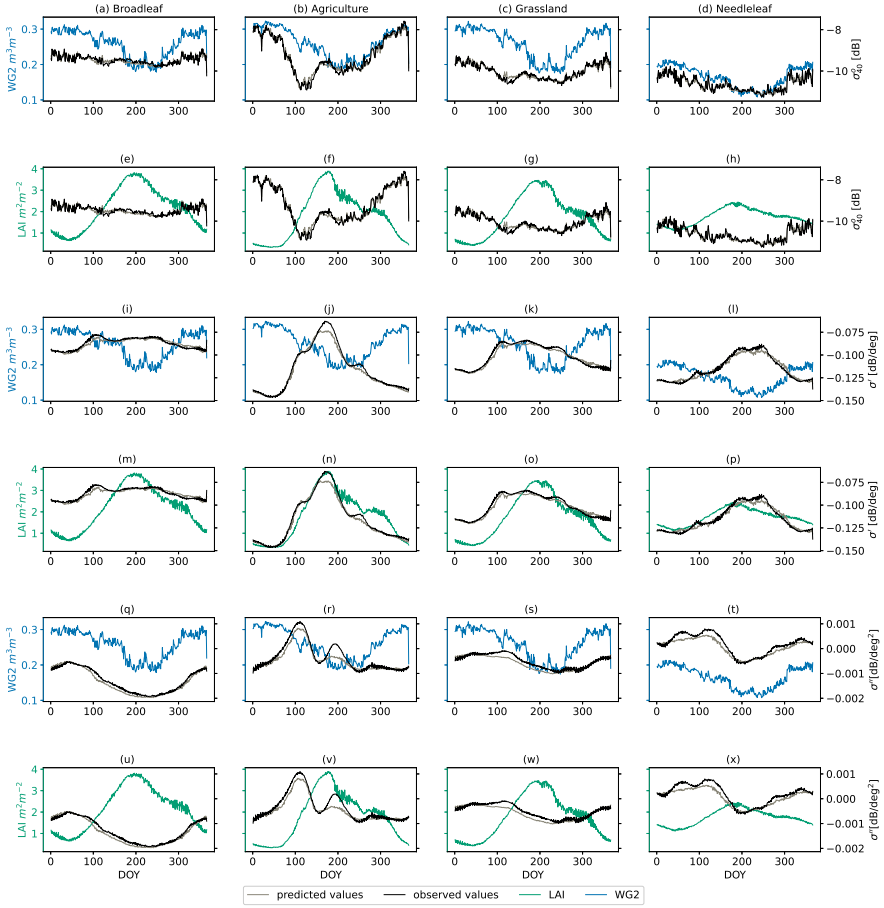

Figure A.7.: Time series of WG2, LAI compared to normalized backscatter, slope and curvature averaged from 2007 to 2019. The predicted value of normalized backscatter (slope, curvature) is the grey line and the observed normalized backscatter (slope, curvature) is the black line. Blue and green lines are LSVs of interest, e.g WG2 (a - d) and LAI (e - l), respectively.

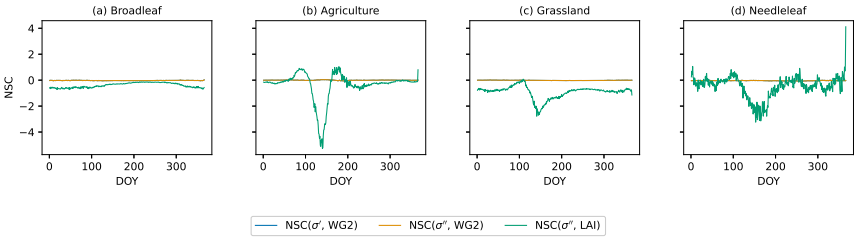

Figure A.8.: Seasonal cycle of NSC values (NSC(slope, WG2), NSC(curv, WG2) and NSC(curv, LAI)) in Broadleaf (a), Agriculture (b), Grassland (c) and Needleleaf (d) regions of interest. The values correspond to the median value among the GPIs, and were calculated using all data from 2007-2019. Note that NSC(slope, WG2) is close to zero, and indistinguishable from NSC(curvature, WG2).

Table A.7.: Statistics of performances of OL and DA using different observation errors (DA [0.15 dB, 0.002 dB/deg] and DA [0.33 dB, 0.005 dB/deg]) of GPIs containing stations Saint Felix de Lauragais, Lahas, Condom and Savenes.

	Experiment	Saint	Lahas	Condom	Savenes
LAI					
ubRMSE (m^2m^{-2})	OL	0.88	0.94	0.8	0.77
	DA [0.15 dB, 0.002 dB/deg]	0.88	1.1	0.8	0.82
	DA [0.33 dB, 0.005 dB/deg]	0.87	1	0.82	0.8
NIC ubRMSE	DA [0.15 dB, 0.002 dB/deg]	-0.11	12	0.96	6.2
	DA [0.33 dB, 0.005 dB/deg]	-1.2	8.4	2.8	3.7
ρ	OL	0.82	0.74	0.8	0.75
	DA [0.15 dB, 0.002 dB/deg]	0.65	0.49	0.49	0.6
	DA [0.33 dB, 0.005 dB/deg]	0.71	0.55	0.64	0.68
NIC ρ	DA [0.15 dB, 0.002 dB/deg]	-21	-35	-38	-19
	DA [0.33 dB, 0.005 dB/deg]	-14	-26	-19	-9.4
SSM (WG3)					
ubRMSE (m^3m^{-3})	OL	0.032	0.038	0.032	0.038
	DA [0.15 dB, 0.002 dB/deg]	0.034	0.044	0.035	0.04
	DA [0.33 dB, 0.005 dB/deg]	0.033	0.04	0.032	0.039
NIC ubRMSE	DA [0.15 dB, 0.002 dB/deg]	6	15	8.1	4.6
	DA [0.33 dB, 0.005 dB/deg]	4.5	4.5	-2.1	2.6
ρ	OL	0.86	0.85	0.85	0.81
	DA [0.15 dB, 0.002 dB/deg]	0.85	0.8	0.82	0.79
	DA [0.33 dB, 0.005 dB/deg]	0.85	0.84	0.85	0.8
NIC ρ	DA [0.15 dB, 0.002 dB/deg]	-1.2	-6.6	-2.9	-2.4
	DA [0.33 dB, 0.005 dB/deg]	-1.2	-1.8	1.1	-1.4
RZSM (WG4&WG5)					
ubRMSE (m^3m^{-3})	OL	0.026	0.044	0.082	0.025
	DA [0.15 dB, 0.002 dB/deg]	0.033	0.05	0.086	0.031
	DA [0.33 dB, 0.005 dB/deg]	0.03	0.047	0.082	0.024
NIC ubRMSE	DA [0.15 dB, 0.002 dB/deg]	26	14	4.6	26
	DA [0.33 dB, 0.005 dB/deg]	12	6.4	-0.57	-3.3
ρ	OL	0.88	0.9	0.9	0.81
	DA [0.15 dB, 0.002 dB/deg]	0.85	0.83	0.88	0.79
	DA [0.33 dB, 0.005 dB/deg]	0.86	0.88	0.91	0.83
NIC ρ	DA [0.15 dB, 0.002 dB/deg]	-3	-7.6	-3.2	-2.4
	DA [0.33 dB, 0.005 dB/deg]	-2.4	-2.3	0.58	2.9

A

Table A.8.: Values of diagnosticis of DA experiments using different observation errors (DA [0.15 dB, 0.002 dB/deg] and DA [0.33 dB, 0.005 dB/deg]) of GPIs containing stations Saint Felix de Lauragais, Lahas, Condom and Savenes.

	Experiment	Saint	Lahas	Condom	Savenes
σ^o_{40} [dB]					
mean(innovations)	DA [0.15 dB, 0.002 dB/deg]	0.046	-0.026	0.03	0.1
	DA [0.33 dB, 0.005 dB/deg]	0.066	-0.061	0.023	0.069
mean(residuals)	DA [0.15 dB, 0.002 dB/deg]	0.044	-0.0041	0.05	0.09
	DA [0.33 dB, 0.005 dB/deg]	0.074	-0.033	0.045	0.078
std(innovations)	DA [0.15 dB, 0.002 dB/deg]	0.38	0.36	0.38	0.35
	DA [0.33 dB, 0.005 dB/deg]	0.4	0.39	0.35	0.36
std(residuals)	DA [0.15 dB, 0.002 dB/deg]	0.25	0.25	0.3	0.27
	DA [0.33 dB, 0.005 dB/deg]	0.34	0.33	0.31	0.33
std(norm innovations)	DA [0.15 dB, 0.002 dB/deg]	1.9	1.8	1.9	1.8
	DA [0.33 dB, 0.005 dB/deg]	1.1	1.1	0.96	1
mean$\lvert O-A/O-F\rvert$	DA [0.15 dB, 0.002 dB/deg]	0.65	0.69	0.75	0.79
	DA [0.33 dB, 0.005 dB/deg]	0.87	0.86	0.9	0.9
C^k_{k+1} of innovations	DA [0.15 dB, 0.002 dB/deg]	0.13	0.28	0.44	0.34
	DA [0.33 dB, 0.005 dB/deg]	0.32	0.46	0.47	0.45
σ' [dB/deg]					
mean(innovations)	DA [0.15 dB, 0.002 dB/deg]	0.00085	0.001	0.00056	0.0011
	DA [0.33 dB, 0.005 dB/deg]	0.0019	0.001	0.00057	0.0022
mean(residuals)	DA [0.15 dB, 0.002 dB/deg]	0.00038	0.00082	0.00032	0.00094
	DA [0.33 dB, 0.005 dB/deg]	0.0014	0.00087	0.00044	0.002
std(innovations)	DA [0.15 dB, 0.002 dB/deg]	0.0034	0.0066	0.0057	0.0055
	DA [0.33 dB, 0.005 dB/deg]	0.0045	0.0075	0.0063	0.0069
std(residuals)	DA [0.15 dB, 0.002 dB/deg]	0.0019	0.0061	0.0049	0.005
	DA [0.33 dB, 0.005 dB/deg]	0.0035	0.0073	0.0057	0.0066
std(norm innovations)	DA [0.15 dB, 0.002 dB/deg]	1.1	3	2.3	2.4
	DA [0.33 dB, 0.005 dB/deg]	0.78	1.5	1.2	1.3
mean$\lvert O-A/O-F\rvert$	DA [0.15 dB, 0.002 dB/deg]	0.48	0.81	0.75	0.87
	DA [0.33 dB, 0.005 dB/deg]	0.76	0.94	0.86	0.96
C^k_{k+1} of innovations	DA [0.15 dB, 0.002 dB/deg]	0.35	0.92	0.78	0.84
	DA [0.33 dB, 0.005 dB/deg]	0.73	0.95	0.88	0.91

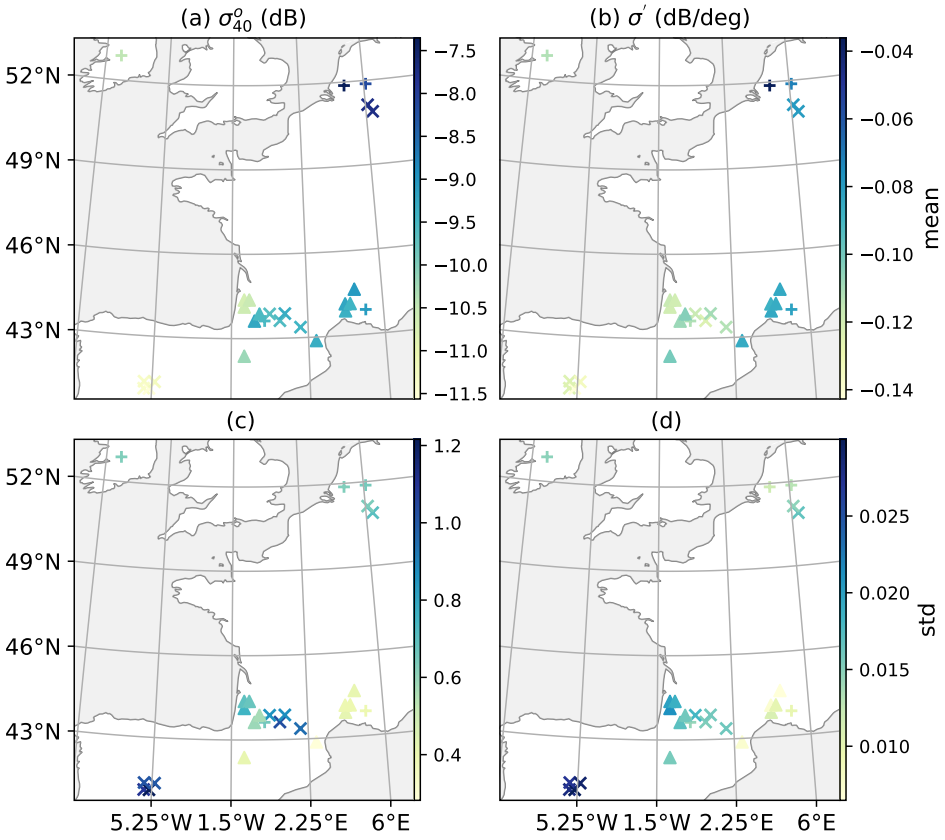

Figure A.9.: Mean and standard deviation values of σ^o_{40} and σ' from 2017 to 2019 in GPIs containing ISMN stations.

A

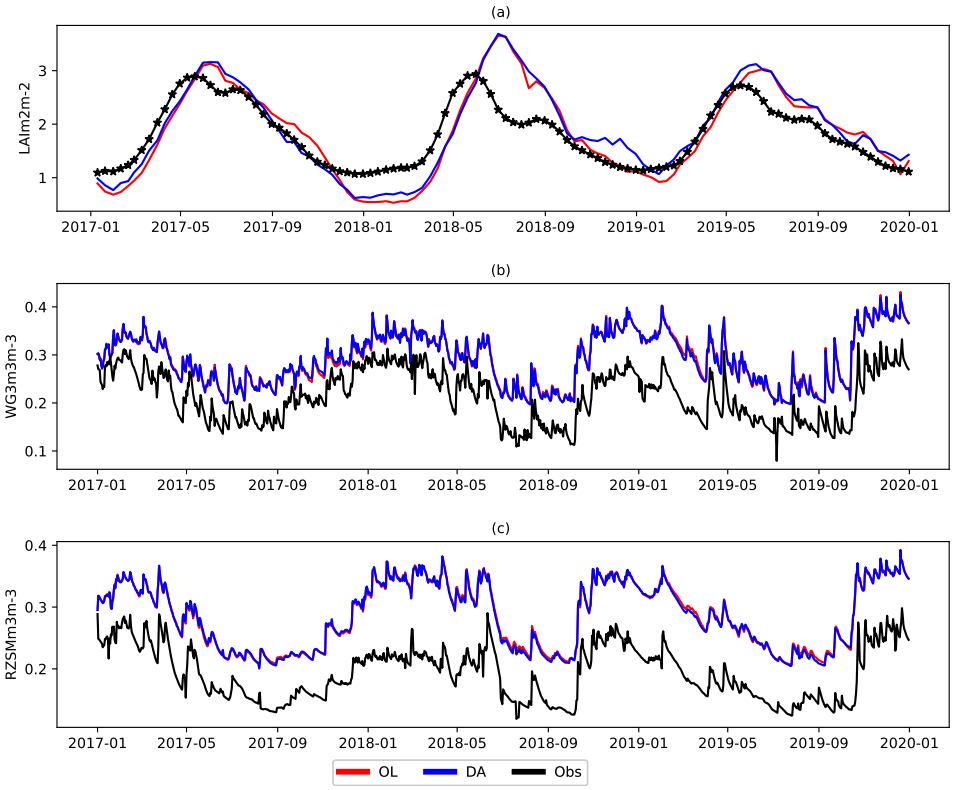

Figure A.10.: Time series of LAI, WG2, WG3, root zone soil moisture (RZSM) for observation, OL and DA(σ^o_{40}, σ') averaged on GPIs of Grasslands from 2017 to 2019.

Table A.9.: Domain median statistics of performances of OL and DA experiments using different observation errors.

	Experiment	ubRMSE	bias	ρ
WG2 m^3m^{-3}				
	OL	0.0265	0.0106	0.951
	DA(WCM)	0.0213	0.00774	0.966
	DA(DNN[WG2,LAI])	0.0230	0.00938	0.963
	DA(DNN[multiple WGs,LAI])	0.0241	0.00940	0.961
WG4 m^3m^{-3}				
	OL	0.0142	0.00249	0.969
	DA(WCM)	0.0132	0.00183	0.973
	DA(DNN[WG2,LAI])	0.0132	0.00243	0.973
	DA(DNN[multiple WGs,LAI])	0.0133	0.00249	0.972
WG6 m^3m^{-3}				
	OL	0.00881	-0.000269	0.982
	DA(WCM)	0.00793	-0.000672	0.985
	DA(DNN[WG2,LAI])	0.00802	-0.00023	0.985
	DA(DNN[multiple WGs,LAI])	0.00846	-0.000343	0.983
WG8 m^3m^{-3}				
	OL	0.00777	-0.000502	0.983
	DA(WCM)	0.00692	-0.000926	0.986
	DA(DNN[WG2,LAI])	0.00693	-0.000538	0.986
	DA(DNN[multiple WGs,LAI])	0.00775	-0.000433	0.983
LAI m^2m^{-2}				
	OL	0.229	-0.000074	0.98
	DA(WCM)	0.212	-0.0647	0.984
	DA(DNN[WG2,LAI])	0.174	0.0039	0.988
	DA(DNN[multiple WGs,LAI])	0.185	-0.00767	0.987

A

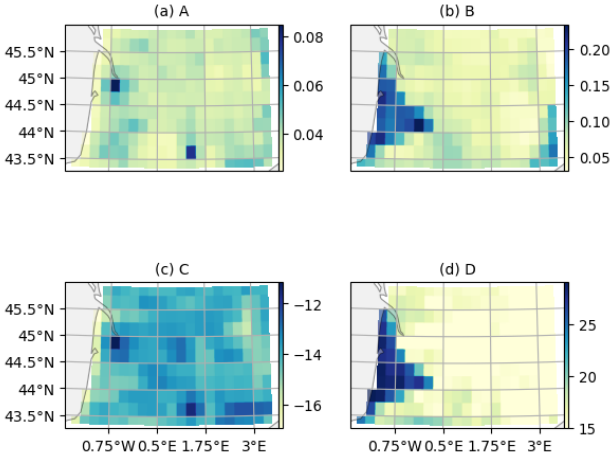

Figure A.11.: Parameters of Water Cloud Model, calibrated from 2007 to 2016 using WG2 and LAI from ISBA OL simulations as input and ASCAT normalized backscatter as output.

ACKNOWLEDGEMENTS

In the past few years, I enjoyed my Ph.D. journey in Delft. This experience would not be possible without the valuable support of my promoter and daily supervisor, Prof. Susan Steele-Dunne. I extend my deepest gratitude to Prof. Susan Steele-Dunne for providing me with this extraordinary opportunity to conduct research within her group. I cannot express enough appreciation for her trust and guidance. I also enjoyed working with her, her prompt and insightful feedback during our discussions consistently sparked new ideas for my research. Beyond my Ph.D. research, Susan also provided immense support for my other academic pursuits.

I am also grateful for the enlightening discussions with Dr. Paco Lopez Dekker, whose sharp insights have significantly contributed to my research. I am also grateful to Prof. Wolfgang Wagner and Dr. Sebastian Hahn from TU Wien, Dr. Jean-Christophe Calvet, Dr. Clement Albergel (now at ESA), and Dr. Bertrand Bonan from Meteo France. My research cannot be done without your valuable input. Discussions with you were always beneficial.

Additionally, I would like to thank my colleagues at TU Delft: Dr. Banafsheh Abdollahi, Dr. Jianzhi Dong, Manuel Huber, Dr. Paco Frantzen, Dr. Saeed Khabbazan, Dr. Sumit Meshram, Dr. Paul Vermunt, Dr. Vineet Kumar, and Dr. Yang Lu, all of whom have provided valuable suggestions and helps throughout my Ph.D. life. I particularly cherished the experiences of the fieldwork with Vineet and Saeed.

I am also thankful for the technical support from Dr. Yongjun Zheng (NUIST), who assisted me with software issues during the early stages of my Ph.D, and Ou Ku and Sonja Georgievska (eScience center), who assisted me with coding issues during my whole Ph.D research. I am also thankful for the generous and timely help from online communities like StackOverflow and Github. My research would be delayed without those discussions.

Meanwhile, many friends brought joy and comfort into my daily life. I would like to thank, in alphabetical order: Chenxu Ma, Cong Wang, Enshan Chen, Haoxiang Zhang, Heye Huang, Hongliang Ma, Huilan Zheng, Jiaming Wen, Jingsen Lian, Junkai Lan, Longmiao Gao, Longlong Yu, Meng Jiang, Mengbin Zhu, Mingzhe Jiang, Nan Bai, Qiaohui Lin, Ruiqing Zhou, Shiduo Xin, Shiyu Wang, Siyuan Wang, Tianhao He, Wanxue Chen, Wenxiao Hou, Xin Tian, Xinyu Liu, Xinyan Ruan, Yaoyi Chen, Yangyang Fan, Yanggu Zheng, Yuancheng Xu, Yuanzhao Ding, Yi Zheng, Yixian Shen, Yiyi Zhang, Yonghui Wang, Yujia Zhai, Yulin Zhang, Yueer Li, Yuzheng Song, Ziao Wang, Zijun Chen, Zhewen Hu, Zheyi Zeng (with Nan Jiang), Zhen Wang, Zhengxuan Liu, Zhenwu Wang, and Zhixin Feng.

Specifically, I want to thank Longmiao Gao, Nan Bai, Shiduo Xin, Tianhao He, Yi Zheng, Yonghui Wang, Zhewen Hu, and Ziao Wang. I enjoyed the time I spent with you. I also greatly benefited from discussions about stocks, options, and other financial knowledge with Meng Jiang, Huilan Zheng, Xinyu Liu, and Yanggu Zheng. Thank you, Shiduo Xin,

Yulin Zhang, Zhen Wang, and Zijun Chen, for the refreshing conversations about mathematics, brain teasers, and algorithms.

During this valuable journey, I deeply appreciate the discussions with seniors (in alphabetical order): Hongliang Ma, Jianzhi Dong, Meng Jiang, Nan Bai, Saeed Khabbazan, and Zheyi Zeng (with Nan Jiang). I learned a lot from your life experiences. I would also like to extend my appreciation to my old friends back in China: Jie Hou, Jie Li, Shan Zhou, Tiankai Pang, and Shuhuai Zhang. I will never forget our valuable friendships.

I must also extend my thanks to my mentors from my undergraduate studies and during my gap period: Dr. Di Long (Tsinghua University), Dr. Hanbo Yang (Tsinghua University), and Dr. Ying Sun (Cornell University). Interestingly, I found the period before the start of my PhD, i.e. my gap period, more suffering than the Ph.D. journey because of the visa issues. Their support was essential during my gap period. Meanwhile, many history books provided me with solace and inspiration, particularly "The Records of the Three Kingdoms." The heroes in this book, with their unwavering bravery and strategic acumen, motivated me to push forward despite the difficulties. Additionally, I am profoundly grateful to my friends who faced similar challenges (visa issues) and stood by me. Their experiences and support were valuable, providing both comfort and strength. I thank them for their unwavering companionship and encouragement, which played a crucial role in my journey.

Lastly, but most importantly, I reserve my deepest gratitude to my parents and my girlfriend, Yujie Bian. The journey so far can be as challenging as it is rewarding, often accompanied by pressures and obstacles that test one's patience. During these times, the support from them becomes not just a comfort but a necessity. They offered me the foundation of love, encouragement, and stability that allowed me to pursue my ambitions fearlessly. Their belief in my abilities is the cornerstone of my success in this PhD journey.

Xu Shan

Delft

CURRICULUM VITÆ

Xu SHAN

03-08-1996	Born in Shanxi Province, China.

EDUCATION

2020–2024	PhD. Civil Engineering and Geosciences Delft University of Technogology, Delft, the Netherlands *Promotor:* Prof. dr. Susan Steele-Dunne *Co-Promotor:* Dr. Paco Lopez Dekker
2014–2018	Bachelor of Engineering, Hydraulic Engineering Tsinghua University
2015–2018	Bachelor of Science, Pure and Applied Mathematics Tsinghua University

AWARDS

2016	Award for Excellence in Rural Research, Tsinghua University
2016	Gold Award of Social Practice (Followers), Tsinghua University
2015	National Endeavor Fellowship (Top 10%), Tsinghua University
2015	Academic Excellence Award (Top 10%), Tsinghua University
2015	Social Practice Award (Leader), Tsinghua University

TALKS AND ORAL PRESENTATIONS

Shan, X., Steele-Dunne, S. C., Hahn, S., Wagner, W., Bonan, B., Albergel, C., ... & Ku, O. (2024, April). On quantifying the uncertainty of the Jacobians of the data-driven methods in the data assimilation framework. NWO NAC 2024 meeting.

Shan, X., Steele-Dunne, S. C., Hahn, S., Wagner, W., Bonan, B., Albergel, C., ... & Ku, O. (2023, December). Constraining plant water dynamics in land surface model through

the assimilation ASCAT backscatter and slope at ISMN stations over Western Europe. EUMETSAT 2023 meeting.

Shan, X., Steele-Dunne, S. C., Hahn, S., Wagner, W., Bonan, B., Albergel, C., ... & Ku, O. (2023, June). Constraining Plant Water Dynamics in Land Surface Model through the Assimilating ASCAT Normalized Backscatter and Slope at ISMN Stations over Western Europe. The 3rd annual Land Data Assimilation (DA) Community Virtual Workshop.

Shan, X., Steele-Dunne, S. C., Hahn, S., Wagner, W., Bonan, B., Albergel, C., ... & Ku, O. (2023, March). Constraining plant water dynamics in land surface model through the assimilation ASCAT backscatter and slope at ISMN stations over Western Europe. NWO NAC 2023 meeting.

Shan, X., Steele-Dunne, S., Hahn, S., Wagner, W., Bonan, B., ... & Ku, O. (2023). Joint assimilation of ASCAT backscatter and slope into the ISBA land surface model at ISMN stations over Western Europe (No. EGU23-8489). Copernicus Meetings.

Shan, X., Steele-Dunne, S. C., Hahn, S., Wagner, W., Bonan, B., Albergel, C., ... & Ku, O. (2022, December). Assimilating ASCAT backscatter and slope into land surface model: case studies at ISMN stations in Europe. In Fall Meeting 2022. AGU.

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., ... & Georgievska, S. (2022). Constraining plant water dynamics in land surface model by assimilating ASCAT dynamic vegetation parameters (No. EGU22-10176). Copernicus Meetings.

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., ... & Georgievska, S. (2021, December). Towards constraining water and carbon cycle processes with radar data through data assimilation. the Netherlands Center for Geodesy and Geo-Informatics (NCG).

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., ... & Georgievska, S. (2021, December). Towards constraining water and carbon cycle processes with radar data through assimilation. In AGU Fall Meeting Abstracts (Vol. 2021, pp. B13B-08).

LIST OF PUBLICATIONS

3. X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. *Evaluating DNN-based observation operators in land data assimilation: Synthetic experiments with land surface model ISBA-A-gs in Southwestern France.* in preparation. 2024

2. X. Shan, S. Steele-Dunne, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, and O. Ku. "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". In: *Remote Sensing of Environment* 308 (July 2024), p. 114167. ISSN: 0034-4257. DOI: 10.1016/j.rse.2024.114167. URL: http://dx.doi.org/10.1016/j.rse.2024.114167

1. X. Shan, S. Steele-Dunne, M. Huber, S. Hahn, W. Wagner, B. Bonan, C. Albergel, J.-C. Calvet, O. Ku, and S. Georgievska. "Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a Deep Neural Network". In: *Remote Sensing of Environment* 279 (Sept. 2022), p. 113116. DOI: 10.1016/j.rse.2022.113116. URL: https://doi.org/10.1016/j.rse.2022.113116