

Assimilation of Sentinel-1 Backscatter to Update AquaCrop Estimates of Soil Moisture and Crop Biomass

**Key Points:**

- First assessment of Sentinel-1 backscatter data assimilation in a crop model integrated into NASA's Land Information System
- Co- and cross-polarization backscatter observations were used to update regional AquaCrop soil moisture and biomass estimates, respectively
- Sentinel-1 data assimilation resulted in improved soil moisture estimates, but further research is needed for optimal vegetation updating

Correspondence to:

S. de Roos,
shannon.de.roos@vub.be

Citation:

de Roos, S., Bechtold, M., Busschaert, L., Lievens, H., & De Lannoy, G. J. M. (2024). Assimilation of Sentinel-1 backscatter to update AquaCrop estimates of soil moisture and crop biomass. *Journal of Geophysical Research: Biogeosciences*, 129, e2024JG008231. <https://doi.org/10.1029/2024JG008231>

Received 8 MAY 2024
Accepted 2 OCT 2024

Author Contributions:

Conceptualization: Michel Bechtold, Gabrielle J. M. De Lannoy
Data curation: Shannon de Roos, Michel Bechtold, Hans Lievens
Formal analysis: Shannon de Roos
Funding acquisition: Gabrielle J. M. De Lannoy
Investigation: Shannon de Roos
Methodology: Shannon de Roos, Michel Bechtold, Hans Lievens, Gabrielle J. M. De Lannoy
Project administration: Gabrielle J. M. De Lannoy
Software: Shannon de Roos, Michel Bechtold, Louise Busschaert, Gabrielle J. M. De Lannoy
Supervision: Michel Bechtold, Gabrielle J. M. De Lannoy
Validation: Shannon de Roos
Visualization: Shannon de Roos

Shannon de Roos¹ , Michel Bechtold² , Louise Busschaert² , Hans Lievens^{2,3} , and Gabrielle J. M. De Lannoy² 

¹Department of Water and Climate, Vrije Universiteit Brussel, Brussel, Belgium, ²Department of Earth and Environmental Sciences, KU Leuven, Heverlee, Belgium, ³Department of Environment, Ghent University, Ghent, Belgium

Abstract This study assesses the potential of regional microwave backscatter data assimilation (DA) in AquaCrop for the first time, using NASA's Land Information System. The objective is to assess whether the assimilation setup can improve surface soil moisture (SSM) and crop biomass estimates. SSM and crop biomass simulations from AquaCrop were updated using Sentinel-1 synthetic aperture radar observations, over three regions in Europe in two separate DA experiments. The first experiment concerned updating SSM using VV-polarized backscatter and the corrections were propagated via the model to the biomass. In the second experiment, the DA setup was extended by also updating the biomass with VH-polarized backscatter. SSM was evaluated with local in situ data and with downscaled Soil Moisture Active Passive (SMAP) retrievals for all cropland grid cells, whereas crop biomass was compared to SMAP vegetation optical depth and the Copernicus dry matter productivity. The assimilation showed mixed results for root mean square error and Pearson's correlation, with slight overall improvements in the (anomaly) correlations of updated SSM relative to independent in situ and satellite data. By contrast, the biomass estimates obtained with backscatter DA did not agree better with reference data sets. Overall, the SSM evaluation showed that there is potential in using Sentinel-1 backscatter for assimilation in AquaCrop, but the present setup was not able to improve crop biomass estimates. Our study reveals how the complex interaction between SSM, crop biomass and backscatter affect the impact and performance of DA, offering insight into ways to optimize DA for crop growth estimation.

Plain Language Summary This study evaluates if using observations from a microwave satellite, Sentinel-1 (S1) can improve model simulations of a crop model AquaCrop, specifically for both soil moisture and crop biomass, over different regions in Europe. For each day in which S1 observations were available over the region, the modeled soil moisture and biomass were "updated" based on these observations, which over time is expected to reduce the model error and uncertainty. This iterative process is called data assimilation (DA) and was executed in a model framework called NASA's Land Information System. Two DA experiments were held. In the first DA experiment, only soil moisture was updated by S1 observations, but the changes in soil moisture were expected to also affect the biomass simulations compared to no DA model runs. In the second DA experiment, both the soil moisture and biomass were updated with S1 data. When comparing the results with independent data sets, the assimilation showed mixed results. The soil moisture showed slight improvements after DA, but the biomass estimates did not improve. Given the complexity of S1 data over agricultural areas, more research is required to optimally perform DA before this setup is able to improve crop growth estimation.

1. Introduction

Soil moisture and vegetation biomass are key components for crop monitoring. In low- and mid-latitude regions, crop development in agricultural systems is largely dependent on soil moisture availability, whereas radiation is more critical at higher latitudes. Yet, the timing and amount of precipitation, and thus the soil moisture availability, are crucial for rain-fed crops over the globe. With the expected growing impact of changing climatic temperature and water distributions, regional perspectives for environmental management and economic planning are needed (Faivre et al., 2009; Jones et al., 2017). In this light, crop growth models are continuously adapted to adequately help inform agronomists and policymakers. Most crop models were initially developed for point-based applications; assuming homogeneous conditions of soil, crop, and management over the field. These crop models are now more frequently used in a spatially distributed way, also known as spatialized or gridded crop models (Busschaert et al., 2022; de Roos et al., 2021; Elliott et al., 2015; Holzworth et al., 2015; Manivasagam & Rozenstein, 2020).

© 2024. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Writing – original draft:

Shannon de Roos

Writing – review & editing:

Michel Bechtold, Louise Busschaert,

Hans Lievens, Gabrielle J. M. De Lannoy

The quality of spatialized crop models is largely dependent on the quality of the spatial input data (Hansen & Jones, 2000), for example, meteorological forcings, crop and soil information. These spatial data sets, especially continental or global products, are difficult to validate with ground truth data. For agricultural fields, the decisions based on the local farmer's knowledge of crop and field management add another level of heterogeneity in space and time that is difficult to capture in large-scale spatialized model setups. At the field scale, spatialized models will therefore mostly be more uncertain than point-based model applications (X. Jin et al., 2018).

One way to correct for model errors is data assimilation (hereafter DA). Satellite-based DA can help reduce both spatial and temporal errors in regional models by updating (state) variables or model parameters. DA into gridded crop models is commonly done using optical-derived indices, such as leaf area index (LAI) or the normalized difference vegetation index (NDVI), to improve model estimates of crop biomass (Claverie et al., 2009; H. Huang et al., 2022) or yield (H. Fang et al., 2008; N. Jin et al., 2022; Ziliani et al., 2022). Alternatively, the assimilation of microwave-based data can be used to update either soil moisture or vegetation. Microwave remote sensing has a significant advantage over optical imagery of being barely affected by cloud cover and providing more regular observations in space and time. Microwaves with longer wavelengths (e.g., L-band) have deeper penetration abilities, making them very suitable for soil moisture updating in crop models (Chakrabarti et al., 2014). Shorter wavelengths, such as C- and X-band, are more sensitive to vegetation and can be valuable for updating crop phenology or production (Reville et al., 2013; Shan et al., 2022; Steele-Dunne et al., 2019; Wiseman et al., 2014), but have also been used in relation to retrieving soil moisture (Bhogapurapu et al., 2022), irrigation (Modanesi et al., 2022) and agricultural drought assessment (Van Emmerik et al., 2015).

The extraction of information from microwave observations is not trivial; the signal is influenced by a variety of land surface elements that are not constant over time. Over croplands, the signal is affected by a combination of plant structure, height and volume, soil surface roughness, texture, and density, and the effects also depend on the local incidence angle. Vegetation scattering generally increases with growing canopy volume and is further affected by the orientation of the vegetation (McNairn & Shang, 2016). Hence, the sensitivity of the microwave signal to soil moisture decreases during the plant growing season, and in case of dense vegetation, the C-band microwave sensitivity to soil moisture can even become negligible.

ESA's and Copernicus Sentinel-1 (hereafter S1) satellite mission provides active microwave synthetic aperture radar (SAR) observations in C-band at a relatively high spatial resolution (5–20 m) in co-polarization (vertical-vertical; VV) and cross-polarization (vertical-horizontal; VH). The VV co-polarized backscatter has been proven to be sensitive to soil moisture (Palmisano et al., 2020; Rains et al., 2021), especially when there is little signal disturbance by vegetation. The S1 observations have therefore been used in several studies to observe spatial and temporal soil moisture patterns (Balzano et al., 2021; Bauer-Marschallinger et al., 2018). The cross-polarization data often show a higher sensitivity to vegetation, as plant structures can cause an increase in volume scattering within the vegetation medium, multiple scattering on vegetation elements, and vegetation-ground scattering interactions, enhancing the depolarization of the incident wave (McNairn & Shang, 2016). Due to their higher complexity, the assimilation of SAR data for updating vegetation is often combined with optically derived indices (Allies et al., 2022; Betbeder et al., 2016; J. Huang et al., 2019; Prévot et al., 2003). The use of SAR alone to update both soil moisture and vegetation biomass in crop models is still relatively unexplored.

For this study, a pilot version of the integration of the Food and Agricultural Organization (FAO) crop model AquaCrop V7.0 in NASA's Land Information System (LIS) is evaluated for the first time, as an assimilation system using Sentinel-1 backscatter observations. LIS is a software framework developed for high-performance land surface modeling and DA (Kumar et al., 2008; Peters-Lidard et al., 2007). LIS V7.4 currently only contains LSMs, which describe large-scale interactions of water and energy fluxes between the Earth's surface and atmosphere, based on detailed physically-based relationships (Fisher & Koven, 2020). Applications of LSMs in agriculture, especially for simulating dynamic crop growth, are not common because the vegetation is considered as a component of the water and energy cycle with limited quantitative information about crop yield. The advantages of current crop models over most traditional LSMs are that they consider more detailed crop development stages, quantify crop production in terms of crop biomass and yield for a large set of crops, and include various land management options. Including a crop model into the LIS framework therefore provides new opportunities for agricultural research.

This study explores the possibilities of S1 backscatter observations for regional crop model updating in a pilot version of the integration of AquaCrop in LIS for the first time and evaluates opportunities to optimize the system.

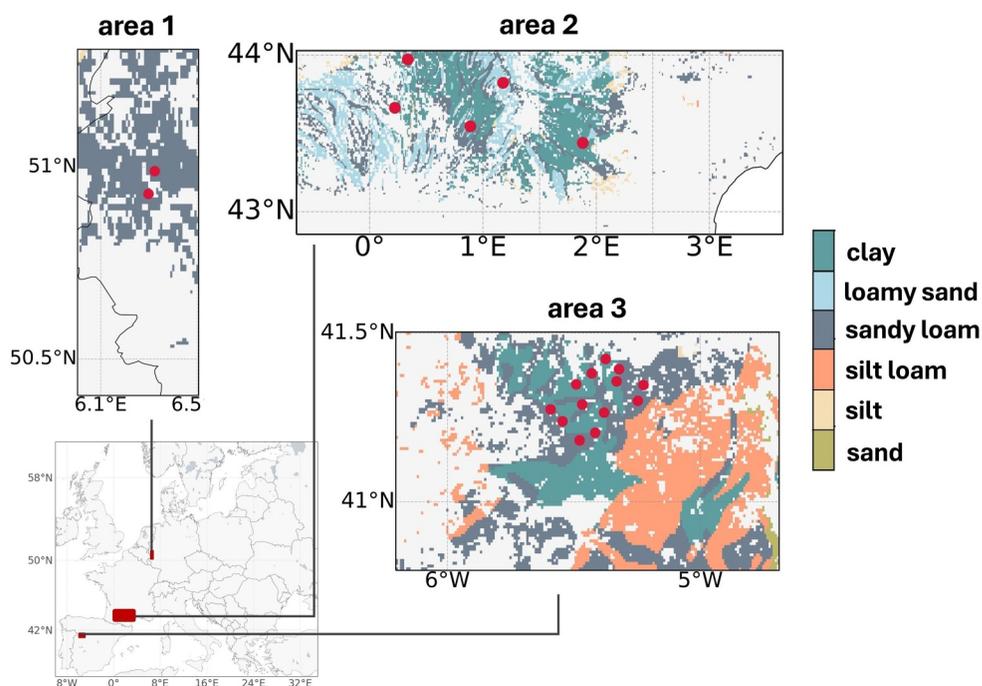


Figure 1. Locations of the three study areas with an indication of their soil textural classes, and in situ points as red dots. A mask has been applied to exclude all land cover classes other than non-irrigated cropland.

The main objectives of the S1 DA experiments are to assess if predictions of soil moisture and biomass can be improved by: (a) soil moisture state updating using backscatter observations in VV polarization, and (b) updating both soil moisture and biomass via joint (but independent) assimilation of respective S1 VV and VH polarized observations.

This paper is structured as follows; Section 2 covers the methodology, describing the model system setup, used data sets, and evaluation metrics. In Section 3, the results of the two experiments are analyzed and discussed. The DA diagnostics are analyzed in Section 4 to gain a better understanding of the system setup, and in Section 5 the study setup and model system are evaluated, and the potentials and limitations are discussed. Finally, a concluding Section 6 summarizes the main findings.

2. Materials and Methods

2.1. Study Setup

2.1.1. Study Areas

Three European study areas were selected based on their different climates and available in situ surface soil moisture (hereafter SSM_{ismn}) measurements from the International Soil Moisture Network (ISMN; Dorigo et al., 2011), as presented in Figure 1. From north to south, the first area (area 1) is situated in the Ruhr region of West Germany, covering in situ data from the TERENO network (Bogena et al., 2018; Zacharias et al., 2011), the second area (area 2) covers southern France with in situ points from the SMOSMANIA network (Albergel et al., 2008; Calvet et al., 2007) and the third area (area 3) is situated in northwest Spain and includes in situ data from the REMEDHUS stations (González-Zamora et al., 2019). The selected areas vary strongly in extent and were simulated with a spatial resolution of 30 arcsec (1 km at equator).

For each area, a mask was applied for pixels containing less than 50% rain-fed agriculture, as defined by the CORINE land cover map of 2018 (Büttner, 2014). The same generic C_3 crop from de Roos et al. (2021) and de Roos et al. (2023), was applied in this study for all areas, which showed to be capable of estimating the daily biomass productivity in Europe over time with good Pearson correlations (mean of 0.8). This approach was selected as an area of 1-km spatial resolution often contains a mix of crops in European fields instead of a single

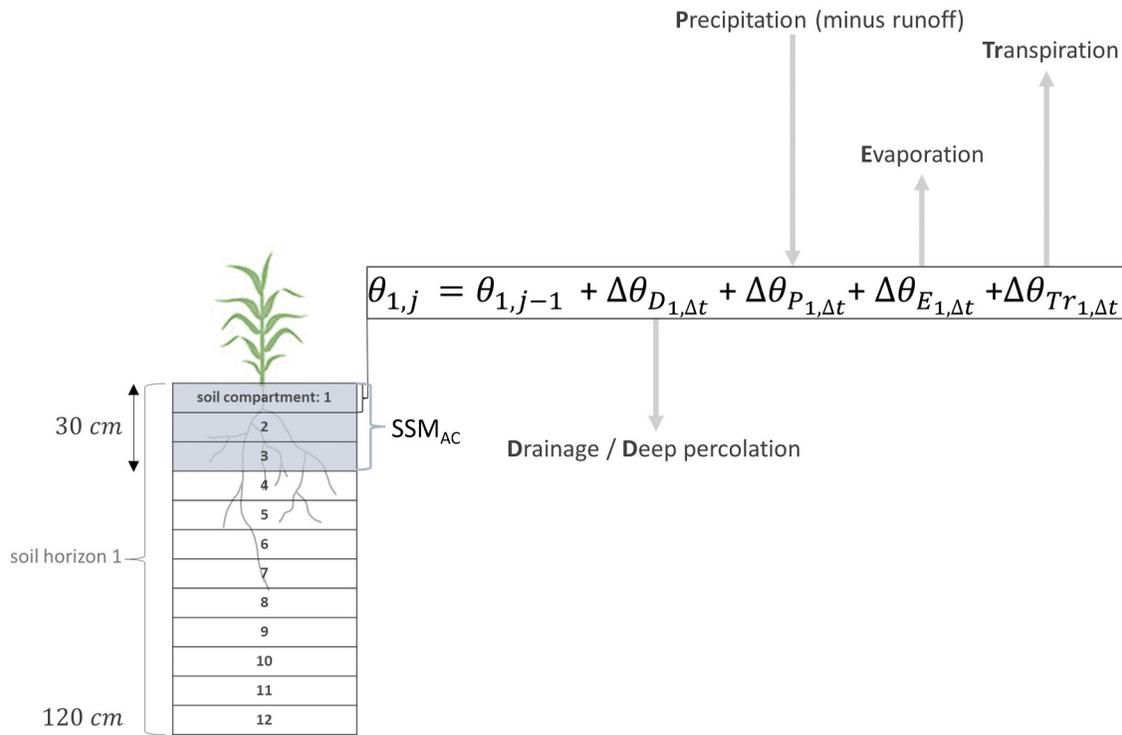


Figure 2. Schematic overview of the AquaCrop soil moisture distribution with incoming and outgoing fluxes relevant to this study (excluding capillary rise and irrigation). The function describes the soil moisture distribution in the first soil compartment, with i as daily time step and Δt as time difference between the simulated day and previous day. The function is adapted from the AquaCrop calculation manual in Raes et al. (2022).

crop type. The option of representing fractional vegetation within a pixel, as is often done in LSMs, was ignored in this study due to the lack of annual crop type maps at the desired scale and resolution. Any systematic biases due to this assumption will be ignored within the DA system, because it is only designed to update the temporal variation in soil moisture and biomass (see Section 2.1.3).

2.1.2. The AquaCrop Model

The AquaCrop model was originally developed by the FAO for field-scale applications to simulate crop development over a homogeneous field. AquaCrop is a water use efficiency model, linking biomass production to crop transpiration (Steduto et al., 2009). The water availability in the root zone is therefore a crucial component of the model as it is used to estimate drought stress. The model uses a conceptual water balance equation as described in Raes et al. (2006, 2009), considering incoming fluxes of precipitation, irrigation and capillary rise, and outgoing fluxes of runoff, evaporation, transpiration and deep percolation at the daily timescale. Losses by leaf interception are not included, as it is expected that the error on the soil water balance and biomass production is larger than the errors caused by a poor estimate of the intercepted water.

To compute the vertical water distribution within the root zone, the soil profile is divided into 12 equally sized compartments. For each compartment, the model computes the volumetric soil moisture content θ (m^3m^{-3}) (Raes et al., 2006) as shown in Figure 2. The drainage, which occurs when the water content exceeds field capacity (θ_{FC}), is based on a function depending on the saturated hydraulic conductivity. For this study, no groundwater table is included, therefore capillary rise is not simulated and the lower boundary of the water balance is free drainage. Irrigation is also not considered for this study. In short, the soil water content θ_{li} in layer l at day i is equal to that of the previous day plus changes ($\Delta\theta$), resulting from the effective incoming precipitation, evaporation and transpiration, and in case θ_{li} exceeds θ_{FC} , also drainage. The water from the precipitation propagates through the subsequent compartments as infiltration, which only happens in wet soil, and the drainage is assumed constant throughout the soil profile for a uniform soil. More detailed information about the water balance equations can be found in Raes et al. (2022) and de Roos et al. (2021).

Crop development is simulated by the fraction of land covered by vegetation, that is, the canopy cover (CC), instead of the more common LAI (Raes et al., 2009). The CC is used to compute transpiration, by adjusting the evapotranspiration with a crop coefficient, which is proportional to CC adjusted for micro-advective effects (CC*). The above-ground cumulative dry biomass (CB) is then derived from the transpiration (Tr) using a water productivity (WP*; g m⁻²) factor, which is a parameter describing the amount of dry biomass produced over a unit area. WP is only significantly different when normalized for climate (by dividing it over the reference ET), between C₃ (15–20 g m⁻²) and C₄ (30–35 g m⁻²) crops. Because the vegetation in this study is simulated as a C₃ crop, the WP parameter is set to 17 g m⁻² (0.17 t ha⁻¹). Apart from normalizing for climate, WP* is normalized for CO₂, considering the positive effect of annual increase in atmospheric CO₂ on crop productivity using a correction coefficient. The calculations for this normalization can be found in Raes et al. (2022). More precisely, CB_{*i*} at day *i* equals

$$CB_i = CB_{i-1} + WP^* \cdot \frac{Tr_i}{ET_{0i}} \quad (1)$$

Daily CB is thus diagnosed from the daily Tr, which is itself determined by CC, the reference evapotranspiration ET₀, and affected by stress factors related to water and temperature. Examples of water stresses in the generic crop are stomatal closure which occurs when 50% of the total water between θ_{FC} and θ_{WP} has been depleted, or canopy expansion that becomes limited when the water availability falls below 70%. The magnitude of the stress becomes larger as time progresses and the effects depend on the growth stage of the canopy. By normalizing the transpiration for climate (i.e., dividing transpiration by ET₀) the WP factor, a crop parameter, only differs significantly between C₃ and C₄ crops.

Crop yield is finally estimated by multiplying the CB with a crop specific harvest index, which is not considered in this study. More details about the model equations can be found in the documentation of AquaCrop (Raes et al., 2022) or in Raes et al. (2006, 2009) and Steduto et al. (2009).

2.1.3. AquaCrop Integration Into LIS

In this study, the compiled modules of AquaCrop were integrated into NASA's LIS version 7.4 (Kumar et al., 2008) and ran at ~1-km (30 arcs) resolution, using an efficient parallelization scheme. LIS is a software framework to facilitate efficient integration of advanced modeling techniques with satellite and ground-based observational data. It encapsulates multiple physical models (LSMs) and allows for utilization of multiple assimilation, optimization and uncertainty estimation algorithms (Kumar et al., 2006, 2008; Peters-Lidard et al., 2007). LIS is written in Fortran90, as it is developed to run efficiently on high performing computing systems. For efficiency reasons, the entire official AquaCrop code was also translated from its original Pascal language into Fortran90 and officially released by FAO as part of version 7 (and higher) of AquaCrop (<https://www.fao.org/aquacrop/en/>). The Fortran90 language of the official AquaCrop sets it apart from unofficial copies in computationally less efficient high-level programming languages like for example, Python. Within LIS, the model can be run in deterministic or ensemble mode, for a range of forcing data sets and spatially distributed satellite DA. In ensemble mode, the ensemble spread (std) is given as output along with the ensemble mean as a measure of uncertainty.

For the experimental setup, the simulation and evaluation period was set from 2015 through 2021, preceded by a 2-year spin-up and the AquaCrop model was run with 24 ensemble members. Meteorological variables (precipitation, 2-m temperature, ET₀) were taken or derived from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017), which was bi-linearly interpolated to the spatial grid of the model simulations (from 0.5°x 0.625° to 30 arcs). Terrain corrections were based on the digital elevation model from the Shuttle Radar Topography Mission (Farr et al., 2007; Farr & Kobrick, 2000). The MERRA-2 precipitation (P) and temperature (T) were converted to the input units required by AquaCrop, that is, mm day⁻¹ and °C. ET₀ (mm day⁻¹) is not a variable directly available from MERRA-2, but was derived with the Penman-Monteith equation, using radiation, mean 2-m temperature, 2-m dew temperature, and 2-m wind speed (Allen et al., 1998).

One homogeneous soil horizon was considered with a depth of 1.2 m, and soil textural properties (θ_{FC} ; θ_{WP} ; θ_{sat}) were used from a simple soil texture classification of 16 classes embedded within the LIS framework, derived

from the global hybrid STATSGO/FAO soil map (Miller & White, 1998; Reynolds et al., 2000). To limit the runoff potential and subsequent loss of water within the $1 \times 1 \text{ km}^2$ pixel, the curve number was set to a very low value of 6, assuming that water lost in one crop field ends up in another field within the pixel considering the mild topography of the study areas.

For the matching between soil moisture simulations and observations, a SSM (SSM_{ac} ; m^3m^{-3}) at each day i was derived for AquaCrop by aggregating the first three soil moisture compartments (Figure 2) using:

$$SSM_{ac,i} = 0.75 \cdot \theta_{1i} + 0.15 \cdot \theta_{2i} + 0.1 \cdot \theta_{3i} \quad (2)$$

The weighted averaging was done to reduce the strong threshold effects of θ_{FC} and θ_{WP} for individual AquaCrop soil compartments as described in de Roos et al. (2021), and to help the DA setup (see below). The strong threshold effects result from the conceptual water balance algorithm of AquaCrop and are not found for SSM estimates of LSMs that use a physically-based water balance, that is, Richards equation (not shown). It should be noted that in AquaCrop, the soil moisture over the entire root zone is used to determine the water availability to the crop, hence the aggregated soil moisture from all compartments. The individual compartments were designed to estimate drainage within the root zone and are more of a simplification compared to a physically-based approach.

2.2. Observation and Reference Products

2.2.1. Sentinel-1

The S1 mission from the European Space Agency and Copernicus measures radar backscatter in C-band (5.4 GHz) (Torres et al., 2012). The mission consists of two satellites; Sentinel-1A (S1-A) and Sentinel-1B (S1-B), with a temporal coverage over Europe of about 2–3 days. The standard acquisition mode over land (outside polar regions) is the Interferometric Wide Swath mode, which provides co-polarized (vertical-vertical; VV) and cross-polarized (vertical-horizontal; VH) backscatter at 5 m by 20 m ground range resolution, from October 2014 onward. The data were processed to the daily timescale, providing observations every ~ 3 days and they were assimilated year-round excluding frozen soil conditions (with masking based on temperatures below 1°C). The processing of the data was done as in Lievens et al. (2019), Modanesi et al. (2021) and de Roos et al. (2023), using ESA's Sentinel Application Platform (SNAP) software. The processing steps consisted of precise orbit file application, border and thermal noise removal, radiometric calibration to backscatter as beta-naught, terrain flattening to backscatter as gamma-naught (γ^0), and range Doppler terrain correction. Urban and water areas with coverage fraction larger than 0.1 (10%) were masked out based on PROBA-V land cover data, after which the data was regridded to match the spatial resolution of the simulated grid at 30 arcs. An orbit-bias correction was applied according to Lievens et al. (2019), in which the mean γ^0 of each orbit was rescaled using the static (long-term averaged) bias of the 175 relative orbits.

2.2.2. In Situ ISMN Stations

Three station networks within the ISMN were considered for SSM evaluation: TERENO, SMOSMANIA and REMEDHUS. Even though the AquaCrop soil moisture compartments from 0 to 30 cm were aggregated to SSM_{ac} , only ISMN SSM (SSM_{ismn}) observation points with soil depths up to 10 cm were selected, as the aggregation was done to partially correct for the relatively large variations in simulated soil moisture in the top compartment, and the simulated top compartment of 10 cm still contributes most (75%) to the total considered soil depth. Only measurements that were flagged as *good quality* were included. Additionally, each soil moisture station was checked manually and suspicious measurements (e.g., very noisy or a soil moisture content reaching $0.0 \text{ m}^3\text{m}^{-3}$) were removed. The data were aggregated to daily data by taking the latest measurement of the day (usually at 23:00 hr). This was done as AquaCrop runs in daily time steps, and the calculated soil moisture represents the value at the end of the day. The observational soil moisture data can contain large time gaps (months to years), hence a minimum threshold was set of 1,100 daily measurements, to ensure data for at least 3 years, covering sufficient seasonal variation. The considered evaluation period between the years 2015 and 2021 is thus not consistent between the locations.

2.2.3. SMAP Surface Soil Moisture

For a spatially complete evaluation, the 1-km downscaled Soil Moisture Active Passive (SMAP) Level 2 product of B. Fang et al. (2022) was used. The 1 km variation in this product is based on an inverse relationship between land surface temperature (derived from Aqua MODIS) and soil moisture.

Data for descending (06:00 hr) and ascending (18:00 hr local time) overpasses are provided separately, but for this study only ascending data were considered, as these were derived from observation at the latest time of the day, and assumed more compatible with AquaCrop daily soil moisture simulations. The data are available from 1 April 2015 onwards and were regridded to match the spatial resolution of the simulated soil moisture, using the nearest neighbor method. As opposed to in situ measurements that can contain large time gaps, SSM retrievals from the SMAP satellite (SSM_{smap}) have a more complete temporal coverage over the evaluation period, with data about every 3–6 days. Therefore, a lower minimum of 800 data points per pixel was set for evaluation with SSM_{smap} data.

2.2.4. SMAP Vegetation Optical Depth

The vegetation optical depth (VOD) product from Konings et al. (2017) was used to evaluate the CB simulations. These data are derived from the SMAP Level 1C brightness temperature observations with the Multi-Temporal Dual Channel Algorithm (MT-DCA) by Feldman et al. (2021). Because the AquaCrop CB output concerns dry and not wet biomass, the VOD was compared against an adjusted cumulative biomass index B_w that includes a correction to wet biomass and a vegetation decline to simulate gradual harvest (harvest factor; HF) as described in de Roos et al. (2023), with the following equation:

$$B_{wi} = \sqrt{CB_i \cdot HF_i} \quad (3)$$

HF_i is the ratio, whereby the minimum temperatures are smoothed using a 40-day averaging window, dividing the minimum smoothed temperature on day i , over a reference minimum temperature window centered around July 1. HF therefore assumes that harvest is linked to declining temperatures. In some regions, harvest is more closely related to water deficiency instead of temperature decline, which is also accounted for by keeping track of the water availability during the growing season, as described in detail in de Roos et al. (2023). Note that in this study the focus is mainly on the first part of the growing season, in which the biomass productivity increases until the maximum is reached, as is described in Section 2.5. The 30-arc AquaCrop simulations were aggregated to match the VOD 9-km EASE grid, by averaging all simulations within each separate 9-km EASE grid cell. Only data were considered when the pixel contained at least 60% agricultural land and only the months May–September were considered to exclude vegetation impacts from crops outside of the summer growing season.

2.2.5. CGLS Dry Matter Productivity

The dry matter productivity (DMP_{cglis} ; $kg\ ha^{-1}\ d^{-1}$) from the Copernicus Global Land Service (Wolfs et al., 2022) was used to evaluate daily DMP_{ac} , which is defined as $\Delta CB = CB_i - CB_{i-1}$ (see Equation 1). The DMP_{cglis} product is based on the fraction of absorbed photosynthetically active radiation (fAPAR) by vegetation, which was derived from a sequence of optical satellites; PROBA-V from January 2014 through June 2020, followed by Sentinel-3/OLCI sensor data, from July 2020 onward (Fuster et al., 2020). The data were used for the evaluation years 2015–2021. Apart from fAPAR, the product makes use of radiation, temperature, and CO_2 from the ECMWF Reanalysis -interim (ERA-interim) atmospheric reanalysis data set. The meteorological variables were derived from ECMWF with a $0.25^\circ \times 0.25^\circ$ spatial resolution (Wolfs et al., 2022). The DMP_{cglis} product provides data three times a month, in which each value is representative of the past dekad (10 days). It should be noted that the product does not make direct use of soil water information and therefore does not capture the effect of short-term water stress on vegetation productivity, but only indirectly includes drought stress through the fAPAR retrieval (Wolfs et al., 2022). The normalized temperature effect is a fraction (0–1) used to multiply with fAPAR, in which the boundaries are based on generalized C_3 vegetation, which has a lower temperature optimum compared to C_4 crops (Wolfs et al., 2022). The product is therefore more suitable to be compared against C_3 vegetation.

The product is available approximately every 10 days at a 300 m spatial resolution and was resampled using bi-linear interpolation to match the spatial grid and resolution of the model simulations at 30 arcs. The DMP is a productivity rate product, considering the absolute productivity per day. To derive productivity rates from the AquaCrop CB, the daily differences were calculated and the mean of DMP_{ac} production was taken for each dekad to match the temporal resolution of the DMP_{egls} product.

2.3. Data Assimilation

The model-only AquaCrop simulations of soil moisture and CB are expected to contain errors, due to coarse meteorological input and generalized crop and soil information. The DA system was set up to sequentially assimilate backscatter γ^0 observations from S1 and to update the diagnostic CB and prognostic model variables θ_1 , θ_2 , and θ_3 , which together compose SSM_{ac} (Equation 2). The DA algorithm asks that both the observations and forecasts are unbiased. Therefore, the link between the model variables and γ^0 forecasts for both polarizations is made with a calibrated forward operator.

2.3.1. Forward Operator

The mapping of AquaCrop model variables SSM_{ac} and CB to the γ^0 space was done with the Water Cloud Model (WCM), described by Attema and Ulaby (1978). The WCM considers the total backscatter with a soil and vegetation backscatter component, in which the vegetation is represented as a water cloud assuming identical water droplets that are randomly distributed. It therefore requires both soil moisture and fresh vegetation as model input variables, which for this study are the SSM_{ac} and B_w . The calibration of the WCM parameters can be found in Appendix A. For a more detailed motivation of the WCM setup and evaluation for the different areas, the reader is referred to de Roos et al. (2023).

2.3.2. Assimilation Algorithm

The 1-dimensional stochastic ensemble Kalman Filter (EnKF; Evensen, 2003) was used to sequentially update either only the prognostic AquaCrop SSM compartment variables θ_1 , θ_2 , and θ_3 composing SSM_{ac} (Equation 2), or also AquaCrop cumulative biomass, CB. Note that the CB is diagnosed output (see Section 2.1), and not a prognostic state variable in AquaCrop, but it is included in the update vector (called control vector below). In the EnKF, a small number of ensemble trajectories are generated based on perturbed instances of the meteorological input and selected control variables. These perturbations should capture the forecast errors in terms of input, structure and parameters. The EnKF can be described as follows:

$$\hat{x}_{i,j}^+ = \hat{x}_{i,j}^- + K_i [y_{i,j} - \hat{y}_{i,j}^-] \quad \text{with} \quad \hat{y}_{i,j}^- = h_i(\hat{x}_{i,j}^-) \quad (4)$$

where i indicates the time step, j the ensemble member, $\hat{x}_{i,j}^+$ the ensemble control vector in the model space after the DA update, $\hat{x}_{i,j}^-$ the ensemble control vector in the model space before assimilation, $\hat{y}_{i,j}^-$ the ensemble observation predictions in the backscatter space obtained with the forward operator, $h_i(\cdot)$, and $y_{i,j}$ the observations (γ_{VV}^0 , or γ_{VV}^0 and γ_{VH}^0 , see Section 2.4), which are also perturbed based on the given standard deviation (SD) of the observation error. The forward operator consists of three steps; (a) the top three soil compartments of AquaCrop are combined to obtain SSM_{ac} , (b) AquaCrop dry CB is converted to B_w (as described above, and in de Roos et al., 2023), to better represent the vegetation as observed by the satellite, that is, wet biomass that includes vegetation decline, and (c) both B_w and SSM_{ac} are used in the WCM to simulate γ^0 .

The EnKF considers the uncertainty of both the forecasts (represented by the ensemble) and the observations, which include instrument errors and representativeness errors, to assign a weight (Kalman gain; K) to the updates. More specifically, K_i is calculated as follows:

$$K_i = COV_{(\hat{x}_i^-, \hat{y}_i^-)} [COV_{(\hat{y}_i^-, \hat{y}_i^-)} + R_i]^{-1} \quad (5)$$

Where $COV_{(\hat{x}_i^-, \hat{y}_i^-)}$ is the error cross-covariance of the forecast in model space (\hat{x}_i^-) and backscatter space (\hat{y}_i^-), $COV_{(\hat{y}_i^-, \hat{y}_i^-)}$ is the forecast error covariance in backscatter space, and R_i is the observation error covariance.

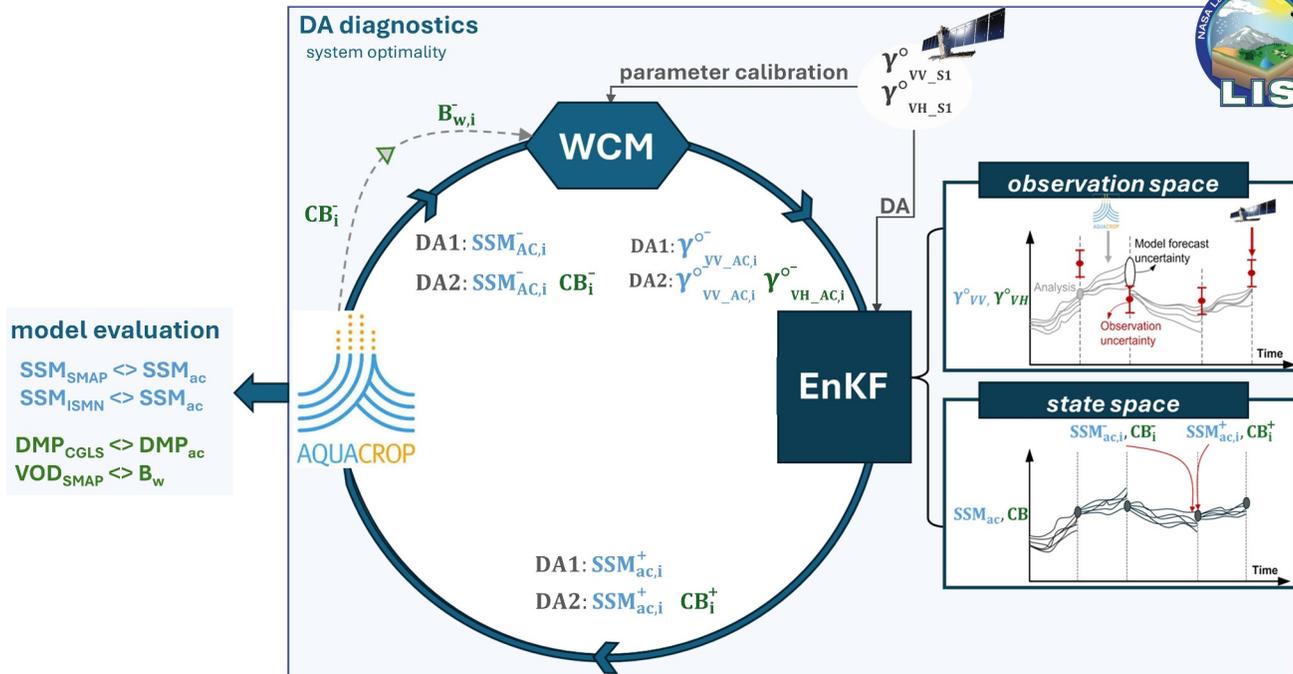


Figure 3. Schematic overview of the study set-up, showing the iterative process of the data assimilation in the Land Information System framework (blue lined box). Green text refers to vegetation variable updating and blue to soil moisture updating. The conversion of AquaCrop CB to B_w , before entering the forward operator of the Water Cloud Model (WCM) is indicated by the dashed arrow. The blue SSM_{ac} is the weighted mean of the three AquaCrop soil water content compartments and is updated for both experiments (DA1, DA2). Negative signs (–) indicate the variable before DA and positive signs (+) refer to updated variables. The Sentinel-1 (S1) backscatter is used for the calibration of the WCM parameters and during the DA algorithm, using the ensemble kalman filter (EnKF). As illustrated with the two time series, the EnKF computes differences between S1 and simulated backscatter in observation space, and maps these differences to updates in SSM and CB in state space.

$COV_{(\hat{x}_i^-, \hat{y}_i^-)}$ governs the mapping from backscatter space to the control variables (three soil water compartments and CB, depending on the experiment). The sizes of the control and observation vectors, and the associated error matrices depend on the DA experiment, and are discussed below.

The model variables were sequentially updated with S1 observations at the points in space and time at which these were available. Not every location was therefore updated with the same frequency of S1 observations. No bias correction was applied between the observed and simulated backscatter, because the WCM calibration presumably removed the long-term bias.

2.4. Experiments

Two DA experiments were run over the three study areas. The experiments use the same ensemble perturbations to meteorological forcings, CB, and θ_1 , θ_2 , and θ_3 . In the first experiment, only the prognostic θ_1 , θ_2 , and θ_3 are updated, whereas in the second experiment also CB is updated with backscatter. The DA system setup of the two experiments is schematized in the flowchart of Figure 3. CB is a diagnostic variable with temporal memory (cumulative) and updates in CB will thus not directly feed back into the AquaCrop model, but CB updates alter the input to the WCM and thereby the mapping of γ_{VV}^0 to soil moisture increments.

Experiment 1 (DA1) - Only γ_{VV}^0 observations were assimilated to update θ_1 , θ_2 , and θ_3 composing SSM_{ac} (Equation 2). The three θ compartments as well as the CB were perturbed, as shown in Table 1. The forecast error perturbation for the CB was assumed to be uncorrelated to that of θ in the 3 compartments. Since the CB is not updated, changes in CB come solely from updated SSM_{ac} . The perturbations to the three compartments of the SSM_{ac} are cross-correlated, with a factor of 0.6 for directly adjacent compartments and a factor of 0.4 between the first and third compartment. Using a weighted average ($\theta_{1,2,3}$) can allow for a larger ensemble spread in the

Table 1
Meteorological (P , T , ET_o), Model (θ , CB) and Observation (γ^0)
Perturbation Parameters, Indicated by the Standard Deviation (SD) of the
Perturbation

Variable	Perturbation type	SD
precipitation (mm day ⁻¹)	multiplicative	0.5
temperature (°C)	additive	1.0
ET _o (mm day ⁻¹)	additive	0.7
$\theta_1, \theta_2, \theta_3$ (m ³ m ⁻³)	additive	0.006
CB (ton ha ⁻¹)	additive	0.01
γ_{VV}^0 (dB)	additive	0.7
γ_{VH}^0 (dB)	additive	0.9

SSM_{ac} as the threshold at θ_{FC} in the top compartment in AquaCrop is easily reached, after which the excess water is transported to the lower compartments. The combination of the soil compartments consequently introduces deeper soil moisture updates, weighted by their respective uncertainty and their fractional contribution to SSM_{ac}. In short, the dimensions of $\hat{x}_i^{-,+}$, $COV(\hat{x}_i^-, \hat{x}_i^+)$, y_i and R_i are (3×1) , (3×1) , (1×1) and (1×1) , respectively.

Experiment 2 (DA2) - Similar to Experiment 1 for updating SSM_{ac}, but with the additional assimilation of γ_{VH}^0 to update CB. The choice was made to update the diagnostic CB and not the prognostic CC for two reasons, (a) S1 backscatter is an indication of total canopy volume, which is more closely related to biomass than fractional vegetation cover and (b) the CC saturates much earlier compared to other phenological indices such as LAI, where CC reaches its maximum in an early crop stage and does not show much variation until senescence has set in, which becomes difficult for updating in the DA

system. The ensemble perturbations of θ_1, θ_2 , and θ_3 , and CB were not changed compared to Experiment 1. The observation error correlations between γ_{VV}^0 and γ_{VH}^0 were ignored, and the update described by Equation 4 can therefore effectively be split into two sequential update equations. One is the same as in Experiment 1 to update θ_1, θ_2 , and θ_3 composing SSM_{ac}, and the second one only has dry CB in $\hat{x}_i^{-,+}$, and γ_{VH}^0 in y_i , so that the error covariance matrices for the second update reduce to scalars. Apart from providing updated dry CB output, this experiment also provides updated B_w estimates to the WCM as the forward model, which should ideally improve the backscatter simulations of both γ_{VV}^0 and γ_{VH}^0 .

Note that γ_{VV}^0 and γ_{VH}^0 are observed by S1 at the same time and frequency, therefore the number of updates in time is the same per location for both experiments. The differences between observed and simulated γ_{VV}^0 or γ_{VH}^0 will be referred to as innovations ($[y_{i,j} - \hat{y}_{i,j}^-]$), whereas the updates to $\theta_{1,2,3}$ or CB will be referred to as increments ($K_i [y_{i,j} - \hat{y}_{i,j}^-]$).

The DA experiments (DA1, DA2) are supplemented with one corresponding open loop (OL) run, in which the model is executed in ensemble mode without assimilation. For all simulations (OL, DA), 24 ensembles were generated with the perturbation standard deviations (SD) listed in Table 1. The perturbations for the model meteorology are generated using a standard normal (additive) error distribution for T and ET_o, and a log-normal (multiplicative) error distribution for P with a mean of 1 (–). This means that for P, zero values are not perturbed and no negative values are allowed. For T and P, common error SDs used in LSMs are 1°C and 0.5 (–) respectively (Kumar et al., 2006; Peters-Lidard et al., 2011; Reichle et al., 2010). ET_o is not an input variable for LSMs, because these LSMs compute ET themselves based on various meteorological variables. Therefore, no reference SD values for the ET_o are directly available in literature. The perturbation SD for ET_o, $\theta_{1,2,3}$ and CB were manually determined based on innovation statistics. Since perturbing the model variables can introduce additional bias caused by strict boundaries of the variables (e.g., θ_{FC} , θ_{WP}), a perturbation bias correction (Ryu et al., 2009) was applied to the model variables ($\theta_{1,2,3}$, CB).

The observation error for γ_{VH}^0 (0.9 dB) was set higher than for γ_{VV}^0 (0.7 dB), because in de Roos et al. (2023) higher root mean square differences (RMSD) values between observed and simulated γ_{VH}^0 were found. The observation error contains both the instrument error and the representativeness error caused by the imperfect WCM. The results of de Roos et al. (2023) showed that the RMSD between the observed and simulated backscatter is quite variable over the areas, and although it is expected that optimizing the observation perturbation per location would give better results, the setup is intended for large-scale applications over Europe. The SD of the perturbations is therefore kept general for all polarizations and experiments.

2.5. Evaluation of DA Impact on Soil Moisture and Biomass

Table 2 summarizes the AquaCrop variables that are evaluated for each of the DA experiments; one related to the soil moisture (SSM_{ac}) and three related to the biomass (CB, DMP, B_w). The evaluation products for the model variables are also summarized in the flowchart of Figure 3.

Table 2
AquaCrop Variable Definitions for Evaluation

AquaCrop variable	Definition	Mathematical definition
$SSM_{ac,i}$ ($m^3 m^{-3}$)	Surface soil moisture	$0.75 \cdot \theta_{1,i} + 0.15 \cdot \theta_{2,i} + 0.1 \cdot \theta_{3,i}$
CB_i ($t ha^{-1}$)	Cumulative dry biomass	$CB_{i-1} + WP^* \cdot \frac{Tr_i}{ET_{0i}}$
$DMP_{ac,i}$ ($t ha^{-1} day^{-1}$)	Daily dry matter productivity	$CB_i - CB_{i-1}$
$B_{w,i}$ (—)	Cumulative wet biomass index	$\sqrt{CB_i \cdot HF_i}$

In the first part of the evaluation (Section 3), the impact of the DA is assessed for both experiments by comparing the DA time series to the model-only (OL) simulations. The SSM_{ac} estimates were compared to in situ (SSM_{ismn}) and SMAP (SSM_{smap}) data in terms of Pearson's correlation (R), anomaly R (i.e., after having subtracted the 7-year based climatology from the original time series) and unbiased root mean square difference (ubRMSD). The anomaly R was used to evaluate both the short-term and inter-annual variation of SSM_{ac} . The SSM_{ac} is evaluated year-round between 2015 and 2021.

The AquaCrop biomass was evaluated in terms of DMP_{ac} against the DMP_{cgl} product, and in terms of B_w against the VOD_{smap} product. The evaluation of DMP was done only from January until the peak in vegetation productivity. The vegetation peak was selected as the end of the evaluation period since Khabbazan et al. (2022) showed that the backscatter signal and vegetation B_w correlate quite well until full biomass is reached, indicating the end of the vegetative stage. The peak was defined as the maximum value of the CGLS product, to make the evaluation independent of the experiments. The comparison with B_w and the VOD time series was done for the growing period between the months May-September, to exclude the effect of other growing seasons. The peak was not selected in this case as the VOD data can be very noisy.

For the evaluation of the DMP, the long-term anomaly correlation was used as evaluation metric, which is calculated by removing the short-term anomaly (window size of 35 days) from the total anomaly. The focus is put here on long-term anomaly correlations because the EnKF algorithm can introduce a spiky pattern in the short-term estimates, which is realistic for SSM_{ac} but not for CB. Furthermore, as mentioned in Section 2.2.5, the DMP_{cgl} product is known to not represent short-term water stress well. The B_w time series are visually compared to the VOD for some points in terms of slope (i.e., increase in vegetation).

2.6. Evaluation of DA Setup With DA Diagnostics

In the second part of the evaluation (Section 4), the DA setup is evaluated in terms of DA diagnostics (Reichle et al., 2010). More specifically, the time series SD of the normalized innovations N_{innov} is used to assess whether the uncertainties in the DA system are well captured. The $N_{innov,i}$ at time steps i is:

$$N_{innov,i} = \frac{y_i - \hat{y}_i^-}{\sqrt{VAR(y_i) + VAR(\hat{y}_i^-)}} \quad (6)$$

Where $y_i - \hat{y}_i^-$ is the ensemble mean observation minus forecast residual (backscatter innovation), $VAR(y_i)$ and $VAR(\hat{y}_i^-)$ are the error variances of the observed backscatter and modeled backscatter, respectively. In an optimal DA system, the SD of the normalized innovations is 1, indicating that the time series differences are accounted for in the assumed ensemble uncertainty, whereas values above 1 indicate that not all the uncertainty is captured, and values below 1 show an overestimation of the uncertainty (De Lannoy & Reichle, 2016). Related to this, an analysis of the forecast uncertainties will explain where and when most DA impact can be expected.

Furthermore, the soil moisture and vegetation updates were related to modeled soil and vegetation properties. The CB increments were studied in relationship to the WCM vegetation parameters and the amount of CB. More specifically, the correlation between the γ_{VH}^0 innovations and the CB increments ($R_{[INCR, INNOV]}$) for different growth stages (levels of CB) were computed for different parts of the growing season. The growing season was divided into an early season (January-March; JFM) during which the CB is still relatively small and the modeled

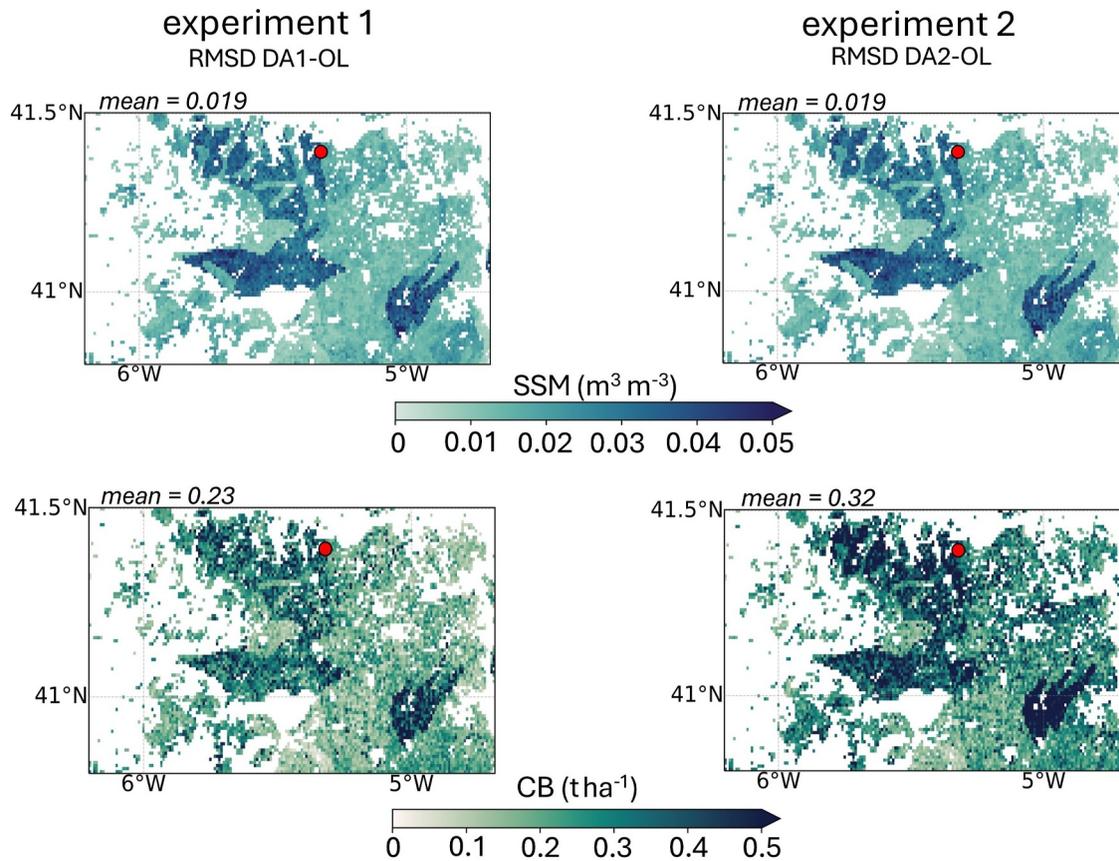


Figure 4. Root mean square differences between DA and OL runs, visualizing the impact of the DA for CB (January–July) and SSM_{ac} (all year) in (left) Experiment 1 and (right) Experiment 2, for area 3. Red dot indicates the location of Figure 5.

γ_{VH}^0 is expected to be dominated by SSM_{ac} , and the late season (March–July; MJJ), during which the CB (almost) reaches its maximum, hence the contribution of vegetation to the total γ_{VH}^0 signal is increased.

3. Results

3.1. Impact Analysis

Figures 4 and 5 give a first impression of the DA effect on AquaCrop SSM_{ac} and CB simulations for the two experiments, spatially (Figure 4) and temporally (Figure 5) for area 3. Time series for locations in the other two areas can be found in Appendix B. The temporal RMSD maps in Figure 4 between the DA (DA1, DA2) and OL runs visualize the main differences between the two experiments. Whereas the impact on the SSM_{ac} is almost equal, the changes in CB are stronger in Experiment 2. For both experiments, the areas with a clay soil texture stand out in the maps, with larger impacts on CB and SSM_{ac} compared to other soil textures.

The SSM_{ac} time series in Figure 5 for Experiment 1 and 2 show no significant differences, as both were updated with γ_{VV}^0 using the same setup (WCM parameters, perturbations). The DA impact on SSM_{ac} is visible at the start or end of the growing season, but becomes negligible when the soil moisture contents are close to or at θ_{WP} in summer, or close to θ_{FC} in winter (Figure 5), due to the boundary restrictions. For Experiment 1, updates of the SSM_{ac} show an effect on the CB for most years, most significantly for the years 2015, 2017, and 2020. The effect of the SSM_{ac} updates on the vegetation is strongly dependent on their timing and magnitude. For the year 2017 for example, in Figure 5, DA1 causes a slight decrease in SSM_{ac} early in the season, resulting in a lower CB. For 2021, DA1 also results in lower SSM_{ac} , but without affecting the CB. This indicates that there is still sufficient water available in the root zone for the early growth stage of the vegetation and no water stress is initiated.

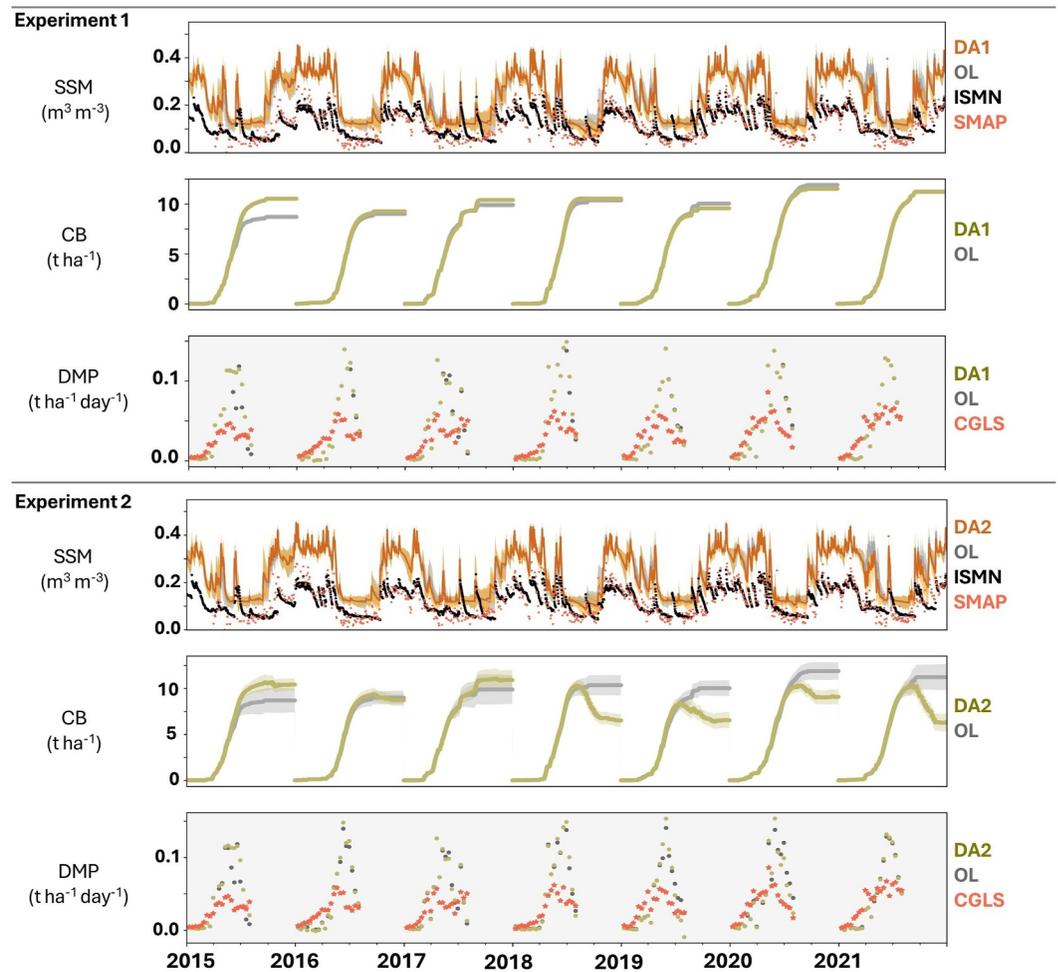


Figure 5. Time series of SSM_{ac} , CB and DMP (OL and DA) for Experiment 1 (top) and Experiment 2 (bottom), at point $41.39^{\circ}N$ and $5.32^{\circ}W$, situated in area 3. Daily simulations of SSM_{ac} and CB are shown as lines, and, if it concerns a directly updated variable, the spread of the model ensembles is shown as the 95% confidence interval. The DMP plots show less data points because they are matched to the 10-daily time steps of the CGLS evaluation product.

The DMP plots for both experiments in area 3 (Figure 5) show an AquaCrop overestimation in both DA and OL relative to the DMP_{cglis} . For area 1 (Appendix 1, Figure B1), there is only a slight overestimation for some years, whereas for area 2 (Appendix B, Figure B2), the ac is generally lower than observed. Comparing the time series of the three study areas, the changes in CB due to DA1 have the most significant effect on study area 2. Out of the three study areas, area 1 shows the lowest impact on the CB resulting from SSM_{ac} updates (Appendix B, Figure B1) and the SSM_{ac} spread is also smaller compared to areas 2 and 3. The main difference between area 1 and the other two areas is that it is situated in a temperate climate region and is therefore less dependent on water availability compared to the other areas, which are situated in a Mediterranean climate region. Additionally, the soil texture plays a significant role in the size of the ensemble spread, which will further be elaborated in Section 4.2. A large ensemble spread results in strong SSM_{ac} updates and subsequent changes in CB and DMP_{ac} (Appendix B, Figure B2).

In Experiment 2, the CB was explicitly updated using γ_{VH}^0 observations, and the stronger effect on the CB in the DA2 run can be seen in the maps (Figure 4) as well as the time series (Figure 5). The ensemble spread in CB becomes significantly smaller than the OL after DA updates, for almost all years (Figure 5). Furthermore, the DA results in different CB curves for the years 2018–2021, where the corrected CB is no longer monotonically increasing within a season, but also shows declines from the corrections to the B_w that is used to describe the vegetation in the WCM as observed by the satellite.

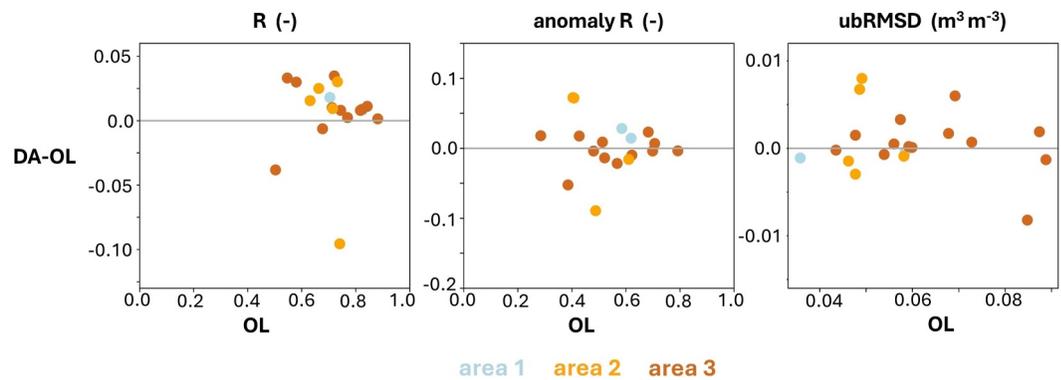


Figure 6. Evaluation of SSM_{ac} with SSM_{ismn} for area 1 (2 sites), area 2 (5 sites), and area 3 (13 sites) over the years 2015–2021. The OL metrics are presented on the x -axis and the change introduced by DA1 on the y -axis.

3.2. Soil Moisture Evaluation

The effect of S1 DA on SSM_{ac} for the two experiments was evaluated for several points with in situ SSM_{ismn} , and over the study area with the SMAP 1-km soil moisture product. Note that the results of Experiment 1 and 2 for SSM_{ac} are minimal.

Figure 6 shows the changes in R , anomaly R and $ubRMSD$ between the DA1 (similar for DA2) and OL simulations at the in situ points for the three areas and both experiments. Overall, the DA shows a small improvement for most in situ points in terms of R regardless of the OL performance. The performance in anomaly R is overall lower for the OL runs compared to R . DA slightly deteriorates the anomaly R for some points in area 2 and 3. The $ubRMSD$ shows no significant improvement through DA.

The results of the performance evaluation against SSM_{smap} are presented in the maps of Figure 7 and the boxplots in Figure 8. The difference maps in Figure 7 further include the results of the corresponding metrics with SSM_{ismn} points. Maps for the other two areas can be found in Appendix C. The areas with a dominant soil type of clay stand out in all the maps, with overall lower correlations but larger positive impacts from the DA. Because of the strong effect of these areas, they were excluded from the boxplot analyses in Figure 8. From the OL maps, it can be seen that the R values are highest in areas without clay soils, around 0.8. The anomaly R values are lower, but still significant (~ 0.6). After the DA, R and anomaly R increase strongly over clay soils, but the effect on the other regions is variable. The $ubRMSD$ is also very dependent on the soil texture, where the highest values are found for clay soils and the lowest values for sandy soils. The difference maps indicate a level of consistency between the SSM_{ismn} and SSM_{smap} . From the boxplots, it can be derived that the R and anomaly R values are highest for area 3, and most variable for area 2, the largest of the three areas. Area 2 shows the lowest overall performance with the SSM_{smap} out of the three areas, but R and anomaly R remain above 0.5. However, the impact of the DA is most notable for areas 1 and 2, with very slight improvements for R , anomaly R and $ubRMSD$. Due to the variable performance of the DA in area 3 (Figure 7), overall no significant changes are found in Figure 8.

3.3. Biomass Evaluation

The OL and DA DMP_{ac} was evaluated against the CGLS product in terms of long-term anomaly R , for January up to the DMP_{cgl} peak, which mostly occurred mid-June. The boxplots in Figure 9 present a comparison between the OL and DA runs. Similarly to the evaluation of SSM , the clay regions are excluded from the boxplot analysis. The simulations in these regions show high water stress in the spring and summer season, as a result of the low infiltration capacity. Consequently, for some years, no CB is simulated, which is not in line with the DMP_{cgl} evaluation product. The change in skill introduced by DA1 is generally very small, and slightly more pronounced for DA2. Area 1 shows reductions in the long-term anomaly R for DA2, and area 2 shows a slight overall improvement for DA2, whereas no impact can be seen for area 3.

Figure 10 shows time series of aggregated AquaCrop B_w for the OL and DA Experiment 2 compared to VOD_{smap} data at three locations in the three study areas. The increase in B_w nicely follows the increase in VOD for area 1 and some regions of area 2 well, but not for area 3. The temporal correlations between OL and VOD_{smap} is over

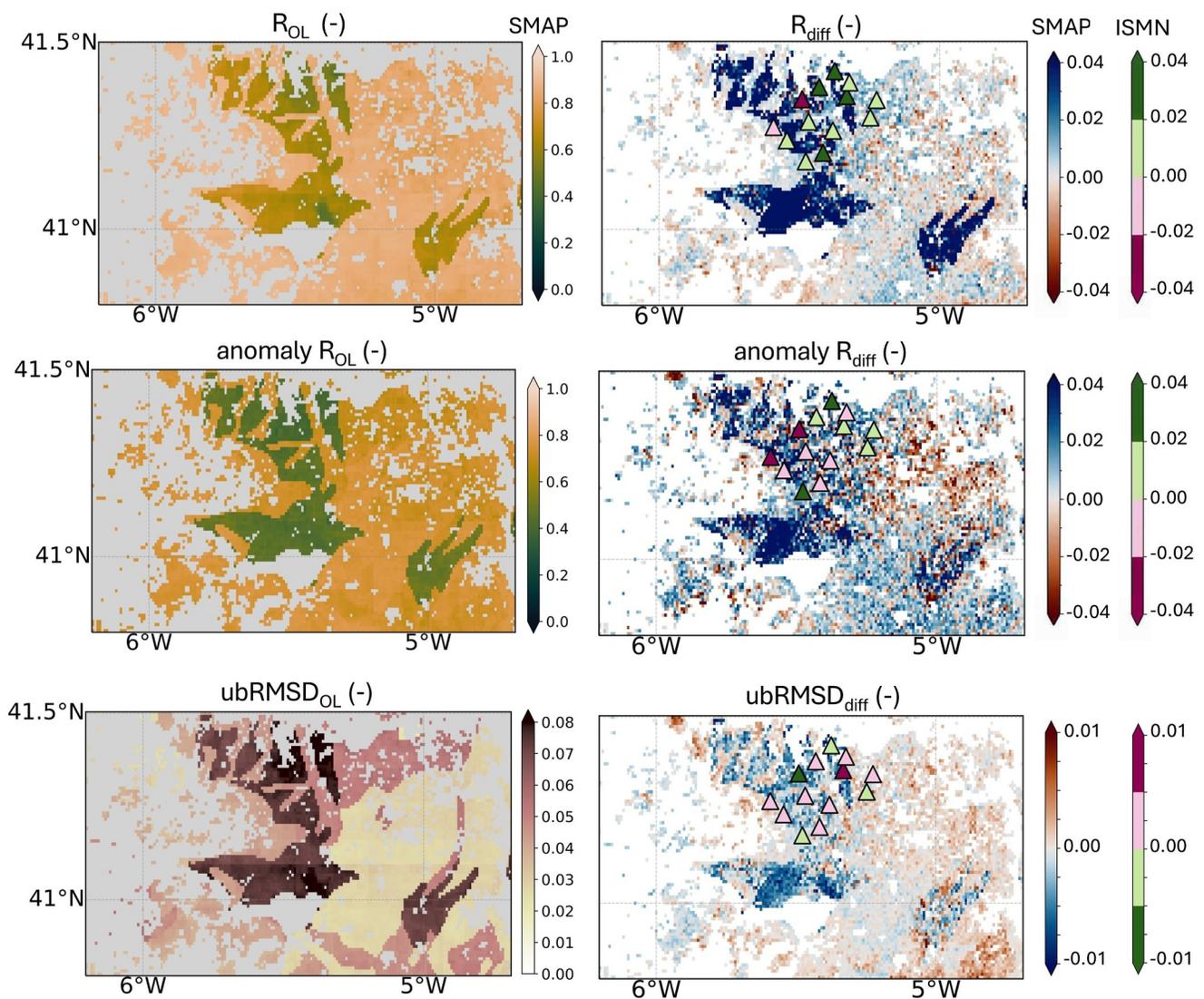


Figure 7. (Left) Spatial maps of Pearson's R, anomaly R and unbiased root mean square difference between OL SSM_{ac} and SSM_{smap} for area 3, and (right) difference maps for DA1 (DA1-OL) for the same metrics. The correlation differences at the individual International Soil Moisture Network points are shown in the difference maps with triangles.

0.7 for area 1, but strongly variable for area 2 and poor for area 3 ($R < 0$). Overall, the DA1 and DA2 runs did not improve the performance compared to the OL runs.

4. DA Diagnostics

4.1. System Optimality

In Figure 11 the standard deviation of the normalized backscatter innovations ($SD(N_{innov})$, Equation 6) of DA2 for the three study areas are presented for γ_{VV}^0 and γ_{VH}^0 . The values are close to one, but overall slightly higher, indicating that the assumed observation and forecast errors are slightly underestimated. This can be partially attributed to the forecast γ^0 uncertainty which is limited by the WCM parameters. This will be explained in more detail in the following Section 4.2. The following sections will explain how the combination of the WCM and both the observation and forecast uncertainties determine the updates to soil moisture and vegetation introduced by γ^0 .

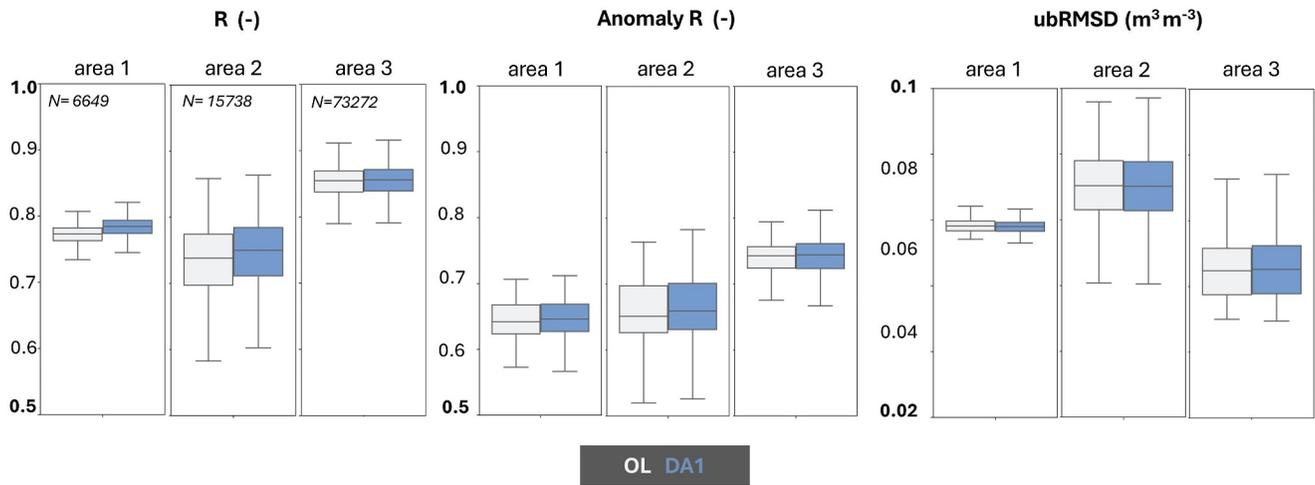


Figure 8. Evaluation of AquaCrop surface soil moisture for OL and DA1 (SSM_{ac}) with (SSM_{smap}) for area 1, area 2, and area 3 for Pearson's R, anomaly R and unbiased root mean square difference. The clay pixels have been removed from the analysis.

4.2. Relationship Between Soil Moisture (Parameters) and γ^0

A larger γ^0 forecast uncertainty leads to stronger updates in SSM_{ac} and CB, and the conversion from γ^0 to soil moisture updates is also largely determined by the WCM equations. The γ^0 forecast uncertainty is influenced by the soil moisture in two ways. First, the γ^0 forecast uncertainty is determined by the ensemble uncertainty in

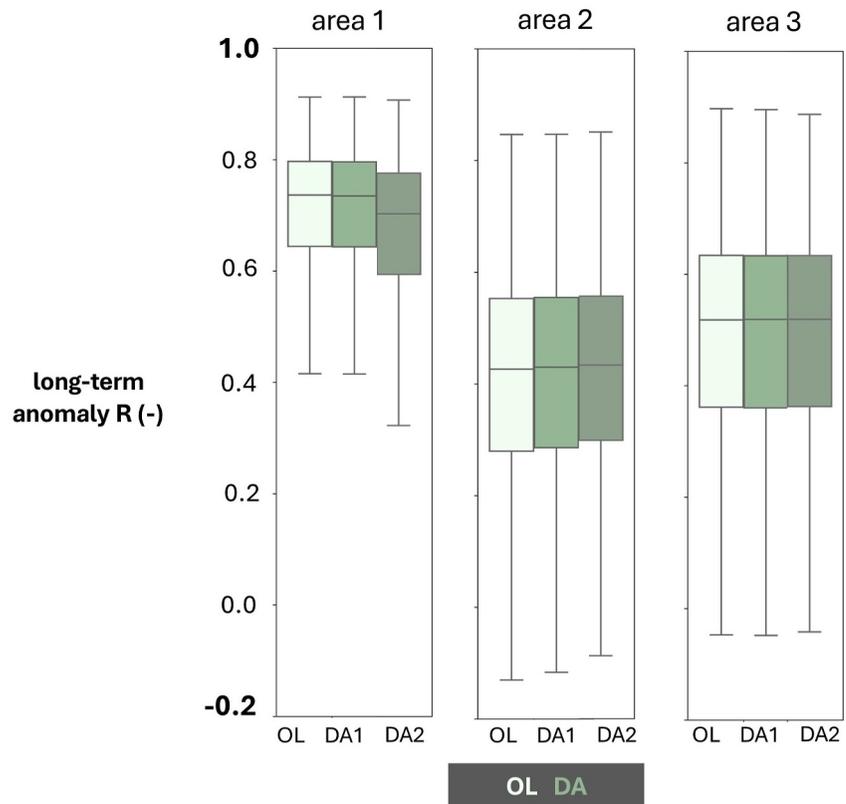


Figure 9. Long-term anomaly correlation of DMP_{ac} (rate) with DMP_{cglis} for area 1, area 2 and area 3. For each area, the left box evaluates the OL, whereas the two dark green boxes evaluate DA1 and DA2 for the months January until the peak of DMP. The clay pixels have been removed from the analysis.

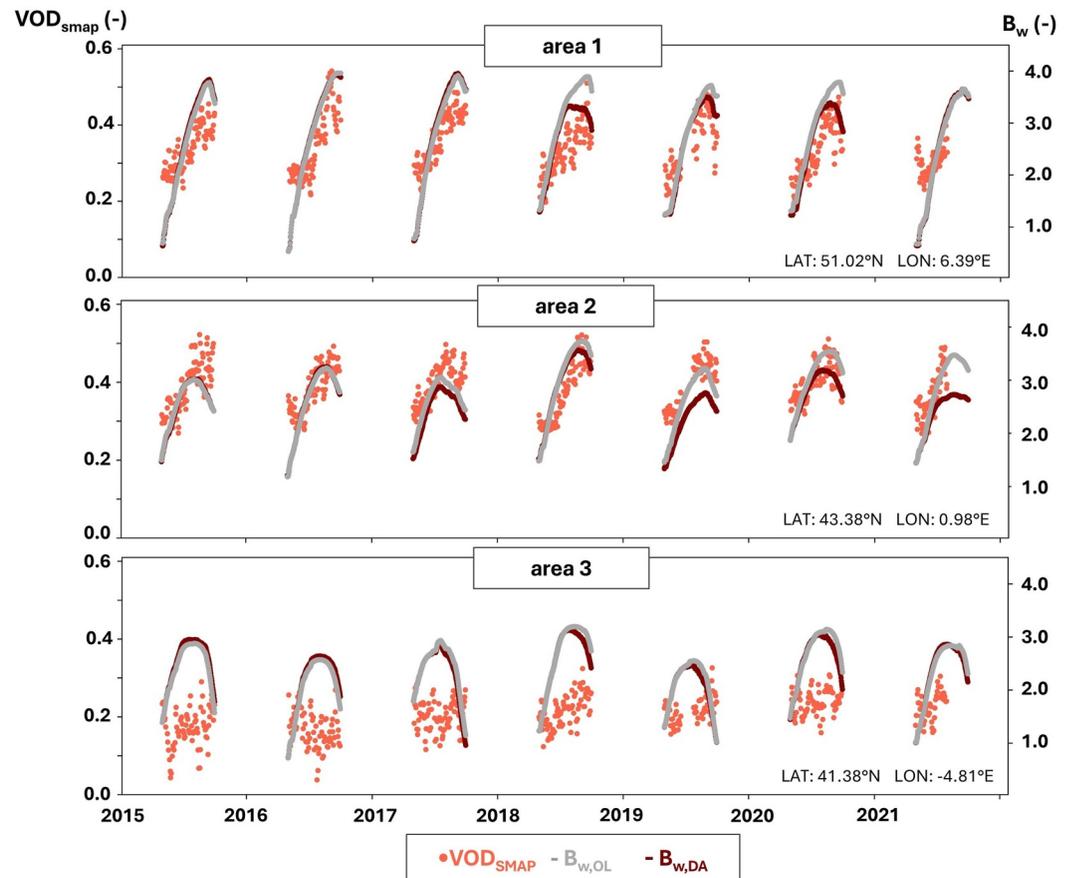


Figure 10. B_w time series for the OL (gray) and DA2 (red) runs of Experiment 2 for each area (individual points), set against the VOD retrievals for the months May–July, to compare the slopes caused by vegetation between the simulations and the evaluation data.

SSM_{ac} , which is in turn determined by soil texture parameters (θ_{FC} and θ_{WP}) of the AquaCrop model. The time-averaged ensemble spread (standard deviation, SD) in SSM_{ac} is visualized on the right side of Figures 12d–12f, revealing the strong differences between different soil textural classes within the areas. The spread is often the largest for pixels with clay dominated soils, mostly in the second and third soil moisture compartments (not shown): clay soils have a larger storage memory due to their low conductivity rate. By design, the state variables with larger spreads attract larger increments, via the $COV(\hat{x}_i^-, \hat{y}_i^-)$ term in the Kalman gain (Equation 5), explaining why a strong DA impact on SSM_{ac} was found over clay pixels.

The second influence on the γ^0 forecast spread relates to the transformation of the SSM_{ac} to γ_{soil}^0 . This relationship as defined by the WCM is described by Equations A1 and A2 and depends on the C and D parameters, which are calibrated for each gridcell individually. Figures 12a–12c reveals a linear relationship between the calibrated slope parameter D of the WCM and the temporally averaged ensemble SD of the forecasted γ^0 from DA2. The higher the value for D , the wider the possible range in γ^0 , that is, the larger $VAR(\hat{y}_i^-)$ in Equation 5, regardless of the ensemble spread in SSM_{ac} or B_w introduced by perturbations. The slope of this linear relationship is dependent on the soil texture and explains the DA impact on soil moisture shown in Figure 5. The upper magnitude of the forecasted γ^0 spread is determined by the strong double boundaries (θ_{WP} and θ_{FC}) of the conceptual water balance. For example, silty soils have a larger range between θ_{FC} and θ_{WP} than other soil textures. In the conceptual AquaCrop water balance, this can result in large fluctuations of the soil moisture in the top compartments, which are much stronger than the fluctuations observed in the γ^0 signal. The effect of SSM_{ac} is then compensated in the WCM calibration by a small D value which subsequently results in a small modeled γ^0 spread. This observation

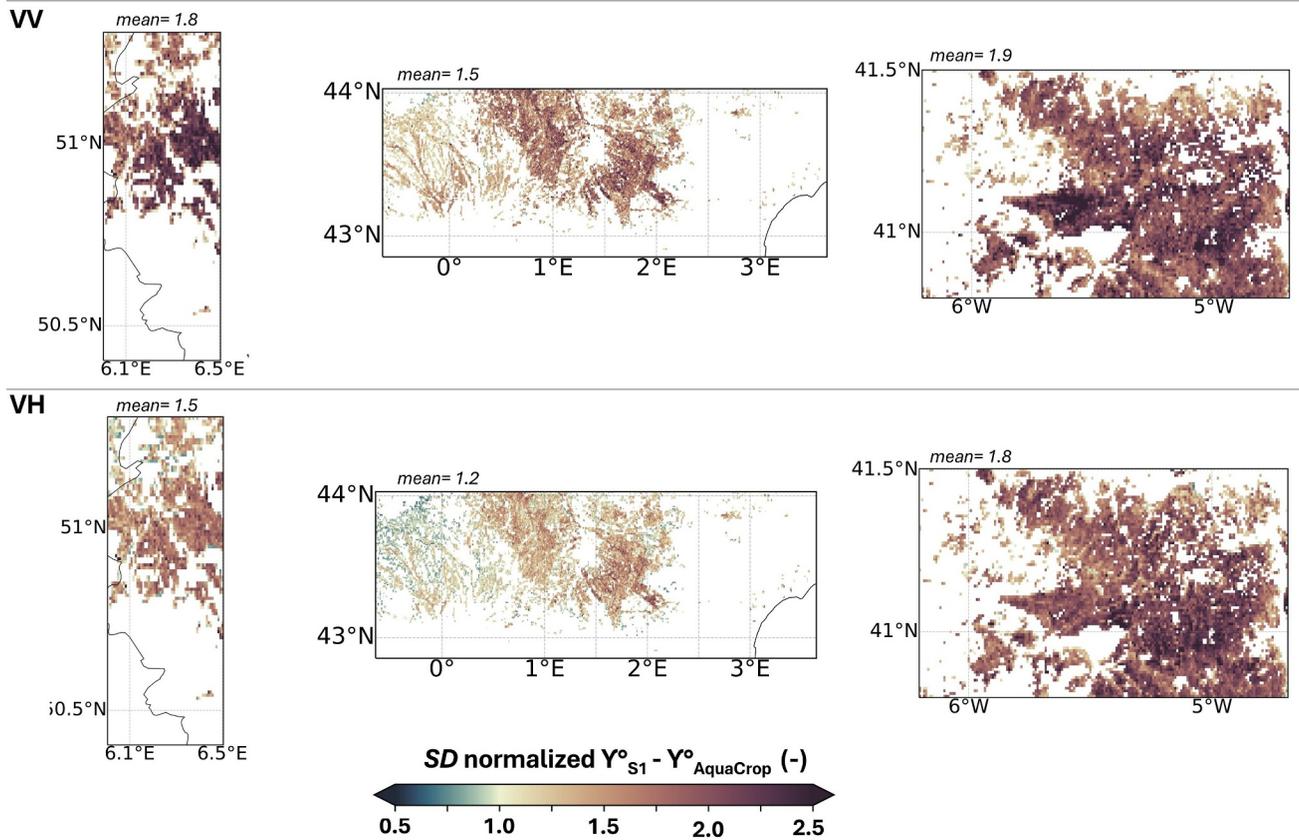


Figure 11. $SD(N_{innov})$ for three study areas, for (a) γ_{VV}^0 and (b) γ_{VH}^0 for DA2, computed for the entire evaluation period (2015–2021).

coincides with the results in de Roos et al., 2023, where the WCM sets a small value for D to counterbalance the considerable AquaCrop soil moisture fluctuations in time.

The large scatter in $SD \gamma_{VV}^0$ (Figures 12b and 12c) for pixels with clay dominated soils is likely due to unrealistic AquaCrop model simulations of SSM_{ac} . The soils steadily dried out over the spring and summer months due to low infiltration capacity, and the WCM optimization maximized the vegetation contribution and minimized soil moisture contribution with low values for the slope parameter D . The low D values further amplify the large DA effect on the SSM_{ac} in clay areas, that is, small changes in observed backscatter lead to large changes in SSM_{ac} .

4.3. Relationship Between Vegetation (Parameters) and γ^0

The conversion from γ^0 to vegetation updates is largely determined by nonlinear and time-varying relationships in the WCM. The WCM contains two vegetation parameters, (a) parameter A related to the direct backscatter contribution of vegetation to the total γ^0 signal and (b) parameter (B) related to the attenuation by vegetation (T^2) of the soil contribution, which can be found in Equations A3 and A4. The parameter values vary per location as all pixels were calibrated individually, and determine how the vegetation affects the γ^0 simulations, or conversely, how the γ^0 DA updates vegetation. No clear relationship between A and the mapping of backscatter innovations to CB increments is found.

The effect of parameter B on the correlation between γ_{VH}^0 innovations and CB increments ($R_{[INCR, INNOV]}$) is visualized in Figure 13, for the early growing season (JFM) and the late growing season (MJJ). The $R_{[INCR, INNOV]}$ indicates if large γ_{VH}^0 innovations actually result in large updates in the DA2 system. At the start of the season, $R_{[INCR, INNOV]}$ are dominantly negative and the relationship between $R_{[INCR, INNOV]}$ and B is negative. When the

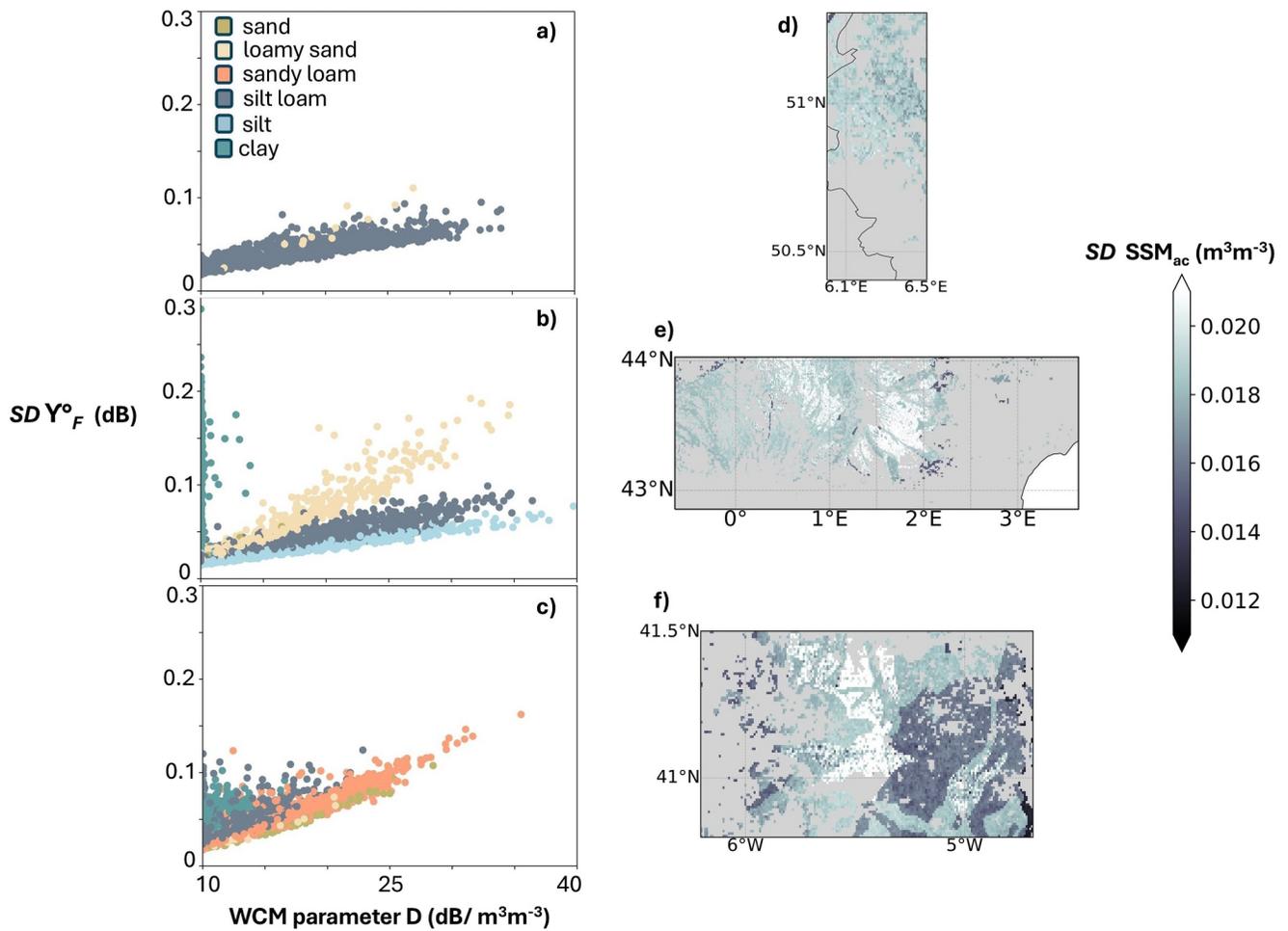


Figure 12. Scatter plots on the left show the forecast backscatter ensemble SD (DA2 SD_{VV}^0) and the Water Cloud Model soil slope D parameter for the three study areas (a–c) for the entire SSM evaluation period, with colors indicating soil textures, and the right side shows the corresponding spatial maps of SSM_{ac} ensemble SD (d–f).

vegetation grows, the $R_{[INCR,INNOV]}$ values are positive and the relationship with the $R_{[INCR,INNOV]}$ becomes positive, as can be seen in the months May–July (MJJ).

This is in line with the behavior of the WCM, described by Bechtold et al. (2023). Low vegetation and high soil moisture (JFM) cause only a small contribution of direct vegetation scatter and only a minimal attenuation of the soil signal, that is, the soil signal dominates. Consequently, the total backscatter is not tightly connected to vegetation and a positive backscatter innovation is not translated into an increase in vegetation. For higher WCM- B parameters, the backscatter attenuation due to vegetation will be relatively stronger than the direct vegetation scattering component (Appendix A, Equation A3), and consequently positive backscatter innovations would be associated with reduced vegetation.

As the CB increases and SSM_{ac} decreases (MJJ), the contribution of direct vegetation increases and becomes larger than the attenuated soil backscatter. A positive backscatter innovation can then be directly related to a positive CB increment. Higher values for WCM parameter B further enlarge the positive relationship by reducing even more the relative soil contribution and attributing most of the total backscatter signal to vegetation (and thus leading to strongly positive $R_{[INCR,INNOV]}$).

5. Discussion

This study presented a first assessment of regional S1 backscatter assimilation within the AquaCrop model integrated into NASA's LIS framework. The quality of the results is largely influenced by both the regional

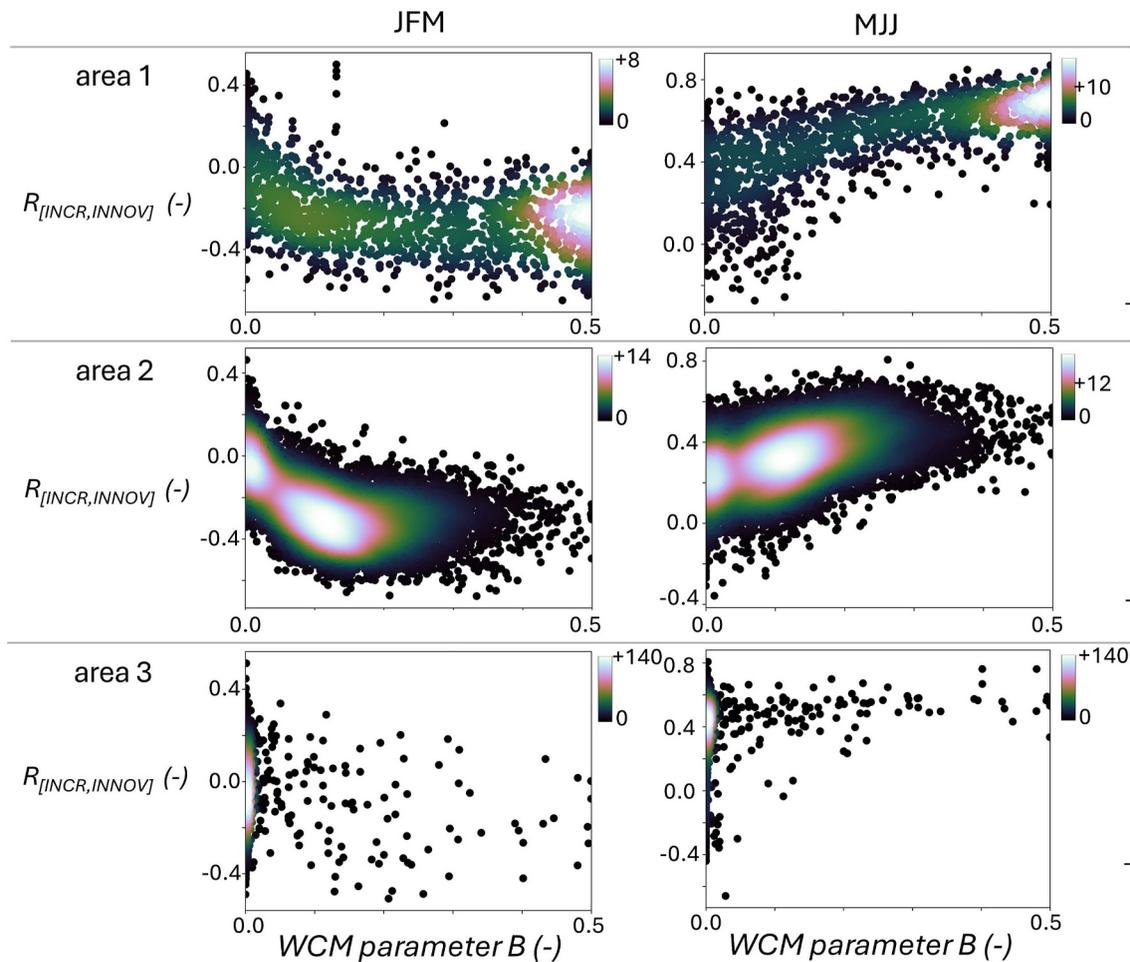


Figure 13. Scatter plots between the correlation of the γ_{VH}^0 innovations with CB increments ($R_{[INCR, INNOV]}$; $-$) and Water Cloud Model parameter B ($-$), for the first part of the growing season; January–March (JFM) and for the later stage of the growing season, March–July (MJJ), for all years, Experiment 2. The pixels with clay soils are removed from the analysis.

AquaCrop model setup and the DA system setup, which will be discussed separately before discussing the DA results for soil moisture and biomass.

5.1. Model Setup

The model setup consisted of several generalizations, most importantly related to homogeneous generic crop parameters considered over all three regions, a simple spatially variable representation of the soil (single soil horizon, no impermeable layers or groundwater table) and downscaled meteorological input data based on $0.5^\circ \times 0.625^\circ$ re-analysis data. These generalizations inevitably led to local model errors, which became apparent for example, over the regions with dominantly clay soils. Apart from the model, the evaluation data sets also contain limitations. The satellite-based reference data are retrievals, based on several assumptions and might observe a quantity that deviates from what the model simulates. The in situ data are point-based and differ in spatial support from that of the simulations.

Despite these limitations and generalizations, overall good temporal correlations were found between the evaluation products and model only (OL) runs, in terms of SSM_{ac} and DMP, for all study areas. The AquaCrop model does not perform well over clay regions in area 2 and 3 because for some years, no vegetation growth could be simulated due to water stress. In addition, the model often simulates overestimated DMP_{ac} compared to the DMP_{egls} product, mostly in the regions with drier climates, and this is likely related to the parameters of

the generic C_3 crop. The evaluation with B_w gave mixed results for the different regions. Especially, for area 3, the growth in B_w did not coincide with the VOD data, resulting in negative temporal correlations. It could be that the seasonal cycle of the generic crop is not representative of the vegetation in the drier climates, but it should also be noted that the VOD data of area 3 was much more noisy and incoherent compared to the other areas.

5.2. Experimental DA Setup

The three top soil moisture compartments of AquaCrop were combined using a weighted average, to represent the SSM_{ac} . The aggregation of the three top compartments facilitates a direct update of soil moisture in three layers, which is beneficial to get updates when the top layer reaches upper or lower boundaries, but it can also be harmful. Due to the free drainage at the bottom of the soil profile in the conceptual water balance of AquaCrop, the spread in the deep layers can become unrealistically high and attract large updates. The weights of the weighted average need further investigation, because they were now empirically set without any calibration to create realistic SSM_{ac} time series compared to some independent soil moisture estimates (trial and error), but the optimal weights will naturally differ with texture class, climatological region and with the soil moisture reference data (e.g., C-band and L-band soil moisture retrievals relate to different surface layer depths).

Overall, the S1 DA introduced SSM_{ac} improvements, mainly for areas 1 and 2. In areas with clayey soils, the updates were strong and could either improve or deteriorate the results. The steady decrease in SSM_{ac} as a result of low infiltration capacity for larger rainfall events in clay soils during the spring and summer resulted in a poor performance of the WCM, in which the effect of the soil moisture on γ^0 simulations for some regions was minimized with low values for D . Low WCM- B parameters, in turn, lead to large increments in soil moisture for small γ^0 innovations.

The success of inferring good increments from S1 γ_{VV}^0 observations generally depends on the performance of the observation operator, including the diagnosis of B_w and the calibrated parameters of the WCM. The generic crop and the associated B_w were set up for crops in the summer growing season, which is not suitable for winter cereals that are in fact commonly cultivated in the Ruhr area, that is, area 1 (Blickensdörfer et al., 2022). The marginal DA impact on the CB in DA2 is further be related to the fact that only small increments are applied at the start of the growing season (Section 4.3). To have a stronger DA impact on the vegetative phase of the crop development, it could be beneficial to also include parameter updating to define the start of the growing season or emergence.

Additionally, the WCM parameters were assumed to remain constant in time, while in reality a certain fraction of crop rotation is to be expected at a 1-km spatial resolution, which would likely have reduced the accuracy of the forward operator. Presenting the model variables in the WCM in the most optimal way is of crucial importance. Alternatively, more detailed forward operators might give better results and additionally the use of machine learning to derive optimal simulated σ^0 has also shown potential (de Roos et al., 2023; Rains et al., 2021).

The DA diagnostics further revealed that the DA system could still be optimized, possibly by calibrating the observation perturbations per location and for each polarization. The observation error could be increased, as it also includes the representativeness error of the WCM (van Leeuwen, 2015). However, as was concluded from de Roos et al. (2023), most of the difference between the simulated and observed backscatter comes from the AquaCrop model errors.

5.3. DA Impact on Soil Moisture and Biomass

The results of DA1 showed overall slight improvements in SSM_{ac} with minimal impact on CB. The small impact on SSM_{ac} can be attributed to (a) the conceptual water balance of AquaCrop, with upper and lower thresholds of wilting point and field capacity limiting updates in very dry or wet periods, and (b) the DA system setup using satellite backscatter instead of retrievals. The assimilation of Sentinel-1 soil moisture retrievals in crop models has shown to be effective in other studies, especially when combined with optical-based remote sensing data for vegetation (Pan et al., 2019; Rojas-Munoz et al., 2023). But the use of SAR observations for correcting crop models is not as common and so far more reported for hydrological applications (Bechtold et al., 2023; Sun et al., 2021). Furthermore, the AquaCrop CB has a particular sensitivity to soil moisture: crop water stress is initiated when a fraction of the water availability between θ_{FC} and θ_{WP} of the root zone is exceeded.

Improvements in the temporal variation of SSM_{ac} due to DA do therefore not automatically result in improvements in CB estimates, because this depends on the total available water content in the root zone, which is strongly dependent on the soil texture.

The use of S1 γ_{VH}^0 to update AquaCrop dry CB simulations (DA2) is particularly challenging because S1 observations are sensitive to the B_w . The updated CB in DA2 is no longer always monotonically increasing (mainly not in warmer areas) and the (negative) changes in CB no longer reflect DMP_{cglis} rates that can be compared to satellite-based DMP_{cglis} estimates. S1 DA can introduce a decrease in CB, because biomass might indeed have been removed from the pixel (e.g., harvest), and also because the DA corrects the (unobserved) CB for any errors in the conversion to B_w as part of the observation operator.

The comparison of time series of B_w (instead of CB) and VOD_{smap} showed potential, but while the agreement between observed VOD and simulated B_w was sometimes high, there was no significant improvement in the inter-annual variation introduced by the DA. The changes in B_w after DA were often small, and the VOD product showed to contain noise, making it difficult to draw conclusions from this evaluation.

This study set up a technical data assimilation framework with the aim to understand how Sentinel-1 data can be used to improve AquaCrop model simulations. The impact analysis showed the effect of DA, but the performance analysis relative to independent data did not show a significant improvement in the results. Unlike the more common assimilation of brightness temperatures from passive microwave satellites, or soil moisture retrievals, the assimilation of Sentinel-1 backscatter to update both soil moisture and vegetation in crop models is still relatively unexplored, and needs further research. However, Sentinel-1 offers the advantages of going to much higher resolutions, which offers potential for new insights in agricultural research. As the integration of AquaCrop into the LIS framework is new, it is anticipated that continued research can further improve the DA system by leveraging the findings of this study on how backscatter and variables in crop models are related.

6. Conclusions

In this study, S1 backscatter data were assimilated into AquaCrop within the NASA LIS system. AquaCrop was run over three areas in Europe, using coarse meteorological re-analysis data, and inevitably uncertain crop and soil parameters, which lead to forecast errors in soil moisture and biomass. Via DA, the aim was to reduce these errors. In one DA experiment, only top soil moisture was updated using γ_{VV}^0 . In a second experiment, both soil moisture and cumulative CB were updated through a combined assimilation of S1 γ_{VV}^0 for soil moisture updating and γ_{VH}^0 for CB updating.

The assimilation of γ_{VV}^0 can improve the simulation of soil moisture in the top soil of AquaCrop. However, problems arose over homogeneous clay soils, which are generally unsuitable soils for most crops, and led to unrealistic modeled crop growth and unrealistic parameters in the observation operator that consequently introduced undesirable soil moisture updates. Furthermore, in general, the strong effects of the soil moisture boundaries of the conceptual water balance of AquaCrop (field capacity and wilting point) caused updates mainly during the beginning and end of the growing season, but only very limited updates during very dry or wet conditions in summer and winter.

Only updating the soil moisture in the first DA experiment did not translate into an improved performance of biomass, based on a comparison the DMP_{ac} simulations against an independent satellite-based DMP_{cglis} . In general, the impact on CB is limited since the CB simulations in AquaCrop are only affected by soil moisture updates in case the water stress threshold is reached during the vegetative growth stage.

The second experiment showed a larger impact on CB, but the results varied spatially. First, linking AquaCrop biomass to γ_{VH}^0 includes two uncertain modeling steps: (a) the conversion from the AquaCrop CB into wet biomass (B_w), and (b) the forward operator, the WCM, relying on relatively simplistic calculations. Additionally, the WCM parameters were assumed to remain constant over time while in reality, cropping practices are variable over time. The modeling deficiencies of the observation operator might result in a poor propagation of the γ_{VH}^0 innovations to CB increments. Second, a large-scale evaluation of the inter-annual variation in dry or wet biomass

with satellite products is non-trivial, because each product represents particular plant features that are not necessarily modeled accordingly by AquaCrop.

It is now technically possible to support a large-scale S1 DA system for the crop model AquaCrop. However, our study indicates that improvements in the local and seasonal characterization of crop and soil parameters is needed to better capture high-resolution soil-plant processes and backscatter interactions, and to eventually improve crop growth through the assimilation of S1 data.

Appendix A: Water Cloud Model

The Water Cloud Model (WCM; Attema & Ulaby, 1978) considers the total γ^0 signal as the sum of the backscatter from the soil (γ_{soil}^0), attenuated by vegetation (T^2) and the backscatter coming from the water-holding medium above the soil, that is, vegetation over cropland (γ_{veg}^0). Similar to de Roos et al. (2023), this study excluded the incidence angle dependency, because the latter was also removed from the S1 observations during the terrain flattening processing step. However, this processing step does not remove potential effects of canopy roughness and differences in travel length of the signal through vegetation, which ultimately will add to the observation error. The equations of the WCM are given below, with Equations A1, A3, A4 written in linear scale and Equation A2 defined in dB (10 log(.)), which is converted back to linear scale when used as input for Equation A1 (unit conversion not shown). The WCM requires the calibration of four parameters, which was done for each grid cell. The parameters C (dB), related to surface roughness, and the slope parameter D (dB/m³ m⁻³) are required for γ_{soil}^0 (Equation A3). The vegetation parameter A (–) relates to vegetation scattering (γ_{veg}^0 ; Equation A3), and B (–) to vegetation attenuation (Equation A4), affecting γ_{soil}^0 . As explained in de Roos et al. (2023), the AquaCrop CB was transformed to wet biomass B_w by considering canopy decline based on temperature and water stress as discussed above (Section 2.2.4), and $V_1 = V_2 = B_w$.

The optimization was done using a Bayesian objective function that minimizes the sum of squared error between the S1 observed γ^0 and the simulated γ^0 , and also uses a prior parameter constraint. The Differential Evolution Adaptive Metropolis (Laloy & Vrugt, 2012; Vrugt, 2016) was used as optimization algorithm. A more detailed description of the WCM optimization can be found in de Roos et al. (2023).

$$\gamma^0 = T^2 \cdot \gamma_{soil}^0 + \gamma_{veg}^0 \quad (A1)$$

$$\gamma_{soil}^0 = C + D \cdot SSM_{ac} \quad (A2)$$

$$\gamma_{veg}^0 = A \cdot V_1 (1 - T^2) \quad (A3)$$

$$T^2 = \exp(-2 \cdot B \cdot V_2) \quad (A4)$$

The WCM calibration was executed for the months January up to the end of July, to capture the soil moisture variations which are more pronounced in the early months (low vegetation cover), and the vegetative stage of the crop later in the main growing season. The uncertain stage of the decline in B_w at the end of the season was hereby largely excluded. The entire evaluation period (2015–2021) was used for calibration (as opposed to only the first 4 years in de Roos et al., 2023) to maximally include the effect of crop variations over time for a single location. The four parameters were calibrated for each grid cell and both γ_{VV}^0 and γ_{VH}^0 separately, using the deterministic AquaCrop model output of SSM_{ac} and B_w , as explained in de Roos et al. (2023).

Appendix B: Time Series for DA Experiments Within Area 1 and 2

Figures B1 and B2 show similar time series as Figure 4 to visualize the effect of DA for both experiments, but for different single locations in study areas 1 and 2.

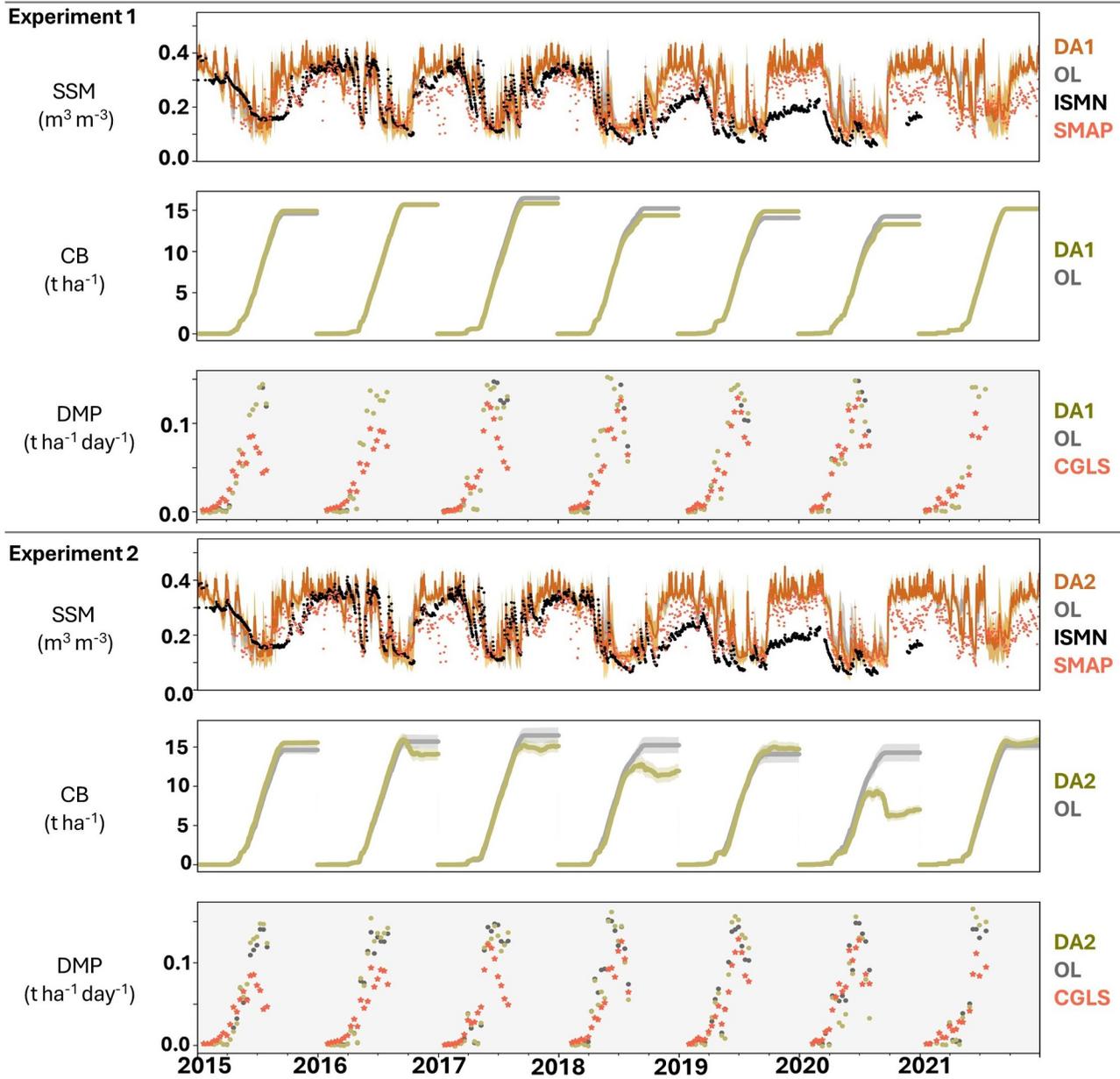


Figure B1. Time series of SSM_{ac} , CB and daily biomass productivity (OL and DA) for Experiment 1 (top) and Experiment 2 (bottom), of point $50.9^{\circ}N$ and $6.32^{\circ}E$, situated in area 1. Daily simulations of SSM_{ac} and biomass are shown as lines, and, if it concerns an updated perturbed variable, the spread of the model ensembles is shown as standard deviation with a 95% confidence interval. The productivity plots (DMP) show less frequent results because they are matched to the 10-daily time steps of the evaluation product.

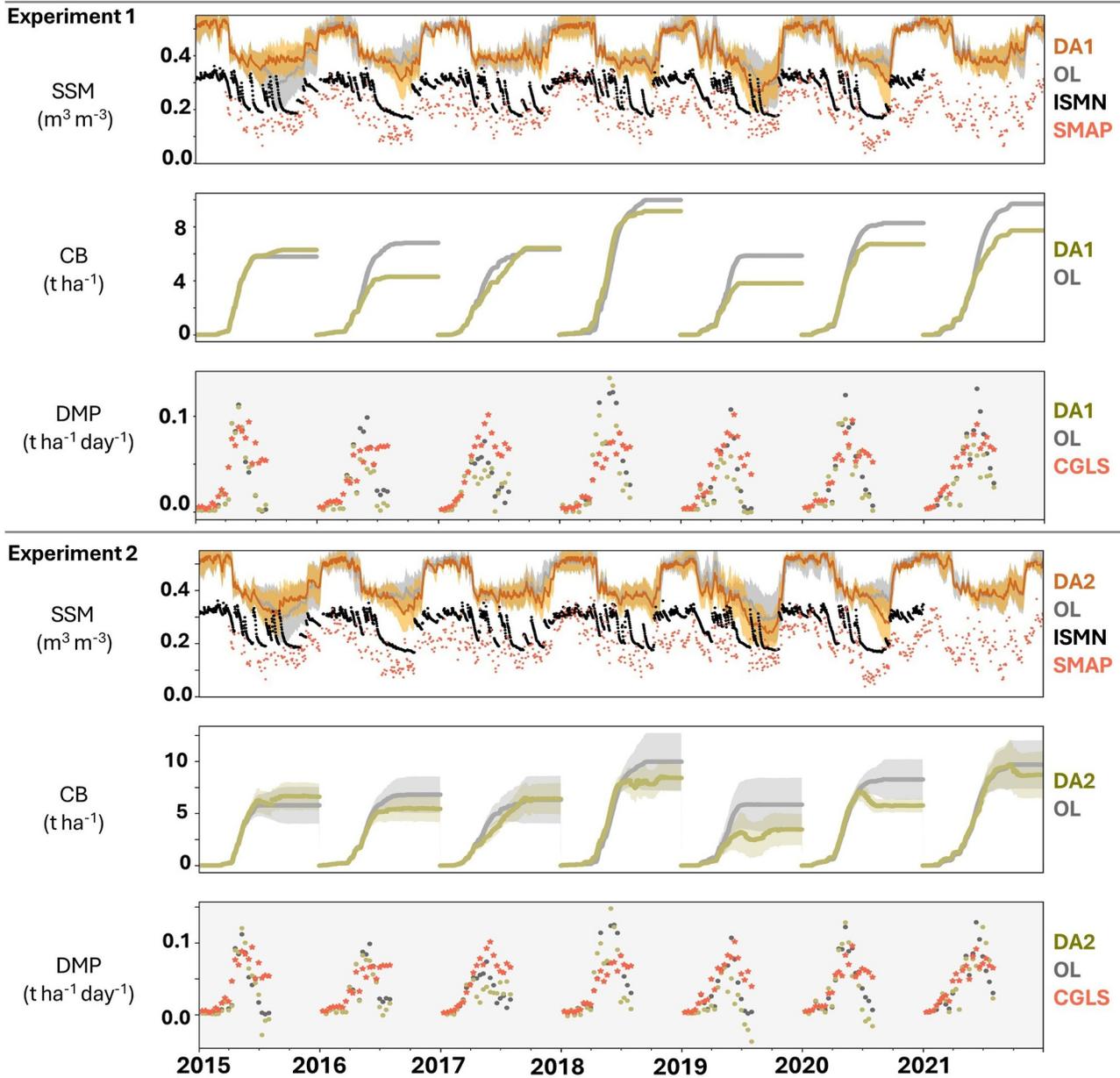


Figure B2. Time series of SSM_{ac} , CB and daily biomass productivity (OL and DA) for Experiment 1 (top) and Experiment 2 (bottom), of point 43.44°N and 1.88°E , situated in area 2 (clay-textured soil). Daily simulations of SSM_{ac} and biomass are shown as lines, and, if it concerns an updated variable, the spread of the model ensembles is shown as standard deviation with a 95% confidence interval. The productivity plots (DMP) are less because they are matched to the 10-daily time steps of the evaluation product.

Appendix C: SMAP Soil Moisture Evaluation

Figure C1 shows SSM correlation difference maps for areas 1 (I) and 2 (II). It compares AquaCrop simulations with SMAP downscaled retrievals and ISMN locations, similar to what is shown in Figure 7 for area 3.

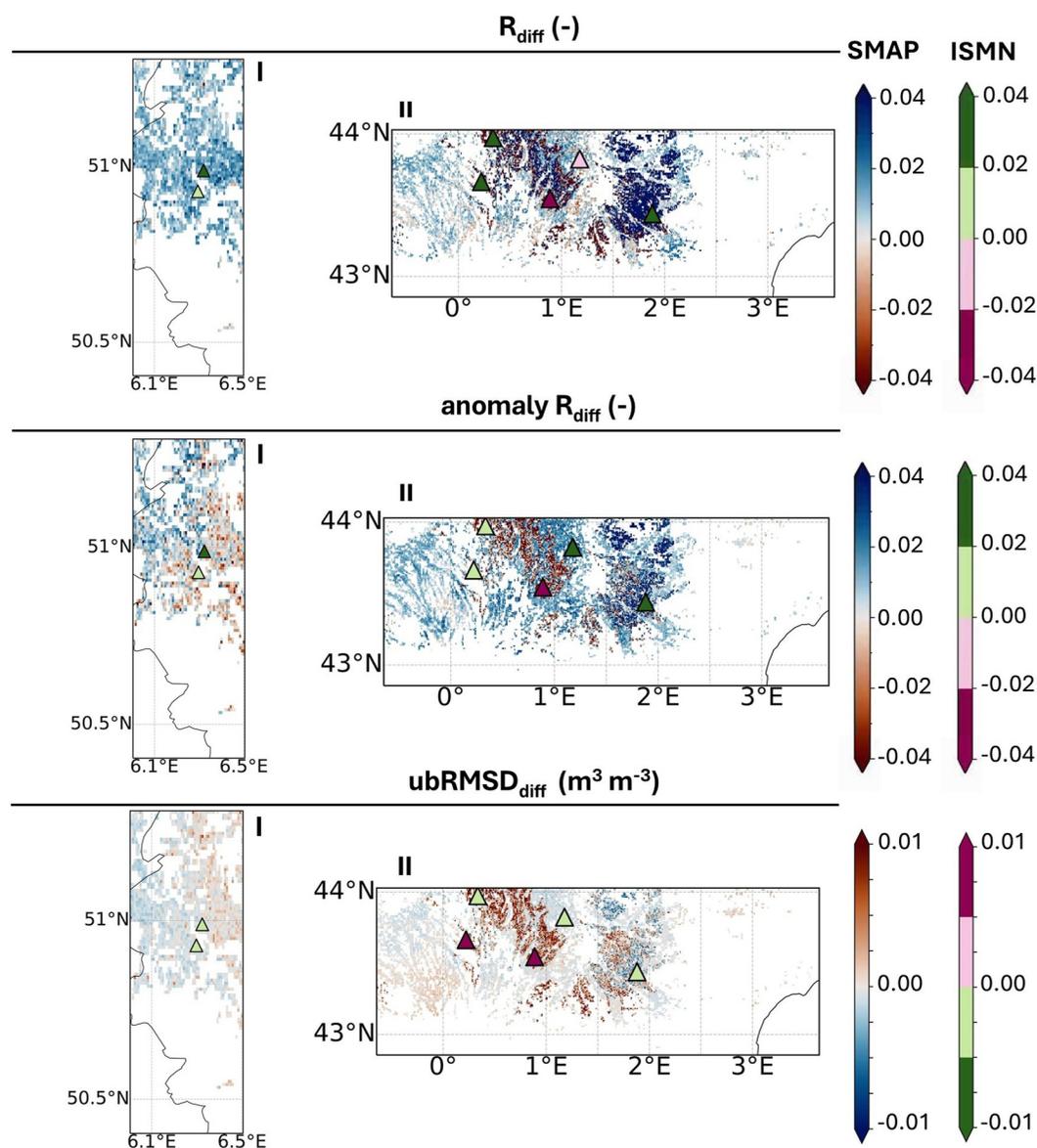


Figure C1. Difference maps (DA-OL) between ensemble mean model simulations and soil moisture active passive for area 1 (I; left) and area 2 (II; right). The differences between SSM_{ac} and SSM_{ismn} for the same metrics are indicated as triangles. Red/pink colors indicate a reduced performance due to DA, blue/green indicate an increased performance in the DA runs.

Data Availability Statement

The integration of the AquaCrop code into LIS has not yet been officially released, but the AquaCrop source code can be found on the FAO website, <https://www.fao.org/aquacrop/en/>. The following repository includes the generic crop file and management file; <https://doi.org/10.1002/2014MS000330>. The water cloud model (WCM) calibration scripts can be found here: https://github.com/KUL-RSDA/obs_operator_calibration. All data that were used for model input and evaluation are freely available online. Please visit the following links for data access. MERRA-2 variables: <https://disc.gsfc.nasa.gov/datasets?project=MERRA-2> (last access: 1 Jan 2022, Global Modeling and Assimilation office, 2015a, <https://doi.org/10.5067/VJAFPLI1CSIV>, 2015b, <https://doi.org/10.5067/RKPHT8KC1Y1T>); the soil mineral classification and organic matter from HWSDv1.2: <http://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/> (last access: 30 August 2019, FAO/IIASA/ISRIC/ISSCAS/JRC, 2012); the CORINE land cover map: <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012?tab=mapview> (last access: 5 September 2019, Copernicus Global

Land Service, 2018); reference data set DMP_{egls}: <https://land.copernicus.eu/global/products/dmp> (last access: 1 June 2023, Copernicus Global Land Service, 2019a); ISMN soil moisture from various observatories: <https://ismn.geo.tuwien.ac.at/en/> (last access: 1 Jan 2023); the SMAP 1-km soil moisture from the NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC): <https://nsidc.org/data/nsidc-0779/versions/1> (last access: 1 July 2023); the VOD_{smap} product, version 4: <https://zenodo.org/records/5579549> (last accessed: September 2023).

Acknowledgments

The authors would like to acknowledge all contributors to the conversion of the AquaCrop source code from Pascal to Fortran; Maxime Van den Bossche, Jonas Mortelmans, Samuel Scherrer and Zdenko Heyvaert, and the help from AquaCrop developer Dirk Raes. We would further like to acknowledge the feedback provided by Susan Steele-Dunne about the data assimilation methodology. Finally, we would like to thank the editors and reviewers for taking the time to review this manuscript. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation–Flanders (FWO) and the Flemish Government. This research has been supported by the European Commission, Horizon 2020 Framework Programme (SHui Grant 773903), internal KU Leuven funds (C14/21/057), and the BELSPO project CropWaves (SR/00/412). Louise Busschaert is funded by FWO Grant 1158423N.

References

- Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., et al. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, 12(6), 1323–1337. <https://doi.org/10.5194/hess-12-1323-2008>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration-guidelines for computing crop water requirements (FAO Irrigation and drainage paper No. 56)*. Food and Agriculture Organization of the United Nations.
- Allies, A., Roumiguié, A., Fieuzal, R., Dejoux, J.-F., Jacquin, A., Veloso, A., et al. (2022). Assimilation of multisensor optical and multiorbital SAR satellite data in a simplified agrometeorological model for rapeseed crops monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 1123–1138. <https://doi.org/10.1109/JSTARS.2021.3136289>
- Attema, E., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. *Radio Science*, 13(2), 357–364. <https://doi.org/10.1029/RS013i002p00357>
- Balenzano, A., Mattia, F., Satalino, G., Lovergine, F. P., Palmisano, D., Peng, J., et al. (2021). Sentinel-1 soil moisture at 1 km resolution: A validation study. *Remote Sensing of Environment*, 263, 112554. <https://doi.org/10.1016/j.rse.2021.112554>
- Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., et al. (2018). Toward global soil moisture monitoring with sentinel-1: Harnessing assets and overcoming obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 520–539. <https://doi.org/10.1109/tgrs.2018.2858004>
- Bechtold, M., Modanesi, S., Lievens, H., Baguis, P., Brangers, I., Carrassi, A., et al. (2023). Assimilation of sentinel-1 backscatter into a land surface model with river routing and its impact on streamflow simulations in two Belgian catchments. *Journal of Hydrometeorology*, 24(12), 2389–2408. <https://doi.org/10.1175/JHM-D-22-0198.1>
- Betbeder, J., Fieuzal, R., & Baup, F. (2016). Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2540–2553. <https://doi.org/10.1109/JSTARS.2016.2541169>
- Bhogapurapu, N., Dey, S., Mandal, D., Bhattacharya, A., Karthikeyan, L., McNairn, H., & Rao, Y. (2022). Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data. *Remote Sensing of Environment*, 271, 112900. <https://doi.org/10.1016/j.rse.2022.112900>
- Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmí, S., & Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment*, 269, 112831. <https://doi.org/10.1016/j.rse.2021.112831>
- Bogena, H., Montzka, C., Huisman, J., Graf, A., Schmidt, M., Stockinger, M., et al. (2018). The TERENO-Rur hydrological observatory: A multiscale multi-compartment research platform for the advancement of hydrological science. *Vadose Zone Journal*, 17(1), 180055–180122. <https://doi.org/10.2136/vzj2018.03.0055>
- Busschaert, L., De Roos, S., Thiery, W., Raes, D., & De Lannoy, G. J. (2022). Net irrigation requirement under different climate scenarios using AquaCrop over Europe. *Hydrology and Earth System Sciences*, 26(14), 3731–3752. <https://doi.org/10.5194/hess-26-3731-2022>
- Büttner, G. (2014). Corine land cover and land cover change products. In I. Manakos & M. Braun (Eds.), *Land use and land cover mapping in europe* (pp. 55–74). Springer. <https://doi.org/10.1007/978-94-007-7969-3>
- Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., & Piguet, B. (2007). In situ soil moisture observations for the CAL/VAL of SMOS: The smosmania network. In *2007 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1196–1199). <https://doi.org/10.1109/IGARSS.2007.4423019>
- Chakrabarti, S., Bongiovanni, T., Judge, J., Zotarelli, L., & Bayer, C. (2014). Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(9), 3867–3879. <https://doi.org/10.1109/JSTARS.2014.2315999>
- Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Keravec, P., Marciel, B., et al. (2009). Spatialization of crop leaf area index and biomass by combining a simple crop model SAFY and high spatial and temporal resolutions remote sensing data. *2009 IEEE International Geoscience and Remote Sensing Symposium*, 3, III-478. <https://doi.org/10.1109/IGARSS.2009.5418296>
- De Lannoy, G. J., & Reichle, R. H. (2016). Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *Journal of Hydrometeorology*, 17(2), 669–691. <https://doi.org/10.1175/jhm-d-15-0037.1>
- de Roos, S., Busschaert, L., Lievens, H., Bechtold, M., & De Lannoy, G. J. (2023). Optimisation of aquacrop backscatter simulations using Sentinel-1 observations. *Remote Sensing of Environment*, 294, 113621. <https://doi.org/10.1016/j.rse.2023.113621>
- de Roos, S., De Lannoy, G. J., & Raes, D. (2021). Performance analysis of regional aquacrop (v6. 1) biomass and surface soil moisture simulations using satellite and in situ observations. *Geoscientific Model Development*, 14(12), 7309–7328. <https://doi.org/10.5194/gmd-14-7309-2021>
- Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The international soil moisture network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, 15(5), 1675–1698. <https://doi.org/10.5194/hess-15-1675-2011>
- Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K., Büchner, M., et al. (2015). The global gridded crop model intercomparison: Data and modeling protocols for phase 1 (v1. 0). *Geoscientific Model Development*, 8(2), 261–277. <https://doi.org/10.5194/gmd-8-261-2015>
- Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367. <https://doi.org/10.1007/s10236-003-0036-9>
- Faivre, R., Leenhardt, D., Voltz, M., Benoît, M., Papy, F., Dedieu, G., & Wallach, D. (2009). Spatialising crop models. *Sustainable Agriculture*, 687–705. https://doi.org/10.1007/978-90-481-2666-8_42
- Fang, B., Lakshmi, V., Cosh, M., Liu, P.-W., Bindlish, R., & Jackson, T. J. (2022). A global 1-km downscaled SMAP soil moisture product based on thermal inertia theory. *Vadose Zone Journal*, 21(2), e20182. <https://doi.org/10.1002/vzj2.20182>

- Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., & Cavigelli, M. (2008). Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-maize model. *International Journal of Remote Sensing*, 29(10), 3011–3032. <https://doi.org/10.1080/01431160701408386>
- Farr, T. G., & Kobrick, M. (2000). Shuttle radar topography mission produces a wealth of data. *EOS, Transactions American Geophysical Union*, 81(48), 583–585. <https://doi.org/10.1029/EO081i048p00583>
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45(2). <https://doi.org/10.1029/2005RG000183>
- Feldman, A., Konings, A., Piles, M., & Entekhabi, D. (2021). The multi-temporal dual channel algorithm. (MT-DCA) (version 5) [Dataset]. <https://doi.org/10.5281/zenodo.5619583>
- Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, 12(4), e2018MS001453. <https://doi.org/10.1029/2018MS001453>
- Fuster, B., Sánchez-Zapero, J., Camacho, F., García-Santos, V., Verger, A., Lacaze, R., et al. (2020). Quality assessment of PROBA-V LAI, fAPAR and fCOVER collection 300 m products of copernicus global land service. *Remote Sensing*, 12(6), 1017. <https://doi.org/10.3390/rs12061017>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- González-Zamora, Á., Sánchez, N., Pablos, M., & Martínez-Fernández, J. (2019). Cci soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain. *Remote Sensing of Environment*, 225, 469–482. <https://doi.org/10.1016/j.rse.2018.02.010>
- Hansen, J., & Jones, J. (2000). Scaling-up crop models for climate variability applications. *Agricultural Systems*, 65(1), 43–72. [https://doi.org/10.1016/S0308-521X\(00\)00025-1](https://doi.org/10.1016/S0308-521X(00)00025-1)
- Holzworth, D. P., Snow, V., Janssen, S., Athanasiadis, I. N., Donatelli, M., Hoogenboom, G., et al. (2015). Agricultural production systems modelling and software: Current status and future prospects. *Environmental Modelling & Software*, 72, 276–286. <https://doi.org/10.1016/j.envsoft.2014.12.013>
- Huang, H., Huang, J., Li, X., Zhuo, W., Wu, Y., Niu, Q., et al. (2022). A dataset of winter wheat aboveground biomass in China during 2007–2015 based on data assimilation. *Scientific Data*, 9(1), 200. <https://doi.org/10.1038/s41597-022-01305-6>
- Huang, J., Scherer, L., Lan, K., Chen, F., & Thorp, K. R. (2019). Advancing the application of a model-independent open-source geospatial tool for national-scale spatiotemporal simulations. *Environmental Modelling & Software*, 119, 374–378. <https://doi.org/10.1016/j.envsoft.2019.07.003>
- Jin, N., Tao, B., Ren, W., He, L., Zhang, D., Wang, D., & Yu, Q. (2022). Assimilating remote sensing data into a crop model improves winter wheat yield estimation based on regional irrigation data. *Agricultural Water Management*, 266, 107583. <https://doi.org/10.1016/j.agwat.2022.107583>
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141–152. <https://doi.org/10.1016/j.eja.2017.11.002>
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., et al. (2017). Brief history of agricultural systems modeling. *Agricultural Systems*, 155, 240–254. <https://doi.org/10.1016/j.agsy.2016.05.014>
- Khabbazan, S., Steele-Dunne, S., Vermunt, P., Judge, J., Vreugdenhil, M., & Gao, G. (2022). The influence of surface canopy water on the relationship between l-band backscatter and biophysical variables in agricultural monitoring. *Remote Sensing of Environment*, 268, 112789. <https://doi.org/10.1016/j.rse.2021.112789>
- Konings, A. G., Piles, M., Das, N., & Entekhabi, D. (2017). L-band vegetation optical depth and effective scattering albedo estimation from SMAP. *Remote Sensing of Environment*, 198, 460–470. <https://doi.org/10.1016/j.rse.2017.06.037>
- Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, 21(10), 1402–1415. <https://doi.org/10.1016/j.envsoft.2005.07.004>
- Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., et al. (2008). A land surface data assimilation framework using the land information system: Description and applications. *Advances in Water Resources*, 31(11), 1419–1432. <https://doi.org/10.1016/j.advwatres.2008.01.013>
- Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try dream (ZS) and high-performance computing. *Water Resources Research*, 48(1). <https://doi.org/10.1029/2011WR010608>
- Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., et al. (2019). Snow depth variability in the Northern Hemisphere mountains observed from space. *Nature Communications*, 10(1), 1–12. <https://doi.org/10.1038/s41467-019-12566-y>
- Manivasagam, V., & Rozenstein, O. (2020). Practices for upscaling crop simulation models from field scale to large regions. *Computers and Electronics in Agriculture*, 175, 105554. <https://doi.org/10.1016/j.compag.2020.105554>
- McNairn, H., & Shang, J. (2016). A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. In *Remote sensing and digital image processing* (Vol. 20, pp. 317–340). Springer. https://doi.org/10.1007/978-3-319-47037-5_15
- Miller, D. A., & White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions*, 2(2), 1–26. [https://doi.org/10.1175/1087-3562\(1998\)002\(0001:ACUSMS\)2.3.CO;2](https://doi.org/10.1175/1087-3562(1998)002(0001:ACUSMS)2.3.CO;2)
- Modanesi, S., Massari, C., Bechtold, M., Lievens, H., Tarpanelli, A., Brocca, L., et al. (2022). Challenges and benefits of quantifying irrigation through the assimilation of Sentinel-1 backscatter observations into Noah-MP. *Hydrology and Earth System Sciences*, 26(18), 4685–4706. <https://doi.org/10.5194/hess-26-4685-2022>
- Modanesi, S., Massari, C., Gruber, A., Lievens, H., Tarpanelli, A., Morbidelli, R., & De Lannoy, G. J. M. (2021). Optimizing a backscatter forward operator using Sentinel-1 data over irrigated land. *Hydrology and Earth System Sciences*, 25(12), 6283–6307. <https://doi.org/10.5194/hess-25-6283-2021>
- Palmisano, D., Mattia, F., Balenzano, A., Satalino, G., Pierdicca, N., & Guarnieri, A. V. M. (2020). Sentinel-1 sensitivity to soil moisture at high incidence angle and the impact on retrieval over seasonal crops. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9), 7308–7321. <https://doi.org/10.1109/tgrs.2020.3033887>
- Pan, H., Chen, Z., de Wit, A., & Ren, J. (2019). Joint assimilation of leaf area index and soil moisture from sentinel-1 and sentinel-2 data into the wofost model for winter wheat yield estimation. *Sensors*, 19(14), 3161. <https://doi.org/10.3390/s19143161>
- Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., et al. (2007). High-performance earth system modeling with NASA/GSFC's land information system. *Innovations in Systems and Software Engineering*, 3, 157–165. <https://doi.org/10.1007/s11334-007-0028-x>

- Peters-Lidard, C. D., Kumar, S. V., Mocko, D. M., & Tian, Y. (2011). Estimating evapotranspiration with land data assimilation systems. *Hydrological Processes*, 25(26), 3979–3992. <https://doi.org/10.1002/hyp.8387>
- Prévoit, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., & Brisson, N. (2003). Assimilating optical and radar data into the STICS crop model for wheat. *Agronomie*, 23(4), 297–303. <https://doi.org/10.1051/agro:2003003>
- Raes, D., Geerts, S., Kipkorir, E., Wellens, J., & Sahli, A. (2006). Simulation of yield decline as a result of water stress with a robust soil water balance model. *Agricultural Water Management*, 81(3), 335–357. <https://doi.org/10.1016/j.agwat.2005.04.006>
- Raes, D., Steduto, P., Hsiao, T., & Fereres, E. (2022). *Aquacrop version 7.0 – Chapter 3: Calculation procedures* (Tech. Rep.). FAO of the United Nations.
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). Aquacrop—The FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, 101(3), 438–447. <https://doi.org/10.2134/agronj2008.0140s>
- Rains, D., Lievens, H., De Lannoy, G. J., McCabe, M. F., de Jeu, R. A., & Miralles, D. G. (2021). Sentinel-1 backscatter assimilation using support vector regression or the water cloud model at European soil moisture sites. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/lgrs.2021.3073484>
- Reichle, R. H., Kumar, S. V., Mahanama, S. P., Koster, R. D., & Liu, Q. (2010). Assimilation of satellite-derived skin temperature observations into land surface models. *Journal of Hydrometeorology*, 11(5), 1103–1122. <https://doi.org/10.1175/2010JHM1262.1>
- Revill, A., Sus, O., Barrett, B., & Williams, M. (2013). Carbon cycling of European croplands: A framework for the assimilation of optical and microwave earth observation data. *Remote Sensing of Environment*, 137, 84–93. <https://doi.org/10.1016/j.rse.2013.06.002>
- Reynolds, C., Jackson, T., & Rawls, W. (2000). Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, 36(12), 3653–3662. <https://doi.org/10.1029/2000WR900130>
- Rojas-Munoz, O., Calvet, J.-C., Bonan, B., Baghdadi, N., Meurey, C., Napoly, A., et al. (2023). Soil moisture monitoring at kilometer scale: Assimilation of sentinel-1 products in ISBA. *Remote Sensing*, 15(17), 4329. <https://doi.org/10.3390/rs15174329>
- Ryu, D., Crow, W. T., Zhan, X., & Jackson, T. J. (2009). Correcting unintended perturbation biases in hydrologic data assimilation. *Journal of Hydrometeorology*, 10(3), 734–750. <https://doi.org/10.1175/2008JHM1038.1>
- Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., et al. (2022). Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a deep neural network. *Remote Sensing of Environment*, 279, 113116. <https://doi.org/10.1016/j.rse.2022.113116>
- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>
- Steele-Dunne, S. C., Hahn, S., Wagner, W., & Vreugdenhil, M. (2019). Investigating vegetation water dynamics and drought using Metop ASCAT over the North American Grasslands. *Remote Sensing of Environment*, 224, 219–235. <https://doi.org/10.1016/j.rse.2019.01.004>
- Sun, L., Belair, S., Carrera, M. L., Bilodeau, B., & Dabboor, M. (2021). Comparing assimilation of synthetic soil moisture versus c-band backscatter for hyper-resolution land surface modeling. *Water Resources Research*, 57(4), e2020WR028921. <https://doi.org/10.1029/2020WR028921>
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). Gmes sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>
- Van Emmerik, T., Steele-Dunne, S. C., Judge, J., & Van De Giesen, N. (2015). Impact of diurnal variation in vegetation water content on radar backscatter from maize during water stress. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7), 3855–3869. <https://doi.org/10.1109/TGRS.2014.2386142>
- van Leeuwen, P. J. (2015). Representation errors and retrievals in linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1612–1623. <https://doi.org/10.1002/qj.2464>
- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273–316. <https://doi.org/10.1016/j.envsoft.2015.08.013>
- Wiseman, G., McNairn, H., Homayouni, S., & Shang, J. (2014). Radarsat-2 polarimetric SAR response to crop biomass for agricultural production monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11), 4461–4471. <https://doi.org/10.1109/JSTARS.2014.2322311>
- Wolfs, D., Swinnen, E., Van Hoolst, R., & Toté, C. (2022). Copernicus global land operations “vegetation and energy” “CGLOPS-1”—product user manual: Dry matter productivity (DMP)—gross dry matter productivity (GDMP)—collection 300m—version 1.1. CGLOPS-1 consortium, Brussels, Belgium.
- Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., et al. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, 10(3), 955–973. <https://doi.org/10.2136/vzj2010.0139>
- Ziliani, M. G., Altaf, M. U., Aragon, B., Houborg, R., Franz, T. E., Lu, Y., et al. (2022). Early season prediction of within-field crop yield variability by assimilating cubesat data into a crop model. *Agricultural and Forest Meteorology*, 313, 108736. <https://doi.org/10.1016/j.agrformet.2021.108736>