L-Band Microwave-based Satellite Data and Model Simulations over the Dry Chaco to Estimate Soil Moisture, Temperature, Vegetation and Soil Salinity

Frederike Vincent[1*], Michiel Maertens[1], Michel Bechtold[1], Esteban Jobbágy[2], Rolf H. Reichle[3], Veerle Vanacker[4], Jasper A. Vrugt[5], Jean-Pierre Wigneron[6], Gabriëlle J. M. De Lannoy[1*]

Abstract— **The Dry Chaco in South America is a semi-arid ecoregion prone to dryland salinization. In this region, we investigated coarse-scale surface soil moisture (*SM*), soil temperature, soil salinity and vegetation, using L-band microwave brightness temperature (T_B) observations and retrievals from the Soil Moisture Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) satellite missions, Catchment Land Surface Model (CLSM) simulations, and in situ measurements within 26 sampled satellite pixels. Across these 26 sampled pixels, the satellite-based *SM* outperformed CLSM *SM* compared to field data, and forward L-band T_B simulations derived from in situ *SM* and temperature performed better than those derived from CLSM estimates relative to SMOS T_B observations. The surface salinity for the sampled pixels was on average only 4 mg/g and only locally influenced the T_B simulations, when including salinity in the dielectric mixing model of the forward radiative transfer model (RTM) simulations. To explore the potential of retrieving salinity together with other RTM parameters to optimize T_B simulations over the entire Dry Chaco, the RTM was inverted using 10 years of multi-angular SMOS T_B data and constraints of CLSM *SM* and temperature. However, the latter modeled *SM* was not sufficiently accurate and factors such as open surface water were missing in the background constraints, so that the salinity retrievals effectively represented a bulk correction of the dielectric constant, rather than salinity per se. However, the retrieval of vegetation, scattering albedo and surface roughness resulted in realistic values.**

Index Terms— **L-band microwave, soil moisture, vegetation, salinity, soil temperature, land surface model, SMOS, SMAP**

I. Introduction

Soil-vegetation processes and their interaction with the atmosphere determine the characteristic features of terrestrial biomes around the world. Soil moisture and vegetation are at the center of the coupling between water, energy and carbon cycles, they regulate land surface fluxes and are constrained by environmental factors such as soil chemical properties, terrain, land use or human interventions [1, 2]. In dry regions, the soil salinity plays an important role in land surface processes, because it influences vegetation growth, and alters the water, energy and carbon budgets [3,4]. Therefore, an integrated large-scale assessment of soil moisture, vegetation and soil salinity is crucial in drylands or biomes with very dry seasons, often found in tropical and subtropical regions.

Large-scale estimates of soil moisture and vegetation are routinely available from land surface and vegetation model simulations and from satellite observations, or combinations thereof [5]. The L-band microwave Soil Moisture Ocean Salinity (SMOS, [6]) and Soil Moisture Active Passive (SMAP, [7]) missions were launched with the explicit purpose to monitor soil moisture globally, and also provide estimates of vegetation optical depth [8, 9, 10].

In contrast to soil moisture and vegetation, large-scale estimates of salinity are scarce and often characterized by a low temporal resolution. Some global estimates are provided by e.g. the Harmonized World Soil Database (HWSDv1.21, hereafter HWSD) [11]. However, those salinity estimates are typically classified in a few categories and the information is time-invariant, often relying on soil samples of decades ago. Remote sensing offers the potential to address these limitations.

Optical remote sensing satellites have localized areas affected by excess soil salinity using spectral salinity indices [12], vegetation proxies [13], or machine learning techniques using a stack of optical remote sensing data and other global datasets [14]. Thermal remote sensing has also been explored to map salinity [15]. Alternatively, microwave remote sensing might prove useful for salinity detection. Passive L-band microwave observations from the SMOS and SMAP missions are routinely used for sea surface salinity estimation [16], but the potential for soil surface salinity monitoring has not been fully explored yet. Given that at L-band, the soil water dielectric

This research was performed with support from the Belspo project REFORCHA (SR/00/338) and C1 KU Leuven (C14/16/045, C14/21/057). The computer resources and services used in this work were provided by the Flemish Supercomputer Center (VSC), funded by the Research Foundation - Flanders (FWO) and the Flemish Government. *(Corresponding authors: G. De Lannoy and F. Vincent).* R. Reichle was supported by the SMAP mission.

1. KU Leuven, Department of Earth and Environmental Sciences, Heverlee, Belgium (Gabrielle.DeLannoy@kuleuven.be, Frederike.Vincent@gmail.com)
2. Universidad Nacional de San Luis & CONICET, Grupo de Estudios Ambientales, IMASL, Argentina
3. NASA Goddard Space Flight Center, Global Modelling and Assimilation Office, Greenbelt, MD, USA.
4. Université Catholique de Louvain, Earth and Life Institute, Georges Lemaître Centre for Earth and Climate Research, Louvain-la-Neuve, Belgium
5. University of California, Irvine, Department of Civil and Environmental Engineering, Irvine, CA, USA
6. INRAE, UMR1391 ISPA, Centre Bordeaux-Aquitaine, F-33140, Villenave d'Ornon, France

The data collected during the field campaign are available at https://doi.org/10.5281/zenodo.4686545. Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

constant is sensitive to salinity [17], it may be possible to improve brightness temperature (T_B) simulations, improve soil moisture retrievals in saline areas, and perhaps even estimate soil surface salinity by including the effect of salinity in the computation of the dielectric constant for water [18, 19], as part of the microwave radiative transfer model (RTM).

Chaturvedi et al. (1983) [20] suggested that a combination of L- and C-band microwave remote sensing data might allow to differentiate between soil salinity and soil moisture effects on the microwave signal. Jackson and O'Neill (1987) [21] used the equations of [19] in the dielectric mixing model of [22] to conclude that salinity below 5 parts per thousand (PPT, mg/g) is most likely undetectable using L-band and would therefore not affect soil moisture retrievals. In contrast, for areas with higher salinity levels, up to 128 PPT, [23] reported significant overestimations of simulated T_B when salinity is not accounted for, leading to errors in soil moisture retrieval.

In short, L-band microwave remote sensing has a proven capability to routinely estimate large-scale soil moisture and vegetation optical depth and it is also affected by salinity, as shown in laboratory and small-scale studies. However, the T_B sensitivity to salinity at the coarse spatial resolution of spaceborne passive microwave radiometers has not yet been widely studied. The key question of this paper is therefore what we can learn from L-band T_B observations about land surface variables such as soil salinity, soil moisture, temperature and vegetation. To this end, the South-American Dry Chaco, with its naturally saline soils [24] at risk of future salinization, forms a unique large-scale testbed.

The objectives of this paper are to (i) evaluate large-scale estimates of soil moisture and temperature in the Dry Chaco, using data from an extensive field campaign, land surface model simulations, SMOS and SMAP retrievals, (ii) simulate L-band T_B with and without accounting for salinity and compare the simulations to satellite-based observations, and (iii) investigate the possibility of jointly retrieving RTM parameters related to soil roughness, vegetation and salinity over the Dry Chaco, using coarse-scale L-band radiometry and modeled constraints of a.o. soil moisture and temperature. The latter constraints help to set the dominant moisture and temperature contributions of the T_B signal apart, and to focus on the estimation of less dominant variables, incl. salinity. Furthermore, the retrieved parameters could be used to optimize the forward RTM to produce improved T_B simulations for a T_B data assimilation system that inherently uses modeled soil moisture and temperature, such as the SMAP Level 4 Surface and Root-Zone Soil Moisture (L4_SM, [5]).

In Section II, the field data are presented, as well as the ancillary data sources and models used for L-band T_B simulation. Section III describes the methods, including the adjustments to the RTM to account for salinity and how the RTM is used in forward and inverse mode to simulate microwave T_B and retrieve land surface properties, respectively. The results are shown in Section IV, a discussion is provided in Section V, and conclusions and recommendations for future research are summarized in Section VI.

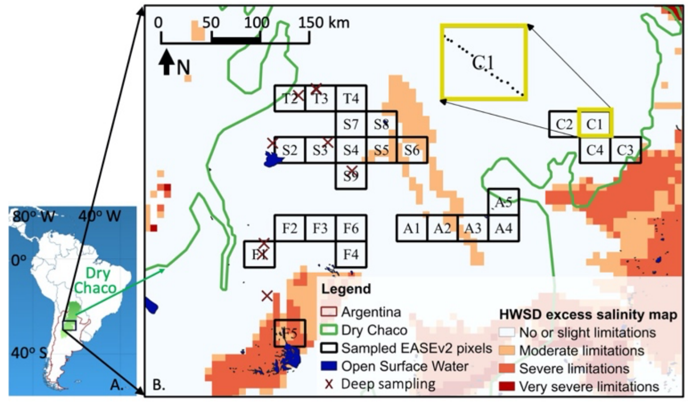

Fig. 1. A. Location of (green polygon) the Dry Chaco in South America, and (blue rectangle) the area of the field campaign. B. Location of (black boxes) EASEv2 pixels with in situ sampling and locations of (x) deep sampling within (green contour) the Dry Chaco region. The background shows the historical excess salinity map from the HWSD and open water areas from the Global Lakes and Wetland Database. Panel B. also provides a (yellow) zoom of one sampling pixel with the corresponding sampling sites.

II. Data and Models

A. Study Area

The South American Gran Chaco hosts a wealth of biodiversity in an area where native dry forests and expanding, mostly rainfed, agriculture for crops and cattle ranching are competing in a changing landscape [24, 25]. Despite its large latitudinal extent and vegetative and climatic variability, the Dry Chaco is a well-delineated ecoregion in the western part of the larger Gran Chaco and covers an area of about 787 000 km^2 in southern Bolivia, western Paraguay and northern Argentina [26], east of the Andean mountain range (Fig. 1A.). The vegetation is dominated by xerophytic shrubs and trees, making up the Earth's largest sub-tropical dry forest, and savannas. East of the Dry Chaco lies the Humid Chaco, characterized by wetlands and a lower tree coverage [25].

The Dry Chaco is one of the planet's largest level plains with slopes < 0.1% and a semi-arid climate. The mean annual temperature ranges between 18 and 26 ˚C but maximum temperatures can easily exceed 40 ˚C [27]. Rainfall varies between 400 and 1000 mm/year and is concentrated in the southern hemisphere summer [26].

Saline soils are common in the region owing to its semi-arid climate, flat topography and shallow groundwater table [24]. Large natural salt deposits occur in the presence of paleo-lakes (e.g. the Salinas de Ambargasta and Salinas Grandes, marked as open water in Fig. 1B.), paleo-channels or floodplains [28]. Large-scale deforestation and forest degradation for agriculture pose a threat for soil salinization in this area. The dry forest conversion began in the early 20th century [29] and reached record high rates at the turn of the century. Multiple regional studies reported an increased downward mobilization of salts towards the water table after deforestation, followed by the rise of the saline water table, leading to the upward movement of salts [30, 25], suggesting the onset of changing hydrological conditions in the Dry Chaco. So far, however, there are no large-scale data to confirm these suggestions.

During the months of July and August 2019, a field campaign was organized in a part of the Argentinean Dry Chaco (Fig. 1B.), discussed in Section II.E. Soil moisture, salinity, temperature and vegetation were sampled near Añatuya, Charata, Frías, Santiago del Estero and Tucuman.

B. SMOS and SMAP Data

L-band T_B (level 1), surface soil moisture (SM) and vegetation optical depth (τ) retrievals (level 2) were collected from the SMOS and SMAP missions. The SMOS mission was launched in 2009 and provides data at a 3-day temporal resolution and a nominal (3 dB) spatial resolution of 43 km [6]. We used the multi-angular SMOS SCLF1C.v620 T_B at horizontal and vertical polarization (Hpol, Vpol), reprojected to the Equal-Area Scalable Earth (EASEv2) 36-km grid, and the SMOS-IC version 2 level 2 SM and τ product [8], resampled to 36 km. The T_B preprocessing and filtering was done as in [31], i.e., excluding pixels where T_B observations are impacted by open water, radiofrequency interference, etc. as provided in the product quality flags. The SMAP satellite was launched in 2015 and collects data at a similar temporal and spatial resolution as SMOS [7]. The SMAP level 1 Hpol and Vpol T_B data were extracted from the SPL1CTB.v004 product, and level 2 SM retrievals were extracted from SPL2SMP.v006, which are both produced on the EASEv2 36 km-grid. The τ retrievals were extracted from the 9-km SPL2SMP_E.004, using the 'option 3' dual-channel algorithm [9], conservatively filtered using the provided quality flags and aggregated to 36 km.

The SMOS and SMAP SM and T_B data for July-August 2019 were compared to in situ and model SM, and to evaluate forward RTM simulations, respectively, at pixels visited during the field campaign. The multi-angular SMOS T_B for 2010-2019 were used for RTM inversions, whereas the time-average <τ> retrieval products for 2015-2019 (both SMOS and SMAP) were used to evaluate the <τ> estimates obtained from the RTM inversion. The SMOS and SMAP τ retrievals were earlier evaluated with a range of independent vegetation datasets [32].

C. Land Surface Modeling

The Catchment Land Surface Model (CLSM, [33]), part of the NASA Goddard Earth Observing System (GEOS) model, was forced with meteorological data from the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2, [34]) to simulate land surface processes over the Dry Chaco on the 36-km EASEv2 grid from 2010 through 2019, after 30 years of spin-up. The CLSM version is similar to that used for the SMAP L4_SM product version 4 [5]. The system simulates soil moisture in the surface (0-5 cm) and root zone (0-100 cm), surface temperature in a layer of 0-5 cm, and soil temperature in 6 layers with the first layer pertaining to 5-15 cm depth. For surface soil moisture, the simulated and in situ sampled layer depth is similar (0-5 cm), and both will be referred to as *SM*. For temperature, the CLSM surface temperature (*Tsurf*, 0-5 cm) and soil temperature in the first layer (*Tsoil1*, 5-15 cm) will be compared to in situ temperature data collected in the upper 5 cm (*T5*). Because the CLSM *Tsoil1* is used as temperature input to the forward RTM in the SMAP L4_SM product, a statistical relationship between *Tsoil1* simulations and in situ *T5* observations (frequently sampled during the day) was established to bias-correct the model *Tsoil1* towards in situ *T5*, and also to extrapolate point-based in situ *T5* measurements to upscaled 36-km EASEv2 estimates at specific times of the day (Section IV.A). These bias-corrected CLSM soil temperature estimates will therefore also be referred to as *T5*. The *SM* and *Tsoil1* (or *Tsurf*) simulations were compared to in situ measurements and satellite retrievals, and then used for forward T_B simulation, or as a constraint to invert SMOS T_B and derive land surface properties (RTM parameters, incl. salinity and τ).

It is important to note some shortcomings of the CLSM. First, explicit deforestation or the vegetation response to salinity is not included in the CLSM, and a historical climatology is used to describe the vegetation instead. The CLSM leaf area index (LAI) values over the Dry Chaco are based on a multi-year average climatology, obtained from GEOLAND2 [35]. This poses no problem for long-term deforested areas, but for recently deforested areas, the vegetation and *SM* may locally deviate from historical climatological conditions [36]. Second, it can be expected that soil salinity alters the water retention curve in a soil and affects the soil moisture profile and evapotranspiration response [37]. That would affect the satellite-observed T_B signal but is not captured in the CLSM or any other state-of-the-art land surface model.

D. L-Band RTM

Given static and dynamic information about the land surface, a zero-order omega-tau RTM was used to simulate L-band T_B at the top of the vegetation at horizontal or vertical polarization p (Hpol or Vpol) as follows:

$$T_{B,p} = T(1 - r_p)A_p + T_C(1 - \omega_p)(1 - A_p)(1 + r_p A_p) \quad (1)$$

Atmospheric contributions were not included, because they were already removed from the L-band satellite data, T [K] is the effective soil temperature (either *T5* or *Tsoil1*), T_C [K] is the effective canopy temperature (assumed to equal T), r_p [-] is the rough surface reflectivity, ω_p [-] is the scattering albedo and A_p [-] is the vegetation attenuation, calculated as follows:

$$A_p = exp\left(\frac{-\tau_p}{cos(\theta)}\right) \quad (2)$$

The vegetation optical depth τ_p [-] is depends on a vegetation-structure parameter, b_p [-], and the vegetation water content (*VWC*), which is the product of *LAI* [m^2/m^2] and leaf equivalent water thickness (*LEWT*) [kg/m^2]:

$$\tau_p = b_p VWC, \text{ and } VWC = LEWT\ LAI \quad (3)$$

The rough surface reflectivity, r_p [-] is defined following [38]:

$$r_p = \left[QR_q + (1 - Q)R_p\right] exp(-h)\, cos^{N_p}(\theta) \quad (4)$$

and requires Q, the polarization mixing ratio [-] (assumed to equal 0 for L-band), θ, the incidence angle [rad], h, a *SM*-dependent roughness parameter [-], and N_p, the angular dependence [-]. The roughness h varies in time between h_{max} for *SM* at or below transition SM and h_{min} for *SM* at saturation. R_p is the smooth surface reflectivity [-], which is calculated using the Fresnel equations:

$$R_H(\theta) = \left|\frac{cos(\theta)-\sqrt{\varepsilon-sin^2(\theta)}}{cos(\theta)+\sqrt{\varepsilon-sin^2(\theta)}}\right|^2, \text{ and}$$

$$R_V(\theta) = \left|\frac{\varepsilon cos(\theta)-\sqrt{\varepsilon-sin^2(\theta)}}{\varepsilon\, cos(\theta)+\sqrt{\varepsilon-sin^2(\theta)}}\right|^2 \quad (5)$$

with ε the complex dielectric constant of the soil, $\varepsilon = \varepsilon' + i\varepsilon''$, where ε' is the real part of the dielectric constant, ε'' is the imaginary part and i is the solution of the equation $i^2 = -1$. As will be discussed below, ε is a function of soil texture, *SM*, *T* and *S*.

This paper used the same RTM structure as that used in Version 4 of the SMAP L4_SM product [5]. In short, the RTM uses dynamic CLSM *SM* and *T* (*Tsoil1* or *T5*) as input, along with seasonally varying climatological LAI, and a set of lookup or calibrated RTM parameters (see below). The lookup parameters were based on the dominant vegetation cover, determined by the MODIS IGBP landcover map. The RTM simulations were also performed for a small sample of in situ *SM* and *T5* observations described below.

E. In Situ Data

During the months of July and August 2019, a total of 26 pixels on the EASEv2 36 km grid (Fig. 1) were sampled to facilitate comparison with satellite and model data. The pixels were chosen based on satellite imagery, deforestation history, expected salinity level (historical mapped data, unmapped natural or upcoming salinity as suggested by literature or satellite imagery), elevation and accessibility. A multitude of sites were sampled within each pixel, each pixel was sampled entirely within a day, and each pixel was visited only once. In the Dry Chaco, the months of July and August are typically dry, and in 2019, there was a clear dry-down after an anomalously wet austral summer. Table I provides an overview of all sampled pixels, the number of sample sites within the pixel and the dominant vegetation type.

We collected extensive surface soil data in areas with diversified levels of salinization, and deeper soil samples at locations where deforestation had taken place either recently or in the past. Surface (0-5 cm) soil data were gathered on soil moisture (*SM*), temperature (*T5*), electrical conductivity (EC), and dielectric constant (ε) at multiple sample sites along a transect within each sampled EASEv2 pixel. The transect was situated to best capture the variability of the pixel, and consisted of minimally three (F6) and maximally 27 (A1) sample sites. An example of a transect is depicted in Fig. 1B. At ten locations, marked with crosses in Fig. 1B., additional deeper soil measurements were taken. Those are discussed in Appendix 1, to frame our findings within the scientific discussion about the salinization potential of the Dry Chaco.

For *SM*, we used two different probes (Stevens HydraGO and the Delta-T Devices ThetaProbe). The HydraGO was used with the factory calibration that is also used for the widely used Soil Climate Analysis Network of the Natural Resource Conservation Service [39]. At each sample site, measurements were taken in two or three different pits (1 m apart) by horizontally inserting the probes into a cleared surface of an undisturbed soil at a depth of approximately 2.5 cm. The probes have a sensing volume that covers 5 cm in the vertical direction, i.e. the measurements represent the top 5 cm of soil. A high Pearson correlation (0.87) and a small bias (0.028 m^3/m^3) were found between all individual *SM* measurements from both probes (931 measurements with each sensor). Per site, an average *SM* was calculated based on the measurements of the HydraGO and the ThetaProbe in all pits. Given the strong agreement between both *SM* sensors and given that the composition of soils in the Dry Chaco is not too different from those used in the factory calibration of the HydraGO sensor, we believe that the averaged data are reliable to serve as reference in terms of relative accuracy metrics. Furthermore, 38 texture samples were collected.

To measure in situ dielectric properties and salinity in the field, several methods were used. The Stevens HydraGO probe was used to measure porewater EC and the real and imaginary parts of ε (ε' and ε'') at 0.05 GHz (i.e., a much lower frequency than the L-band frequency of interest to this paper). Two other probes, a Hanna sensor and the YSI proDSS, were used to measure the EC of a soil-water paste (hereafter referred to as EC) of a mixed soil sample that was taken from the top 5 cm at each sample location along the transect. With the soil sample, a soil-water mixture was prepared on a 1:1 ratio; adding 50 ml of water to 50 g of the sample. The 1:1 ratio was chosen to minimize probe sensitivity difficulties because of dilution (which occurs more at the typical 1:5 ratio) and to allow reproducibility. After mixing and resting, the EC was measured with both probes. The measurements by both probes correlated well (R=0.94), with a slight underestimation by the Hanna probe (bias = -0.82 dS/m), due to saturation at 20 dS/m. In the following, only the YSI proDSS EC measurements were used. Where needed, these in situ EC measurements in dS/m were converted to salinity (*S*) in parts per thousand (PPT) in the Practical Salinity Scale, following the YSI proDSS's internal conversion method, i.e. via regression equations (PSS 78, [40]) that only consider the EC and the temperature of the soil-water mixture. The equations are based on the salinity of seawater, and thus not fully representative for the ionic content of the Dry Chaco soil water, but the conversion error is expected to be small and consistent across all samples.

In situ surface soil temperature was measured with the HydraGO within the top 5 cm of the soil. Per sample site, *T5* measurements in all pits were taken within less than 15 minutes and averaged to one value. All measurements were taken during the daytime, and a special effort was done to sample additional *T5* data close to the SMOS and SMAP satellite overpass times.

TABLE 1
OVERVIEW OF SAMPLED EASEv2 PIXELS IN JULY AND AUGUST 2019: NAME, NUMBER OF SAMPLE SITES PER PIXEL (#), SAMPLE DATE (DAY/MONTH) IN THE YEAR 2019 AND DESCRIPTION OF THE LAND COVER. NAMES ARE BASED ON THE MUNICIPALITY CLOSEST TO THE PIXEL LOCATION: A = AÑATUYA, C= CHARATA, F=FRÍAS, S=SANTIAGO DEL ESTERO AND T=TUCUMAN.

name	#	date	description	name	#	date	description	name	#	date	description
A1	27	13/08	Forest	**F1**	20	22/08	Forest and pastures	**S5**	16	25/07	Mainly forest, cotton fields
A2	19	12/08	Mainly forest, some bushes	**F2**	19	24/08	Forest and pastures	**S6**	13	26/07	Mainly forest
A3	25	14/08	Forest and bushes	**F3**	19	23/08	Forest	**S7**	16	29/07	Forest, bushes (halophytic)
A4	22	16/08	Agriculture	**F4**	20	21/08	Mainly bushes	**S8**	20	31/07	Mainly forest
A5	6	17/08	Bushes, close to salt lakes	**F5**	5	18/08	Forest and bushes near saline lakes	**S9**	21	01/08	Agriculture
C1	20	06/08	Post-harvest agriculture	**F6**	3	23/08	Bushes	**T2**	20	26/08	Agriculture
C2	20	07/08	Post-harvest agriculture	**S2**	11	19/07	Agriculture and forest	**T3**	20	27/08	Agriculture, some bushes
C3	21	08/08	Post-harvest agriculture	**S3**	19	23/07	Mainly forest and bushes	**T4**	16	30/07	Mainly dense forest
C4	19	09/08	Post-harvest agriculture	**S4**	19	24/07	Mainly forest				

Vegetation type was visually assessed at each location along the transect, and sampling was done in all different vegetation types that were representative for that pixel, and in zones with different elevations and expected salinity, i.e., the sample sites were chosen after close inspection of satellite imagery and ancillary spatial datasets (see below).

F. Pixel-Scale In Situ Data

To allow for comparison with satellite observations and model simulations on the 36-km EASEv2 grid, point measurements of surface *SM*, *S* (measured as EC in the field) and ε were upscaled via a weighted average to obtain in situ-based 'observations' that are representative of a pixel area. The values of the limited number (mostly > 15) of samples for the one pixel were weighted based on elevation from the 30 arc-second Multi Error Improved Terrain (MERIT, [41]) digital elevation model. This choice was based on the notion that both *SM* and *S* vary spatially with depth to the water table and the proximity of streams, which in turn are linked to elevation.

The upscaling of in situ soil temperature (*T5*) measurements leveraged CLSM simulations (Section II.C). In short, the simulated diurnal cycle of CLSM *Tsoil1* at the 36-km pixel scale was compared to *T5* site measurements taken at many different times in the day, to establish a mapping function between both. This allowed to bias-correct the 36-km *Tsoil1* simulations to in situ *T5* observations, and to obtain a pixel-scale "in situ-based" *T5* estimate at any time of the day.

The visual inspection of vegetation (Table 1) within each sampled pixel was aggregated to a dominant impression of the landscape. For each visited EASEv2 pixel, a dominant vegetation class was selected by 3 observers during the field campaign from the 17 possibilities of the International Geosphere-Biosphere Program (IGBP) land cover classification [42].

Fig. 2 summarizes all surface soil data gathered during the field campaign. In panels A.-C., the within- and between-pixel variability in *SM*, *S* (EC in dS/m), and *T5* are shown. The variability also includes temporal variability, which is limited within a pixel for *SM*, because each EASEv2 pixel was always completely sampled within a day. In contrast, the variability in *T5* samples is largely driven by diurnal and daily temperature fluctuations. The blue dots indicate the upscaled values for *SM* and *S* (EC), whereas the red dots are *T5* estimates at 6 am local time on the sample day (extrapolated to 6 am, discussed below). Except for *T5*, the upscaled values for all other variables are close to the simple pixel median values (center line of boxplot). For *T5*, the median value is associated with the sampled subset of the diurnal temperature variation, whereas the upscaled value is at a fixed time. The figure indicates that there is a large variability between pixels. For certain pixels and variables, there is also a large within-pixel variability. Our study only focuses on the variability between – and not within – pixels.

G. Ancillary Data for In Situ Pixels

Additional data sources complement the in situ data collection. For every sampled 36-km EASEv2 pixel, an LAI value was calculated from 4-day 500-m MODerate resolution Imaging Spectroradiometer (MODIS; onboard the NASA Terra and Aqua satellites) retrievals (version 006) [43] for July and August 2019. Fig. 2D. shows the pixel-average LAI and the in situ observed vegetation class.

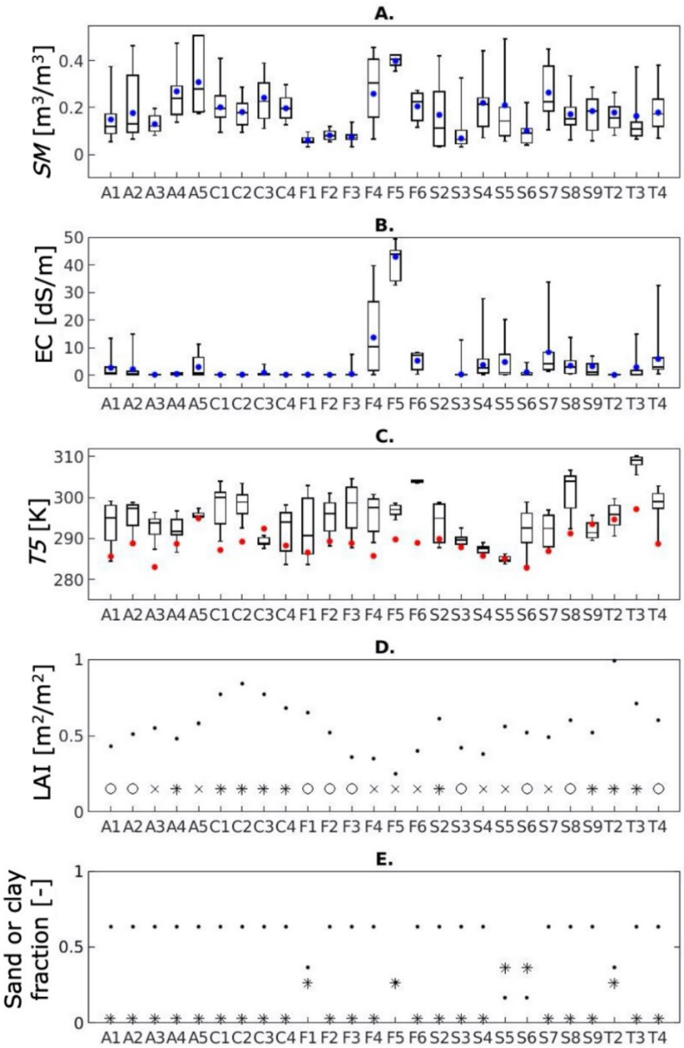

Fig. 2. Boxplots showing the within-pixel variability of in situ measurements for the visited EASEv2 pixels, for A. soil moisture (*SM*), B. salinity (EC) and C. soil temperature (*T5*). The blue and red dots indicate the upscaled EASEv2 pixel value, with the red dots representing soil temperature extrapolated to 6 am. Also shown are D. July-August 2019 average of MODIS-derived LAI (**.**) and in situ observed vegetation class, with o = mixed forest, * = agriculture and x = open shrubland, and E. HWSD-derived sand (**.**) and clay (*) fraction. Sampling dates differ across pixels and range from July 19 to August 27, 2019 (Table I).

Soil texture information was derived from the HWSD, and soil hydraulic parameters were derived as in [44]. Fig. 2E. gives the HWSD sand and clay fraction per EASEv2 pixel. A comparison with 38 soil samples (laboratory analysis, not shown) within 20 of the 26 sampled EASEv2 pixels showed high sand and low clay fractions in both datasets. Given the small set of in situ texture samples, we relied solely on HWSD estimates in the remainder of this paper.

H. Regional Reference Data

We further consulted the HWSD excess salts map (Fig. 1B.) as a reference for salinity over the entire Dry Chaco. This map indicates the historically observed and expected severity of growth limitations due to salinity, sodicity or both, and does not account for possible recent salinization. Fig. 1B. illustrates that the occurrence of excess salinity is higher in proximity of (paleo-)lakes and streams. The static open water areas in the figure originate from the Global Wetland and Lake Database [45]. The 25-km dynamic open water estimates retrieved from the Advanced Microwave Scanning Radiometer 2 (AMSR2) [46] were also consulted to evaluate the results.

As a reference dataset for vegetation, 1-ha above ground biomass (AGB) estimates for the year 2010 [47] were used. The 1-ha AGB data were obtained from a combination of radar, optical and lidar data and were aggregated to the EASEv2 36-km resolution. Since AGB is the oven-dry weight of the woody parts (stem, bark, branches and twigs) of all living trees excluding stump and roots, it represents a very different quantity than the L-band τ, which is a microwave-based index that strongly varies with vegetation water content. However relative spatial patterns in AGB estimates can be reliably related to L-band τ patterns [48].

III. Methodology

A. Accounting for Salinity in the RTM

In the RTM, the soil dielectric constant ε is calculated using a dielectric mixing model. Soil, water and air components all contribute to the dielectric properties of the soil mixture. Salinity also has an influence on the dielectric properties of a soil, and reduces T_B, but is generally not included in global land RTMs. In line with the SMAP L4_SM product, we used the empirical [22] dielectric mixing model to calculate ε based on the dielectric constants of air (ε_a), rock (ε_r), ice (ε_i), and water (ε_w) with a distinction between bound and free water (see Appendix 2 for detailed equations, similar results were found with the [49] model). The dielectric constant of free water (ε_w) can be calculated by the Debye expression:

$$\varepsilon_w = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1+(i\omega t)^{1-\alpha}} - i\frac{\sigma}{\omega\varepsilon_0}, \quad (6)$$

where ε_∞ is the dielectric constant at an infinite frequency (set to $\varepsilon_\infty = 4.9$), $\omega = 2\pi f$ with f the frequency [Hz], ε_0 is the permittivity of free space (equal to $8.854 \cdot 10^{-12}$) and α is an empirical parameter describing the distribution of relaxation times, and is set to zero [19].

In most operational RTMs, ε_w is calculated for pure water with the ionic conductivity σ equal to zero. The remaining variables in Eq. 6, i.e., the ionic conductivity σ [mhos/m], the relaxation time t [s] and the static dielectric constant ε_s [-], are calculated based on regressions equations using only T as an explanatory variable [49].

In this study, we replaced these equations for $\sigma(T)$, $t(T)$ and $\varepsilon_s(T)$ in pure water with $\sigma(S,T)$, $t(S,T)$ and $\varepsilon_s(S,T)$, i.e. with the regression equations of [19] for saline water in the Debye expression (Eq. 6) to calculate ε_w for the dielectric mixing model of [22]. These regression equations of [19] build further on the equations of [18], which are valid for salinity ranging from 4 to 35 PPT. However, the authors note that the lower limit is not restrictive, and that only for distilled water a different set of equations should be used. For the detailed regression equations, we refer to [19].

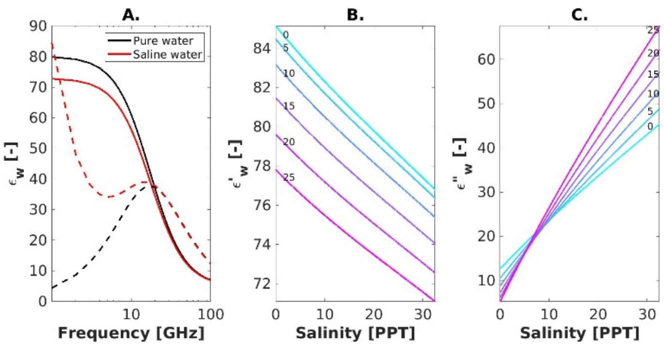

Fig. 3. A. Real (solid lines) and imaginary (dashed lines) dielectric constant of pure (0 PPT) and saline (32.5 PPT) water based on [18] for frequencies between 0 and 100 GHz and for a temperature of 20 ºC. Effect of salinity on the (B.) real and (C.) imaginary part of the dielectric constant of water at L-band (1.4 GHz) for a realistic range of soil water temperatures (˚C).

Fig. 3A. illustrates the $\varepsilon_w{}'$ and the $\varepsilon_w{}''$ of pure (0 PPT) and saline (32.5 PPT) water at frequencies ranging from 0 to 100 GHz following [19]. Increasing salinity leads to a decrease in the real part and an increase in the imaginary part of the dielectric constant of saline water, in line with [17]. Fig. 3B. and C. illustrate the influence of soil water temperature on the sensitivity of ε'_w and ε''_w to salinity at L-band (1.4 GHz). Fig. 3C. indicates that ε''_w, becomes more sensitive to salinity at higher soil water temperatures.

Note that adding salinity via ε_w in the free water solution is a simplification of reality, as salts can be found as ions in the bound water fraction or in the form of salt crystals. We assume that the salinity measured in the field equals the soil water salinity that influences ε_w, and consequently the dielectric constant of the soil. In sensitivity tests (not shown), we found that the saline-water Double Debye dielectric model [50] created by William Ellison [51] produced very similar results to [19].

B. Forward T_B Simulations

To understand the forward (and inverse) simulations with the adjusted RTM, a global sensitivity analysis of the modelled forward T_B to various RTM input parameters was performed, considering first order interactions between the various RTM input parameters. Sobol's sensitivity indices [52] were computed to quantify how much an individual input parameter contributes to the variance in T_B (either at H- or Vpol), using Monte Carlo simulation with 10^5 samples. The framework of [53] was used to calculate Sobol indices for eight input parameters (*SM*, *T5*, *S*, porosity (*P*), wilting point (*wp*), VWC, h_{min} and ω), with the margins based on the field campaign or literature ('Lit2', [56]). The global sensitivity analysis is followed by a local sensitivity analysis focusing only on the influence of *S* on T_B.

Next, forward T_B simulations were performed using either upscaled in situ data or CLSM input of *SM*, *T* and vegetation for the sampled EASEv2 pixels. For both in situ-based and model-based T_B simulations, the RTM ran twice: once with and once without salinity as an input variable. Hpol and Vpol T_B simulations at 40˚ incidence angle were evaluated with SMAP and SMOS T_B observations at the time of the satellite overpass. The difference between both sensors is small (< 3 K difference between SMAP and SMOS T_B over land, [54, 55]) and therefore SMOS and SMAP T_B were used together to ensure sufficient observations for the evaluation of T_B simulations at the time of field sampling.

The input of upscaled in situ and CLSM *SM* and *T* (*T5* and *Tsoil1*) was as described in Section II.F and II.C, respectively. When T_B was simulated with in situ-based input, the LAI was obtained using MODIS imagery at the sampling time, whereas the T_B simulation with CLSM input used CLSM LAI values based on the multi-year average seasonally varying GEOLAND2 climatology (Section II.C). For in situ-based simulations, the vegetation-related lookup RTM parameters were based on the vegetation class observed in the field, whereas for the model-based simulations, lookup RTM parameters were based on the MODIS IGBP vegetation class used in the CLSM RTM. For both in situ- and model-based simulations, the soil-related RTM parameters were based on the HWSD (Section II.G), and T_B simulations that account for salinity use the in situ salinity measurements. The lookup table values for the parameters b_p, *LEWT*, ω_p, h, and N_{rp} per vegetation class were compiled from literature and are referred to as 'Lit2' in [56].

C. Retrieval of Land Surface Properties

To retrieve land surface properties over the entire Dry Chaco, including RTM parameters related to salinity, vegetation optical depth τ, microwave roughness h, and scattering albedo ω, the RTM was used in inverse mode for each pixel individually. To estimate multiple unknowns, multiple types of independent observations and constraints are needed. An important complexity is that the presence of water and salinity both affect the ε and subsequent T_B simulations. Therefore, the estimation of salinity requires a priori knowledge or strong constraints for the other (more dominant) variables that influence T_B.

The multi-angular (7 incidence angles [30º, 35º, …, 60º]) and dual-polarized SMOS T_B of the previous decade (July 2010 - November 2019) were used as observational constraint, and CLSM-based *SM*, LAI and *T* input data (either with or without CLSM *Tsoil1* bias correction, i.e. *T5* or *Tsoil1*) were used as modeled background constraint (i.e. assumed to be known) to (i) find parameters that are consistent with the modeling system and suitable for a subsequent forward RTM application, and (ii) to exclude the dominant influence of *SM* and *T* from the T_B signal to retrieve less dominant factors. A set of parameters (α [-]) was calibrated (or 'retrieved') to minimize the multi-angular and multi-polarization misfit between long-term mean values and standard deviations of SMOS observed T_B (T_{Bo}) and forecasted T_B ($T_{Bf}(\alpha)$), following the procedure of [57, 58], i.e., using a Bayesian optimization with inclusion of priori parameter constraint. The $T_{Bf}(\alpha)$ were forecasted using the RTM (Eq. 1) with CLSM-based input data.

The calibrated parameters included $h_{\min}$ and Δh=$h_{\max} - h_{\min}$ (Eq. 4) to parameterize a *SM*-dependent h, the scattering albedo ω (Eq. 1) (here polarization-independent), b_{h} and Δb=$b_{\mathrm{v}} - b_{\mathrm{h}}$ (Eq. 3) to parameterize the LAI-dependent τ_{p} (i.e., 5 parameters), either with or without two additional parameters that mimic the presence of salinity: s_a and s_b. In line with field experiments [21], the s_a and s_b parameters describe that over time, when *SM* (m^3/m^3) decreases, *S* (PPT) increases at a single pixel:

$$S = s_a + s_b\, SM \tag{7}$$

The initial values and ranges were set to s_a = 5 [0 35] PPT and s_b = -10 [-88 0] PPT/(m^3/m^3) based on trial-and-error [59]. The proposed equation does not consider temporal variation of salt mass in the soil caused, for example, by ongoing evaporation. A physically more complex alternative salt model formulation could not be justified, for lack of large-scale salinity observations to support it (and it will be shown later that the calibrated *S* does not actually reflect salinity but a bulk correction to the dielectric constant). After calibration of s_a and s_b, Eq. 7 can be used to calculate *S* per pixel at any time, given its *SM* value. This allows for an evaluation of *S* for specific days of the field campaign, but our main evaluation will focus on spatial patterns in long-term 10-year average <*S*> estimates. In line with $S = f(s_a, s_b, SM)$, it should be noted that $h = f(h_{min}, \Delta h, SM)$ depends on time-varying *SM*, and $\tau_{\mathrm{p}} = f(b_{\mathrm{p}}, LAI)$ depends on time-varying LAI, and we will evaluate the 10-year average <*h*> and <*τ*>, where the latter is also an average of τ_{p} at Hpol and Vpol.

Four calibration cases were considered. In a reference case, the same 5 parameters (α) were calibrated as in [57, 58] without inclusion of salinity or any correction to CLSM input variables, similar to what is used for the RTM calibration of the SMAP L4_SM product [5]. The three other cases either (i) include an in situ-based bias-correction to CLSM *Tsoil1* input, or (ii) include the calibration of 2 additional parameters related to salinity (equivalents), or (iii) include both (i) and (ii). The optimization was performed with a Markov Chain Monte Carlo algorithm, i.e., DiffeRential Evolution Adaptive Metropolis with parallel direction and snooker sampling from past states (DREAM$_{(ZS)}$, [60]), with the same settings as in [58]. A major advantage of this method is that it provides the entire posteriori density distribution of the parameter estimates, i.e. with access to the maximum a posteriori density (or 'best') and ensemble mean parameter estimates, and the associated ensemble standard deviation. The latter quantifies the uncertainty of the retrievals.

The multi-temporal retrieval approach with strong background constraints of modeled *T* and *SM*, and the imposed inverse temporal relationship between *SM* and *S* limit the possibilities for equifinality: if *S* and *SM* would be retrieved simultaneously at individual time steps without strong background constraints, multiple combinations of *SM* and *S* would be found to be equally good. However, imposing modeled background *SM* also holds the risk that the *S* retrievals compensate for biases in the background data, as will be discussed below. Keeping the multi-temporal approach and including a long-term *T* and *SM* bias estimation in the retrieval is feasible and recommended for future research.

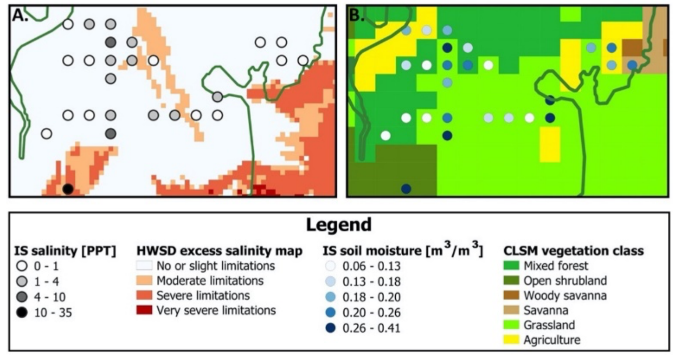

Fig. 4. In situ (IS) upscaled values of A. salinity and B. soil moisture. The circles represent the centroid of the EASEv2 pixels sampled at various dates during the field work. In the background, A. the HWSD excess salinity map and B. the model-based MODIS IGBP vegetation classification and the boarders of the Dry Chaco are shown (green contour).

IV. Results

A. Data Analysis: Satellite Pixels with In Situ Sampling

We start with an inspection of the surface *S*, *T* (*Tsurf, Tsoil1* or *T5*), *SM* and vegetation in the sampled EASEv2 pixels, observed in situ or with satellite data, and simulated with CLSM. Given that the temperature shows a strong diurnal cycle, whereas the other variables are nearly constant within a day in this region during the dry season, we analyzed *Tsurf, Tsoil1* and *T5* at individual sample times within a day, whereas in situ SM and salinity samples were first aggregated to an upscaled daily value prior to comparison with model and satellite data as discussed in Section II.F.

For *S*, Fig. 1 and Fig. 4A. show that two pixels with the highest in situ values (F4, F5) are indeed located in or near areas identified as being limited due to salinity and/or sodicity in the HWSD. However, there is only a limited agreement between the in situ data and the general pattern of the HWSD, indicating that the HWSD estimates likely do not reflect surface salinity only or may not be representative of the current situation. Fig. 4B. shows the associated *SM* at the times and locations of sampling. Pixels with a high in situ *S* (Fig. 4A.) are also characterized by a high in situ *SM* (Fig. 4B.).

For *SM*, Table II gives an overview of spatio-temporal accuracy metrics, for various *SM* products for the 26 sampled EASEv2 pixels, where in situ *SM* is upscaled and each 36-km EASEv2 pixel is sampled only once. CLSM *SM* does not correlate well with in situ data, SMOS or SMAP retrievals. The in situ measurements, on the other hand, have a high correlation (R~0.7) with SMOS and SMAP *SM* retrievals. The RMSD values are similar for model or in situ *SM* compared to satellite observations. Fig. 5A. further illustrates that the bias between CLSM and SMAP *SM* is only 0.02 m^3/m^3 when computed across the years 2015-2019 and all 26 sampled EASEv2 pixels.

For some individual pixels, however, large biases between in situ *SM*, retrievals and model simulations exist, with a noteworthy deviation (significantly higher bias of 0.09 m^3/m^3) between CLSM and SMAP estimates for saline pixels (> 4 PPT, indicated in blue).

TABLE II
COMPARISON OF DIFFERENT SM DATA PRODUCTS AT 26 IN SITU PIXELS SAMPLED IN JULY AND AUGUST 2019. THE BIAS IS RELATIVE TO THE REFERENCE PRODUCT IN THE CORRESPONDING MAJOR ROW HEADER. CORRELATIONS IN BOLDFACE ARE SIGNIFICANT AT A LEVEL OF α=0.05.

	Model			**In situ**		
	R [-]	**RMSD [m^3/m^3]**	**Bias [m^3/m^3]**	**R [-]**	**RMSD [m^3/m^3]**	**Bias [m^3/m^3]**
In situ	0.09	0.11	0.06	-	-	-
SMOS	0.32	0.08	-0.01	**0.71**	0.08	0.05
SMAP	0.14	0.06	-0.03	**0.73**	0.07	0.03
Model	-	-	-	0.09	0.11	0.06

Fig. 5B. shows all individual point in situ *T5* observations at various times in the day within all EASEv2 pixels against the model equivalents of *Tsurf* and the deeper *Tsoil1* closest in time and space (closest hour, overlying 36-km pixel). There is a high correlation (R=0.84) between in situ observed *T5* and simulated *Tsurf*, and the model *Tsurf* is only slightly colder (less than 2 K) than in situ *T5*. Compared to the modeled *Tsoil1*, the correlation is still high (R=0.83) and the model *Tsoil1* is lower than in situ *T5* (bias = -5.53 K), due to the spatial (horizontal and vertical) representativeness mismatch. The model *Tsoil1* is associated with a deeper 5-15 cm depth for an entire 36-km EASEv2 pixel, whereas the in situ *T5* is only representative of the top 5 cm at a point location. The diurnal amplitude of *Tsoil1* is thus expected to be smaller than the in situ *T5*, and *T5* will be warmer during the daytime. Because *Tsoil1* is used as input to the RTM in the SMAP L4_SM product, *Tsoil1* is used as the reference temperature in the following, and a relationship was established between model *Tsoil1* and in situ measured *T5* at all in situ measurement times. A stratification by vegetation class over the area of the field campaign did not further refine the relationship between in situ *T5* and model *Tsoil1* and therefore, all data were used to derive the following linear regression between in situ *T5* [K] and model *Tsoil1* [K] (based on daytime samples with 281.22 K < *Tsoil1* < 300.17 K):

$$T5 = 1.20\ Tsoil1 - 52.44 \qquad (8)$$

This Eq. 8 was thus used to rescale the model *Tsoil1* at SMOS or SMAP satellite overpass times (at ~ 6 am or 6 pm local time) to obtain either extrapolated “in situ-based” *T5* estimates for each visited EASEv2 pixel, or “bias-corrected” model *Tsoil1* estimates in forward or inverse RTM simulations, as introduced in Section II.F.

Finally, the observed vegetation at the sampled pixels is either dominantly agriculture, open shrubland or mixed forest, as summarized in Fig. 2D. Fig. 4 illustrates that the sampled area is in the transition zone between agricultural, shrubland and broadleaf dry forest. It is thus not surprising that only for 6 out of the 26 sampled pixels, the assignment of the model-based vegetation classes agrees with the field-based classification. Furthermore (not shown), the MODIS observed LAI is typically higher for the agricultural areas than for the dry forest during the field campaign.

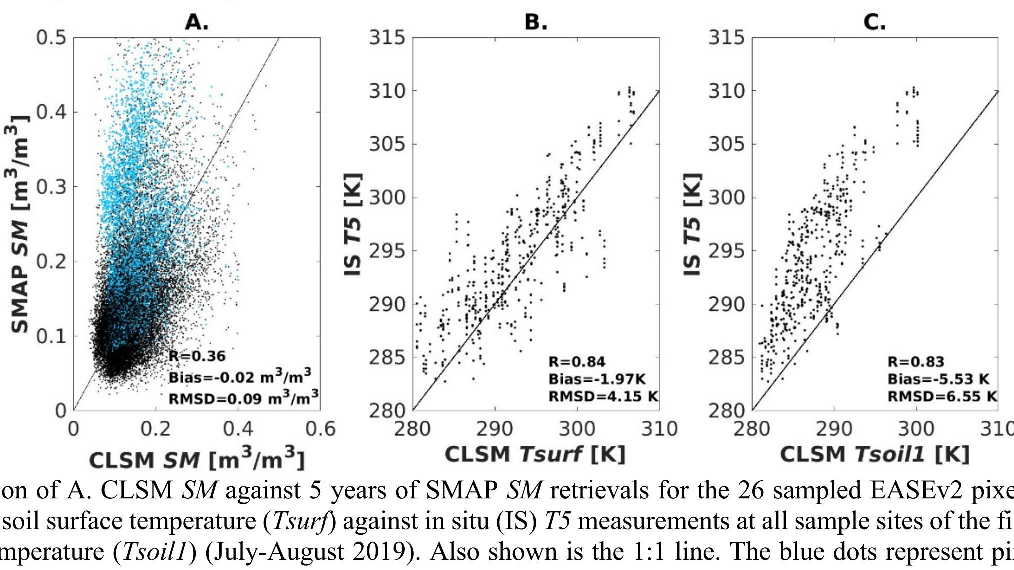

Fig. 5. Comparison of A. CLSM *SM* against 5 years of SMAP *SM* retrievals for the 26 sampled EASEv2 pixels (August 2015-2019), B. CLSM soil surface temperature (*Tsurf*) against in situ (IS) *T5* measurements at all sample sites of the field campaign and C. CLSM soil temperature (*Tsoil1*) (July-August 2019). Also shown is the 1:1 line. The blue dots represent pixels (3 of the 26) with a *S*>= 4 PPT.

B. Data Analysis: Dry Chaco Region

Satellite observations of T_B for the entire Chaco area give an integrated view of the land surface. Fig. 6A. and B. show 10-year averages of SMOS Hpol and Vpol T_B at 40° incidence angle, and the corresponding (cross-masked) 10-year averages of CLSM simulations of *SM* (Fig. 6C.) and *Tsoil1* (Fig. 6D.). The model *Tsoil1* simulations show a smooth gradient over both the Dry Chaco and the area east of it (Humid Chaco) with higher temperatures in the northern Chaco and little differences between east and west. The 10-year average *SM* simulation pattern is dominated by texture (areas with higher porosities have higher *SM* values) and climatological rainfall patterns. For example, the area east of the Dry Chaco is wetter both because of higher precipitation amounts and higher soil porosities. The long-term mean SMOS T_B (both H- and Vpol) observations combine real *SM* and temperature, along with vegetation information, resulting in a smoother spatial pattern than the modeled *SM* (which is patchy due to distinct soil hydraulic parameters associated with sharp soil class delineations), and a much stronger delineation between the Dry Chaco (high T_B) and the wetter region east of it (low T_B) than what could be expected based on CLSM *SM* or *Tsoil1* alone. In the northern Chaco, the denser forest vegetation in combination with the higher temperature contribute to the higher observed T_B values.

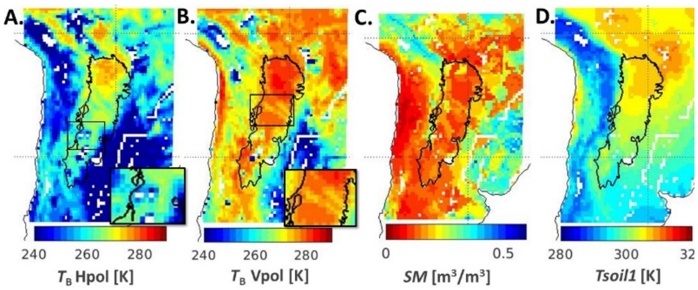

Fig. 6. Ten-year average of observed A. SMOS T_B at Hpol, B. SMOS T_B at Vpol, and simulated C. CLSM soil moisture and D. CLSM soil temperature (*Tsoil1*), after cross-masking to qualitative SMOS data (e.g., excluding frozen times in the Andes). Zoomed regions in A. and B. are discussed in the text.

Some interesting features in the T_B patterns are highlighted in Fig. 6A. and B. The zoom in Fig. 6A. is situated in the mid-west part of the Dry Chaco, and shows local low T_B and some missing pixels surrounding the Tucuman and Santiago Del Estero area of the field campaign. Pixels in the presence of a lake are filtered in the satellite data (Fig. 1), and land conversion in the area might contribute to the local decreases in T_B. The zoom in Fig. 6B. shows linear features spanning from east to north-west in the northern part of the Dry Chaco, matching the floodplains of the Bermejo and Pilcomayo River and areas with high excess salinity in the HWSD map. The same features can be found in the simulated *SM* map (Fig. 6C.), but less distinctly than in the T_B map, nicely illustrating how the T_B pattern gives an integrated view of the land surface.

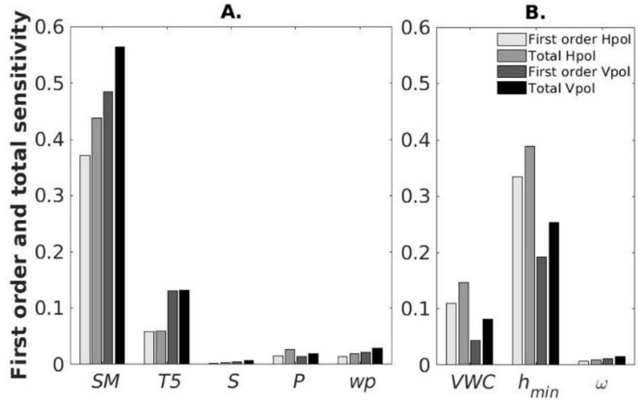

Fig. 7. First order and total sensitivity for A. soil-related and B. vegetation-related RTM input variables and parameters. Indices are calculated using a Monte Carlo simulation with 10^5 samples.

C. Forward T_B Simulations

Fig. 7A. and B. give an overview of the first order and total Sobol indices for T_B at H- and Vpol for soil- and vegetation-related RTM parameters respectively.

This global sensitivity analysis confirms that *SM*, *T5*, vegetation and roughness parameters have a significant contribution to the output T_B variability, whereas *S* only has a small influence. Therefore, we will use strong constraints of *SM* and *T5* (model background assumed to be known) in the following retrieval analysis to potentially tease out the marginal impact of the *S* signal.

The above analysis was supplemented with a local sensitivity analysis. Table III gives an overview of the simulated T_B sensitivity to *S* for a range of other RTM input variables. It should be noted that the tested salinity range of ΔS=35 PPT is a stretch for most soil-water mixtures, i.e. values above 30 PPT were only found in soils near salt lakes within the Dry Chaco. The change in T_B (ΔT_B) for a ΔS of 35 PPT is limited to -3.6 K at Hpol and -2.8 K at Vpol for average field conditions marked as 'initial settings' (open shrubland, loamy soil, T= 288 K, *SM* = 0.2 m^3/m^3, LAI = 0.3), and using uncalibrated 'Lit2' values [56] for RTM parameters such as e.g. $h_{\min}$, $h_{\max}$, ω, $b_{\rm h}$ and $b_{\rm v}$. However, the sensitivity $|\Delta T_B|$ increases to ~ 7 K with increasing T, increasing *SM* (in correspondence with literature, [61]), and decreasing LAI. Note that if the canopy temperature Tc is varied independently from T, the sensitivity of T_B to salinity increases with decreasing Tc (not shown). Vegetation classes with a denser canopy cover, such as forests, lower the T_B sensitivity to soil salinity.

Fig. 8 illustrates the forward simulation via full time series of input variables and simulated T_B for one random pixel of the field campaign (EASEv2 pixel A1) for (A.) ten years with SMOS data and (B.) the months of the field campaign. While model-based T_B simulations correctly follow the pattern of the satellite T_B observations, the absolute value is underestimated by as much as 20 K, because the model *Tsoil1* is not bias corrected here and the RTM parameters are not locally tuned. In contrast, the T_B simulation using in situ data at the day of field sampling, agrees well with the satellite T_B.

TABLE III

HPOL T_B (40°) SENSITIVITY TO SALINITY FOR VARIOUS RTM INPUT VARIABLES, WITH AN INDICATION OF THE INITIAL SETTING AND THE TESTED RANGE OF EACH VARIABLE. RTM PARAMETERS DEPEND ON THE VEGETATION CLASS AND ARE TAKEN FROM LITERATURE-BASED LOOKUP TABLE 'LIT2' IN [53]. THE LAST COLUMN QUANTIFIES THE DIFFERENCE IN T_B (ΔT_B) FOR A DIFFERENCE IN S (ΔS) OF 35 PPT.

Variable	Initial	Range	Effect on T_B	Effect on sensitivity of T_B to salinity (0 PPT – 35 PPT)
T ** [K]	288	[275 300]	T_B increases with T	Sensitivity increases with rising T; ΔT_B is minimally ~-7 K at $T5$ = 300 K
SM [m^3/m^3]	0.2	[0.1 0.4]	T_B decreases with SM	Sensitivity increases with rising SM; ΔT_B is minimally ~-5 K at SM = 0.4 m^3/m^3
LAI [m^2/m^2]	0.3	[0.1 3]	T_B increases with LAI	Sensitivity decreases with rising LAI; ΔT_B is minimally ~-4 K at LAI = 0.1 m^2/m^2
Vegetation class	7: Open shrubland	[1 16]	Mixed effect	Lower sensitivity in denser vegetation covers; ΔT_B is minimally ~-6 K for grass- and wetlands
Soil class	2: Loamy sand	[1 12]	Mixed effect	Sandy classes have slightly higher sensitivity; ΔT_B is minimally ~-4 K for silt loam texture

** Changing soil temperature T simultaneously changes canopy temperature Tc and the soil water temperature, which are all assumed equal.

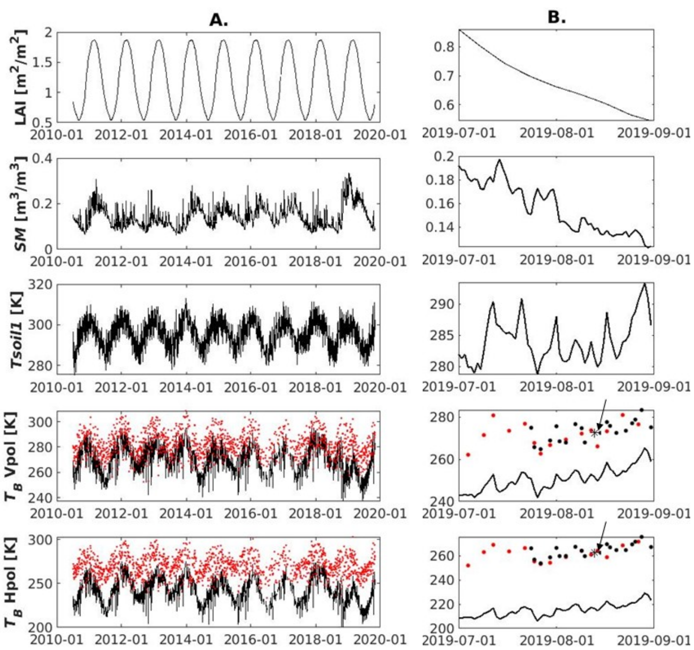

Fig. 8. Time series illustrating the concept of forward simulation with the RTM for EASEv2 pixel A1, for A. 10 years and B. the months of July and August 2019, when the field campaign took place. The top three panels show the CLSM LAI, SM and $Tsoil1$. The lowest two panels depict the observed Vpol and Hpol T_B by (red dots) SMOS, (black dots) SMAP, the (*, indicated by the arrows) in situ-based T_B simulation on the exact day of measurement in the field, and the (black line) model-based T_B simulations.

Both the model- and in situ-based T_B simulations at the time and location of the sampled EASEv2 pixels were compared with the closest SMOS (~ 6 am/pm local time) and SMAP (~6 pm/am local time) T_B on the sample day. Table IV gives a summary of the (spatio-temporal) accuracy metrics using a total of 31 satellite T_B observations (both SMOS and SMAP, am and pm overpasses were combined to ensure a sufficiently large sample size) for the visited EASEv2 pixels. For some of the 26 pixels, no satellite T_B observation was found at the day of field sampling, for other pixels multiple overpasses were available. The in-situ-based forward T_B simulations correlate better with satellite T_B observations than model-based T_B simulations, especially at Vpol. The low correlation between model-based T_B simulations and T_B observations is most likely due to the low correlation between model and in situ or satellite SM, and because the (soil- and vegetation-related) lookup RTM-parameters are associated with model-based vegetation classes that differ from those observed in situ. $Tsoil1$ bias correction adds bias to the Vpol and reduces the bias in the Hpol, but generally brings the model-based T_B simulations closer to the in situ-based simulations. Across the 31 sample points, Table IV indicates a relatively small average T_B bias compared to literature [57], but the ubRMSD suggests that large compensating differences are found across the sampled pixels.

In line with the low sensitivity of T_B to S, there is a slight but consistent increase in T_B simulation performance when S is included. At Hpol, the R value increases from 0.77 to 0.79 and at Vpol from 0.80 to 0.83. For saline pixels, defined as EASEv2 pixels with in situ S > 4 PPT, accounting for S causes an increase in R from 0.66 to 0.70 at Hpol and from 0.72 to 0.79 at Vpol. Given the small sample size, no statistically significant model improvement can be proven, but this finding shows that S can have a locally important impact on T_B.

TABLE IV
SPATIO-TEMPORAL PERFORMANCE METRICS FOR THE FORWARD T_B SIMULATIONS. SIMULATIONS WERE PERFORMED WITH IN SITU (IS) AND MODEL DATA (M) INPUT, EITHER WITHOUT OR WITH (+ S) IS SALINITY INPUT VALUES. T_B SIMULATIONS MARKED WITH * USE BIAS-CORRECTED *TSOIL1*. THE REFERENCE T_B OBSERVATIONS (OBS) INCLUDE ALL AVAILABLE SMOS AND SMAP (TOGETHER) T_B AT 40° INCIDENCE ANGLE COLLECTED FOR THE TIMES AND EASEV2 PIXELS OF THE FIELD SAMPLING. CORRELATIONS IN BOLDFACE ARE SIGNIFICANT AT A LEVEL OF α=0.05.

	R [-]	RMSD [K]	Bias [K]	ubRMSD [K]
Hpol				
T_B IS	**0.77**	12.70	-0.53	12.69
T_B IS + s	**0.79**	12.62	-0.44	12.61
T_B m	0.18	25.74	-9.35	23.98
T_B m + s	0.17	25.85	-8.89	24.24
T_B m + s*	0.18	24.76	-4.33	24.38
Vpol				
T_B IS	**0.80**	9.49	6.23	7.17
T_B IS + s	**0.83**	9.24	6.20	6.86
T_B m	0.13	14.67	0.77	14.56
T_B m + s	0.12	14.84	1.02	14.80
T_B m + s*	0.13	16.12	6.07	14.93

D. Retrieval of Land Surface Properties

The above T_B simulations were limited to some in-situ sampled EASEv2 pixels, and indicated that the spatial pattern of CLSM input and lookup RTM parameter values were not ideal to represent the observed T_B. In a next step, the long-term (2010 - 2019) discrepancy between simulated and observed T_B was leveraged to estimate some RTM properties at each pixel of the entire Dry Chaco. More specifically, vegetation, soil roughness and possibly salinity, and their uncertainty, were estimated via RTM inversion, using CLSM background information of *T5* and *SM* (which have a dominant effect on T_B, Fig. 7A.). The resulting estimates can be used as parameters in future forward RTM applications with consistent CLSM background information, or they can be interpreted as retrievals which are constrained by model background information (on *SM* and *T*).

Fig. 9 shows the spatial distribution of the 'best' and ensemble mean estimate for four diagnosed parameters retrieved (i.e. calibrated) with DREAM(ZS) over the 569 EASEv2 pixels located within the Dry Chaco, for two calibration cases. The reference calibration optimizes 5 parameters without inclusion of *S* or *Tsoil1* bias correction. The second case calibrates 7 parameters, including s_a and s_b (to diagnose *S*), and applies a *Tsoil1* bias correction. For the latter, the spatial average best or ensemble mean value ± the spatial average ensemble standard deviation is also shown. The ensemble standard deviation is an indication of the retrieval uncertainty. In any case, the RTM inversion yields spatially continuous parameter values, unlike the literature-based values associated with the few different vegetation classes in the region. Not explicitly shown is that the difference in distributions for the two calibration experiments is due to the *Tsoil1* bias correction, whereas the calibration of *S* has no significant impact on the spatial distribution of the retrieved parameters.

Overall, the inversion (or RTM calibration) yields realistic estimates of vegetation (long-term LAI-based vegetation optical depth <τ> and scattering albedo ω) and roughness (long-term *SM*-based <*h*>) with a low associated uncertainty. Including a bias-correction of *Tsoil1* results in lower <*h*> and higher <τ> (via higher *b* parameters) and ω values.

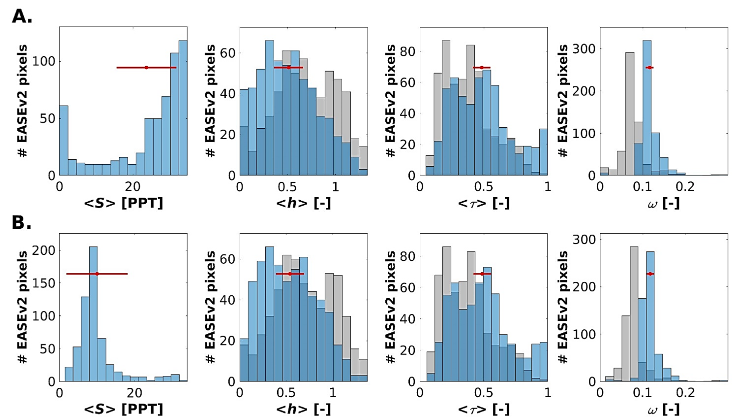

Fig. 9. Spatial distribution of four parameters calibrated with DREAM(ZS) for the Dry Chaco, with (red) an indication of the uncertainty. Distribution of (A.) the 'best' estimate and (B.) the ensemble mean value for each parameter, when (gray) calibrating 5 parameters without *Tsoil1* bias correction, and (blue) calibrating 7 parameters, incl. salinity equivalents, with *Tsoil1* bias correction. Also shown is (red) the spatial average (A.) 'best' and (B.) ensemble mean value ± the spatial average ensemble standard deviation corresponding to the blue parameter distribution.

The <τ> estimates were evaluated against independent vegetation products, i.e. 2015-2019 average <τ> retrievals from SMOS and SMAP, and 2010 AGB estimates. Fig. 10 illustrates that both (A.) without and (B.) with *Tsoil1* bias correction, a high spatial correlation is found between long-term best estimates of <τ> and long-term SMOS-IC <τ> (R~0.95), long-term SMAP <τ> (R=0.85-0.90), and 2010 AGB estimates (R~0.80), at locations where SMAP retrievals are assumed to be of high quality (i.e., excluding areas with too dense vegetation, indicated in Fig. 10C.). To assess the usefulness of the AGB dataset for 2010 to evaluate climatological <τ> retrievals, we also computed the correspondence between the retrieved <τ> with the SMOS-IC <τ> for the year 2010, resulting in R=0.92 and RMSD=0.07 [-]. Land use changes in the area typically result in local mosaic patterns that probably have not significantly changed the coarse-scale τ patterns over the last decade. Again, the differences between Fig. 10A.-B. are mainly due to the *Tsoil1* bias correction, and the results only marginally differ with or without inclusion of *S* calibration.

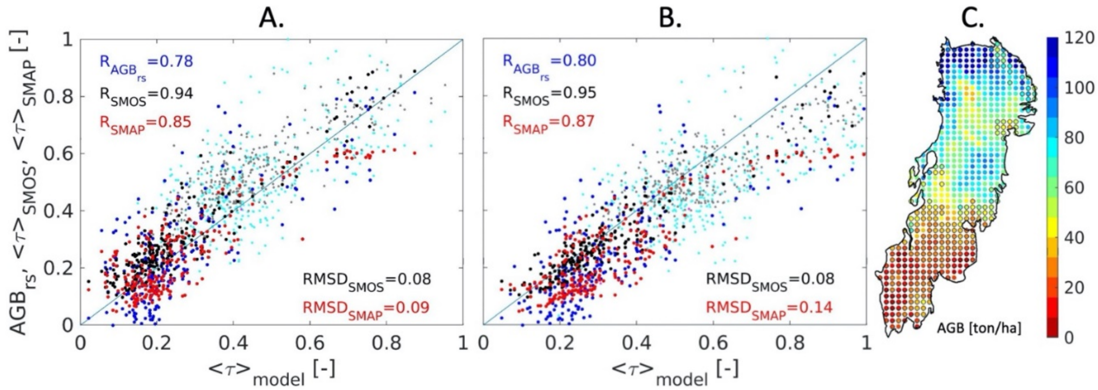

Fig. 10. Comparison of long-term mean retrieved $<\tau>$ ($<\tau>_{model}$) (A.) without and (B.) with inclusion of *Tsoil1* bias correction against (black) SMOS $<\tau>$, (red) SMAP $<\tau>$, and (blue) linearly rescaled 2010 AGB (AGB_{rs}). *S* calibration is included, but does not affect the skill metrics. The 'best' $<\tau>$ values are shown; the metrics are the same for the ensemble mean $<\tau>$ values. The lighter and smaller markers correspond to (cyan) AGB and (gray) SMOS estimates where SMAP retrievals are masked out, and these are not included in the skill metrics. C. Distribution of 36-km AGB values. Markers with black edges correspond to grid cells where quality SMAP retrievals are found.

The smaller and lighter markers in Fig. 10 A.-B. show that in dense vegetation areas (high $<\tau>$ values) the inclusion of a *Tsoil1* bias correction possibly leads to an overestimation of absolute $<\tau>$ values, esp. compared to the SMAP $<\tau>$ (RMSD metrics not reported; too few high τ values left after applying quality control). This finding, combined with locally increased misfits between long-term mean T_B observations and T_B simulations in case of *Tsoil1* bias correction (not shown), leads to the conclusion that our in situ-based *Tsoil1* bias correction is not suitable everywhere, and that the use of *Tsoil1* as input for the SMAP L4_SM RTM is justified for this area.

Even though the calibration yields realistic estimates of $<\tau>$, ω, and $<h>$ with a low associated uncertainty, Fig. 9 shows that the (10-year average) $<S>$ takes on unrealistically high values in the Dry Chaco, especially so when the 'best' values for s_a and s_b are used in the calculation of *S*. The retrieved space-time average value for the pixels sampled during the field campaign is ~12 PPT when $<S>$ is calculated using the ensemble mean s_a and s_b values and ~ 28 PPT when the best values for s_a and s_b are used, whereas the average surface salinity measured in situ, which comprised only a small part of the Dry Chaco, was 4 PPT. Furthermore, the *S* estimates computed at the sample days do not at all correlate (R=-0.19) with in situ measurements for the EASEv2 pixels of the field campaign. Fig. 9 also highlights a large ensemble uncertainty on the $<S>$ estimates and a large discrepancy between the 'best' estimate and the ensemble mean estimate, which is indicative of a wide and skewed a posteriori distribution of unreliable *S* estimates, and thus highly uncertain estimates. Fig. 11A. gives the ensemble mean retrieved 10-year average $<S>$ pattern in and near the Dry Chaco and its near surroundings, calculated using the ensemble mean s_a and s_b parameters, calibrated without *Tsoil1* bias correction. The retrieved $<S>$ values are too high to represent natural or human-induced salinity, are often of the same magnitude as their uncertainty, and likely represent an integrated correction on the dielectric constant in terms of salinity equivalents, rather than salinity itself. High values of $<S>$ thus compensate for shortcomings in the RTM input variables, other than salinity. Some of the ensemble mean $<S>$ pattern can for example be related to the presence of periodic open water fractions shown in Fig. 11 B.

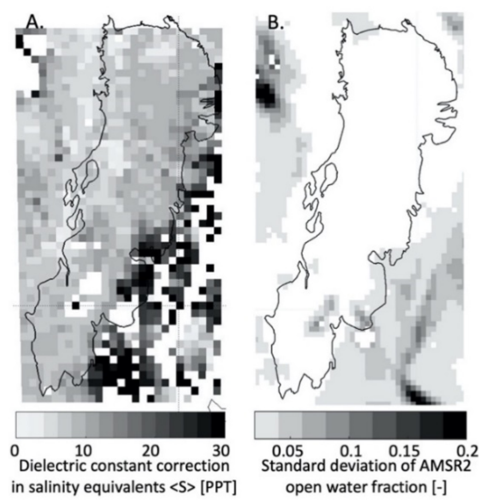

Fig. 11. A. 10-year averaged retrievals of the ensemble mean correction of the dielectric constant in terms of salinity $<S>$ [PPT], calibrated without *Tsoil1* bias correction. B. 2015-2017 standard deviation of AMSR2 open water fraction.

V. Discussion

An evaluation of satellite-based *SM* retrievals using in situ measurements at the large 36-km pixel scale has so far only been done for a few dedicated calibration and validation (Cal/Val) sites [62]. For this study, in situ measurements were collected within 26 different satellite pixels during the dry season in the Dry Chaco and -despite the small sample size- they confirmed that SMOS and SMAP *SM* retrievals agree well (R~0.7), and better than CLSM simulations, with ground measurements. However, when including all SMAP retrievals across the years 2015-2019, the CLSM simulations only showed a small bias of 0.02 m^3/m^3 across the sampled pixels, and a noteworthy wet bias in the retrievals for saline pixels (Fig.

5). The latter indicates that missing salinity may affect the *SM* retrievals, because the salinity decreases T_B by a few K, especially for high *T* and low LAI, as in the Dry Chaco. As a rule of thumb, a 2-3 K decrease in T_B (40°) corresponds to an increase of 0.01 m^3/m^3 in *SM* retrievals [63]. The sensitivity analysis (Table III) showed that the influence of salinity on T_B is only a few K, and thus is close to the uncertainty on individual (single angle) SMOS T_B observations (~4 K), and only marginally above the T_B differences (< 3 K) between SMOS and SMAP sensors. This necessitates the combined use of multiple T_B observations and strong modeled background information to tease out the influence of salinity on the T_B.

The long time series of multi-angular and multi-polarization SMOS T_B data was used to obtain reliable retrievals of vegetation (τ, ω) and microwave roughness (*h*), but the retrieval of salinity (*S*) did not return realistic values and did not significantly affect the retrieval of other land surface properties. The unrealistic high *S* values contrast with our own fieldwork and the level of dryland salinization indicated in literature, that is, the observed surface salinity is still low at this time, whereas the salinity is higher mainly in deeper layers (see Appendix 1). Because of the low sensitivity of the T_B simulations to *S*, the low observed surface *S,* and the low spatial resolution of L-band radiometers, sub-pixel heterogeneity makes it hard to disentangle *S* from other inaccurate sources of input to the T_B signal, which explains the uncertainty in the *S* retrievals. The *S* estimates compensate for errors introduced in the computation of the dielectric constant ε by inaccurate data other than salinity. For example, southeast of the Dry Chaco, the calibrated salinity equivalents are almost certainly compensating for missed open water, organic material and underestimated CLSM *SM*. Fig. 11B. confirms that the area southeast of the Dry Chaco experiences temporal ponding, which is not included in the CLSM background simulations, and which was not persistent enough to flag SMOS T_B prior to the inversion. The high *S* equivalents are effectively increasing ε'', resulting in a higher magnitude of ε, similar to how higher water amounts would affect ε'.

The retrieval approach could be further elaborated by improving the strong model background constraints, or by adding more observational constraints e.g. by including information on water ponding. In our study, long-term T_B signature statistics were optimized, using (i) simulated *T* and *SM*, and (ii) imposed time series relationships between *S* and *SM*. While the first constraint ensures temporally and spatially consistent fields of RTM parameters that are readily applicable for forward modeling in a land surface data assimilation system [5], it may also introduce bias in the *S* (and other) retrievals. The bias in the large-scale CLSM *SM* and *T* could originate from the MERRA-2 forcings or CLSM parameters, and such biases are almost unavoidable, unless extensive in situ observations would be available for model optimization. However, including a calibration of the long-term *SM* and *T* background bias as part of the inversion may reduce this problem. We verified (results not shown) that by keeping the modeled *SM* dynamics, but including a rescaling factor (bias correction) for *SM* (and consistent porosity) in the calibration, the *S* values significantly decreased (by on average 5 PPT), and the inclusion of an *S* retrieval has a local impact on the magnitude of the *SM* rescaling factor. The latter reinforces that *SM* retrieval might locally be affected when high *S* is present. If available, it can be recommended to include knowledge about *S* as ancillary information (constraint) to improve *SM* retrieval accuracy over salt-affected areas. The constraint on the relationship between *S* and *SM* could also be improved by a more elaborate physically-based model, which would then allow to better estimate the temporal variation of *S*.

Finally, the low spatial resolution of passive microwave remote sensing is great for large-scale ecosystem monitoring, but not ideal for agricultural applications. Even the largest farms in Argentina are on average only half the size of the SMOS and SMAP pixel scale [64]. Finer resolution active microwave data could solve that problem of spatial resolution, but decomposing the backscatter signal is not trivial.

VI. Conclusion

The Dry Chaco is a biogeographical region with a distinct land surface characterized by dry forest and expanding agriculture, possibly threatened by a changing water distribution and salinization. This paper highlights limitations and possibilities of various data sources in capturing coarse-scale surface soil moisture (*SM*), soil temperature, soil salinity (*S*) and vegetation in the Dry Chaco. More specifically, we examine L-band microwave brightness temperature (T_B) observations and retrievals from the Soil Moisture Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) satellite missions, Catchment Land Surface Model (CLSM) simulations, and in situ measurements within 26 satellite pixels covering a part of the Dry Chaco.

First, data from an intensive field campaign in July-August 2019, CLSM simulations, and SMOS and SMAP *SM* retrievals are compared. Across the 26 pixels sampled in the field campaign (each observed once, on different days), the CLSM-based *SM* does not correlate well with the in situ measurements, whereas SMOS and SMAP *SM* retrievals correlate much better with in situ data. The CLSM daytime surface soil temperature (*Tsurf*) is slightly colder (2 K) than in situ temperature (*T5*) measurements (both 0-5 cm), and CLSM's first layer soil temperature at 5-15 cm (*Tsoil1*) is colder by ~5 K compared to the 0-5 cm in situ *T5* data, due to differences in spatial (horizontal and vertical) representativeness. When comparing satellite-based *SM* retrievals to CLSM *SM* across the years 2015-2019 for the sampled pixels, a wet bias in the *SM* retrievals for saline pixels was detected.

Next, the effect of *S* and other land surface variables on forward L-band T_B simulations is quantified. To this end, we implemented the equations of [18] for saline water into the dielectric mixing model of [21] to estimate the dielectric constant of the soil-water-salinity mixture. When propagating these dielectric constant estimates through the L-band radiative transfer model (RTM), T_B shows an overall low sensitivity to *S* (decrease by ~4 K when *S* increases from 0 to 35 PPT under average field campaign conditions, and using literature-based RTM parameters), with increases in sensitivity when *SM* increases, vegetation decreases or soil temperature increases. Because the 26 in-situ sampled satellite pixels have on average only a *S* of 4 PPT, the forward T_B simulations only change marginally on average when accounting for *S* in the RTM, but local stronger impacts are found. In contrast, T_B simulations using in situ-based soil temperature and *SM* greatly outperform those using CLSM-based input data, when compared to SMOS and SMAP T_B observations. In line with the SMAP L4_SM product, CLSM *Tsoil1* is an input to the RTM. An optional *Tsoil1* bias correction using the *T5* data of our field campaign only marginally improves the forward L-band RTM.

Finally, we use the RTM in inverse mode to estimate time-average vegetation (τ, ω), microwave roughness (h) and salinity and their uncertainty at each pixel in the Dry Chaco, using Markov chain Monte Carlo simulations, 10 years of multi-angular and dual-polarization SMOS T_B observations and constraints of CLSM *SM*, temperature and leaf area index. The RTM inversion retrieves consistent spatial patterns for h_{min}, Δh (related to microwave roughness h), b_h, Δb (related to vegetation opacity τ), and ω (related to vegetation). The calibrated pattern of 10-year average $<\tau>$ agrees very well with independent SMOS and SMAP $<\tau>$ retrievals (R>=0.9), and with above ground biomass (AGB) estimates. The inclusion of an in situ-based *Tsoil1* bias correction in the retrieval is not recommended for the entire Dry Chaco. The inclusion of *S* in the retrieval does not significantly alter the values of the retrieved vegetation and roughness parameters. However, the retrieved *S* values themselves are unrealistically high with a large associated uncertainty over the Dry Chaco, but they help to slightly reduce the differences between simulated T_B and SMOS T_B observations. The *S* retrievals should thus not be seen as *S* estimates as such, but rather as a bulk correction of the dielectric constant that also compensates for inadequate CLSM *SM* values, neglected open water contributions, or other model deficiencies.

To retrieve soil surface *S* from the microwave L-band signal, the *S* levels should be high enough and the uncertainty on the 'known' RTM input variables should be minimized to maximize the sensitivity to *S*. For future research, we suggest improving the model background information, e.g. by including both local and seasonal soil temperature and *SM* bias corrections (incl. the impact of open water fraction) and more accurate soil texture information. Future research would also benefit from a study area where soil surface salinization is in a further stage than in the Dry Chaco, to overcome T_B sensitivity issues at low *S*.

Appendix A

At ten locations (Fig. 1B.), deeper soil measurements of EC and pH were collected. Like the surface soil EC measurements, the deeper EC measurements were performed in a home laboratory setting, and based on a 1:1 soil-water mixture sample. Eight of the sample locations were chosen based on deforestation history and were situated at the interface between forest and agriculture. At those locations, measurements were taken along three transects: one in the agricultural area, one in the forest and one moving from agriculture to forest. Each of the transects consisted of two to five sample sites. A pit of approximately 40 cm deep was dug at every sample site and soil moisture, temperature and dielectric properties were measured with the HydraGO and ThetaProbe at 5 – 10 cm and 20 – 40 cm depth. With a soil auger, deeper soil samples at 80 – 100 cm and 200 cm depth were collected for salinity analysis with the Hanna and YSI proDSS probes. Preparation and analysis of those samples followed the same steps as discussed for the surface soil samples. Following the same data collection strategy, a transect of six sample sites along a river and another one along an elevation gradient was sampled. Fig. A1 shows that, currently, the surface soil salinity in the Dry Chaco is still low, whereas deeper soil layers have significantly higher salinity values, and the correlation between surface and deeper salinity decreases with depth, i.e., R=0.70 (184 samples) at 80 – 100 cm depth and R=0.42 (34 samples) at 200 cm depth.

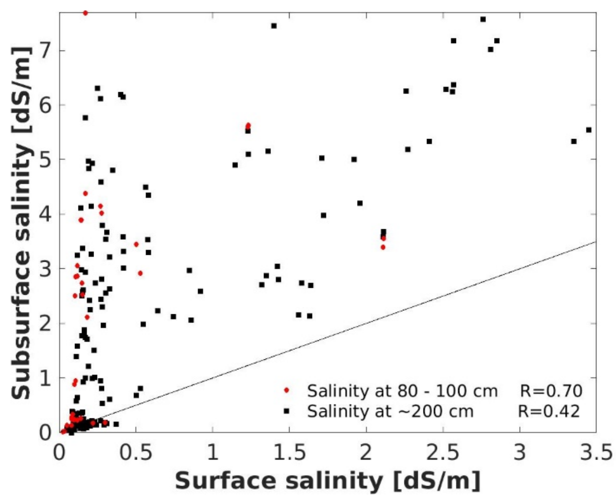

Fig. A1. Salinity measurements at 80 - 100 cm and at approximately 200 cm depth versus soil surface (0 - 20 cm depth) salinity measurements. R is the Pearson correlation coefficient. The line represents the 1:1 line. Data were taken at the deeper soil sample sites, indicated in Fig. 1B.

Appendix B

The dielectric mixing model of [21] calculates the dielectric constant of the soil based on the different constituents of the soil:

- air: $\varepsilon_a = 1$
- rock: $\varepsilon_r = 5.5 + 0.2i$
- tightly bound water approximated by the dielectric constant of ice: $\varepsilon_i = 3.2 + 0.1i$
- free water ε_w, based on our modified model of [18].

Further, the model differentiates between the dielectric constant of the soil when the soil moisture content (*SM*) is below or above a certain transition level, because the permeability to electricity differs for tightly bound water from free water. The transition water content W_t is calculated as:

$$W_t = 0.49WP + 0.165 \quad \text{(B1)}$$

where WP [m^3/m^3] is the wilting point of the soil. When the soil moisture content SM [m^3/m^3] is lower than W_t, the dielectric constant of the soil can be calculated as:

$$\varepsilon_{soil} = SM\varepsilon_x + (P - SM)\varepsilon_a + (1 - P)\varepsilon_r \quad \text{(B2)}$$

where P is the soil porosity and ε_x is the dielectric constant of the initially absorbed water, calculated as:

$$\varepsilon_x = \varepsilon_i + (\varepsilon_w - \varepsilon_i)\frac{SM}{W_t}y \quad \text{(B3)}$$

where y is a fit parameter: $y = -0.57WP + 0.481$. When SM is higher than W_t, the dielectric constant of the soil can be calculated as:

$$\varepsilon_{soil} = W_t\varepsilon_x + (SM - W_t)\varepsilon_w + (P - SM)\varepsilon_a + (1 - P)\varepsilon_r \quad \text{(B4)}$$

with $\varepsilon_x = \varepsilon_i + (\varepsilon_w - \varepsilon_i)y$.

Acknowledgment

We thank Ignacio Gasparri, Martin Requena, Raul Giménez, Javier Houspa and Samuel Massart for their assistance with the field campaign. We also thank the anonymous reviewers for their constructive feedback.

References

[1] D. Entekhabi, D., Rodriguez-Iturbe, I., & Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. Journal of Hydrology, 184(1–2), 3–17.

[2] V. Humphrey, *et al.* (2021). Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature*, 592(7852), 65–69.

[3] S.V. Briggs, and N. Taws. (2003). Impacts of salinity on biodiversity - Clear understanding or muddy confusion? *Australian Journal of Botany*, 51(6), 609–617.

[4] D. L. Corwin (2021). Climate change impacts on soil salinity in agricultural areas. *European Journal of Soil Science*, 72(2), 842–862.

[5] R. H. Reichle, *et al.* (2019). Version 4 of the SMAP Level-4 Soil Moisture Algorithm and Data Product. *Journal of Advances in Modeling Earth Systems*, 11(10), 3106–3130.

[6] Y. H. Kerr, *et al.* (2010). The SMOS mission: new tool for monitoring key elements of the global water cycle. Proceedings of IEEE, 98(5), 666-687.

[7] D. Entekhabi, *et al.* (2015). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE,* 98(5), 704-716.

[8] J.-P. Wigneron, *et al.* (2021). SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sensing of Environment*, 254, 112238.

[9] J. Chaubell, *et al.* (2020). Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil Moisture. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6), 3894–3905.

[10] X. Li, J.-P. Wigneron, L. Fan, F. Frappart, S. H. Yueh, A. Colliander, A. Ebtehaj, L. Gao, R. Fernandez-Moran, X. Liu, M. Wang, H. Ma, C. Moisy, P. Ciais (2022). A new SMAP soil moisture and vegetation optical depth product (SMAP-IB): Algorithm, assessment and inter-comparison. *Remote Sensing of Environment*, 271, 112921.

[11] FAO/IIASA/ISRIC/ISSCAS/JRC. (2012). Harmonized world soil database (version 1.2). Rome, 19 Italy and Laxenburg, Austria: FAO and IIASA.

[12] X. Fan, Y. Weng, and J. Tao (2016). Towards decadal soil salinity mapping using Landsat time series data. *International Journal of Applied Earth Observation and Geoinformation*, 52, 32–41.

[13] G. I. Metternicht and J. A. Zinck, J. A. (2003). Remote sensing of soil salinity: Potentials and constraints. *Rem Sens Environ.*, 85(1), 1–20.

[14] A. Hassani, A. Azapagic, and N. Shokri, N. (2020). Predicting long-term dynamics of soil salinity and sodicity on a global scale. *Proceedings of the National Academy of Sciences*, 117(52), 33017–33027.

[15] K. Ivushkin, H. Bartholomeus, A. K. Bregt, A. Pulatov, B. Kempen, and L. de Sousa (2019). Global mapping of soil salinity change. *Remote Sensing of Environment*, 231, 111260.

[16] J. Boutin, N. Martin, X. Yin, J. Font, N. Reul, and P. Spurgeon, P. (2012). First assessment of SMOS data over open ocean: Part II-sea surface salinity. *IEEE Trans. on Geoscience and Remote Sensing*, 501662–1675.

[17] M. T. Hallikainen, F. T. Ulaby, M. Dobson, M. El-Rayes, and L. Wu (1985). Microwave Dielectric Behavior of Wet Soil-Part I: Empirical models. *IEEE Trans. Geoscience and Remote Sensing*, GE-23(1), 25–34.

[18] A. Stogryn (1971). Equations for Calculating the Dielectric Constant of Saline Water. *IEEE Trans Microwave Theory and Techniques*, 5, 733–736.

[19] L. A. Klein and C. T. Swift (1977). An Improved Model for the Dielectric Constant of Sea Water at Microwave Frequencies. *IEEE Journal of Oceanic Engineering*, 2(1), 104–111.

[20] L. Chaturvedi, *et al.* (1983). Multispectral Remote Sensing of saline seeps. *IEEE Trans Geoscience and Remote Sensing,* GE-21(3), 239-251.

[21] T. J. Jackson, P. E. O'Neill (1987). Salinity Effects on the Microwave Emission of Soils. *IEEE Trans Geoscience and Remote Sensing*, GE-25(2), 214–220.

[22] J. R. Wang and T. J. Schmugge (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans Geoscience and Remote Sensing*, GE-18(4), 288–295.

[23] K. McColl, D. Ryu, V. Matic, J. P. Walker, J. Costelloe, J., and C. Rüdiger, (2012). Soil Salinity Impacts on L-Band Remote Sensing of Soil Moisture. *IEEE Geoscience and Remote Sensing Letters,* 9(2), 262–266.

[24] E. G. Jobbágy, R. Giménez R., V. Marchesini, Y. Díaz , B. Jayawickreme and M. Nosetto. (2020). Salt Accumulation and Redistribution in the Dry Plains of Southern South America: Lessons from Land Use Changes, in: R. Lavado and E. Taleisnik (editors), *Saline and Alkaline Soils in Latin America*. Springer, Switzerland, p. 51-70.

[25] V. Marchesini, M. Nosetto, J. Houspanossian, J., and E. Jobbágy (2020). Contrasting hydrological seasonality with latitude in the South American Chaco: The roles of climate and vegetation activity. *Journal of Hydrology,* (587), 124933.

[26] M. Vallejos, J. Volante, M. Mosciaro, L. Vale, M. Bustamante, and J. Paruelo (2015). Transformation dynamics of the natural cover in the Dry Chaco ecoregion: A plot level geo-database from 1976 to 2012. *Journal of Arid Environments*, 123(26), 3–11.

[27] I. Gasparri, and G. Baldi (2013). Regional patterns and controls of biomass in semiarid woodlands: Lessons from the Northern Argentina Dry Chaco. *Regional Environmental Change*, *13*(6), 1131–1144.

[28] M. Moretti, H. Morrás, F. Pereyra, and G. Schulz (2019). Soils of the Chaco region. In: Rubio, G., Lavado, R.S. and Pereyra, F. (editors), *The soils of Argentina.* 149-160.

[29] P. Boletta, A. Ravelo, A. Planchuelo, and M. Grilli (2006). Assessing deforestation in the Argentine Chaco. *Forest Ecology and Management,* 228(1–3), 108–114. https://doi.org/10.1016/j.foreco.2006.02.045

[30] R. Giménez, J. Mercau, M. Nosetto, R. Páez, and R. Jobbágy (2016). The ecohydrological imprint of deforestation in the semiarid Chaco: insights from the last forest remnants of a highly cultivated landscape. *Hydrological Processes,* 30(15), 2603–2616.

[31] G. De Lannoy, and R. Reichle (2016) Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences*, 20, 4895–4911.

[32] L. Gao, A. Ebtehaj, M. Julian, M. Sadeghi, X. Li, and J.-P. Wigneron (2021). Reappraisal of SMAP inversion algorithms for soil moisture and vegetation optical depth. *Remote Sensing of Environment*, 264, 112627.

[33] R. Koster, M. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000). A catchment-based approach to modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration. *Journal of Geophysical Research Atmospheres*, 105(D20), 24823–24838.

[34] R. Gelaro, *et al.* (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454.

[35] S. Mahanama et al. (2015) Land Boundary Conditions for the Goddard Earth Observing System Model Version 5 (GEOS-5) Climate Modeling System - Recent Updates and Data File Descriptions. NASA Technical Report Series on Global Modeling and Data Assimilation 104606, Vol. 39, NASA Goddard Space Flight Center, MD, USA, 51p.

[36] M. Maertens, G. J. M. De Lannoy, S. Apers, S. Kumar, S. Mahanama (2021). Simulating the Dry Chaco Hydrology under Deforestation Pressure. *Hydrology and Earth System Sciences, 25(7),*4099-4125.

[37] P. Bobrov (1999). Estimating of soil salinity by passive microwave observations at C band. International Geoscience and Remote Sensing Symposium (IGARSS), 2, 1105–1107.

[38] B. Choudhury, T. Schmugge, A. Chang, and R. Newton (1979). Effect of surface roughness on the microwave emission from soils. *Journal of Geophysical Research*, 89(C9), 5699–5706.

[39] M. Seyfried, L. Grant, E. Du and K. Humes. (2005). Dielectric Loss and Calibration of the Hydra Probe Soil Water Sensor. *Vadose Zone Journal*, 4(4), 1070–1079. https://doi.org/10.2136/vzj2004.0148

[40] Lewis, E. L., and Perkin, R. G. (1981). The practical salinity scale 1978: conversion of existing data. Deep Sea Research Part A, *Oceanographic Research Papers*, 28(4), 307–328.
[41] D. Yamazaki, *et al.* (2017). A high accuracy map of global terrain elevations. *Geophysical Research Letters*, vol.44, pp.5844-5853.
[42] S. Kim (2013). Ancillary Data Report: Landcover Classification. Jet Propulsion Laboratory California Institute of technology. JPL, SMAP Science Document no. 042, D-53057.
[43] R. Myneni, Y. Knyazikhin,, and T. Park - Boston University and MODAPS SIPS - NASA. (2015). MOD15A3H MODIS/Combined Terra+Aqua Leaf Area Index/FPAR Daily L4 Global 500m SIN Grid. NASA LP DAAC. http://doi.org/10.5067/MODIS/MOD15A3H.006
[44] G. J. M. De Lannoy, R. Koster, R. Reichle, S. Sarith, and Q. Liu (2014). An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *JAMES*, 6(4), 957-979.
[45] B. Lehner, and P. Doell (2004): Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* (296)1-4, 1-22.
[46] J. Du, J. Kimball, L. Jones, Y. Kim, J. Glassy, and J.D. Watts (2017). A global satellite environmental data record derived from AMSR-E and AMSR2 microwave earth observations, 2017. *ESSD*, 9, 791-808.
[47] M. Santoro, *et al.* (2018). GlobBiomass global above-ground biomass and growing stock volume datasets, available on-line at http://globbiomass.org/products/global-mapping
[48] N. Rodríguez-Fernández *et al.*(2018). An evaluation of SMOS L-band vegetation optical depth (L-VOD) data sets: high sensitivity of L-VOD to above-ground biomass in Africa. *Biogeosciences,* 15(14), 4627–4645.
[49] M. Dobson, F. Ulaby, M. Hallikainen, and M. El-rayes (1985). Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models. *IEEE Trans Geoscience and Remote Sensing*, GE-23(1), 35–46.
[50] F. Ulaby, and D. Long (2014). Microwave dielectric properties of natural earth materials. In *Microwave radar and radiometric Remote Sensin*g, 119-159. University of Michigan Press. Ann Arbor.
[51] C. Mätzler, C. (2006*). Thermal Microwave Radiation: Applications for Remote Sensing*. The Institution of Electrical Engineers, London UK.
[52] I. M. Sobol (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation,* 55, 271–280.
[53] S. Marelli, S. and B. Sudret, (2014). UQLab: A Framework for Uncertainty Quantification in MATLAB In The 2nd International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2014), University of Liverpool, United Kingdom, July 13-16, pp. 2554–2563.
[54] R. Bindlish, T. Jackson, S. Chan, A. Colliander, and Y. Kerr (2019). Integration of SMAP and SMOS L-band observations. IGARSS, Yokohama Japan, July 28- August 2.
[55] X. Li, *et al.* (2021). Global long-term brightness temperature record from L-band SMOS and SMAP observations. IGARSS, Brussels.
[56] J. Quets, *et al*. (2019). Uncertainty in soil moisture retrievals: An ensemble approach using SMOS L-band microwave data. *Remote Sensing of Environment*, 229, 133–147.
[57] G. J. M. De Lannoy, R. Reichle, and V. Pauwels (2013). Global Calibration of the GEOS-5 L-Band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations. *Journal of Hydrometeorology*, 14(3), 765–785.
[58] G. J. M. De Lannoy, R. H. Reichle, and J. A. Vrugt (2014). Uncertainty quantification of GEOS-5 L-band radiative transfer model parameters using Bayesian inference and SMOS observations. *Remote Sens. Environ.*, 148, 146-157.
[59] Vincent, F. (2020). Salinity detected from space: a case study for the Dry Chaco, KU Leuven, 76 p [M.S. thesis dissertation].
[60] J. Vrugt, C. ter Braak, C. Diks, B. Robinson, J. Hyman, and D. Higdon (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Intl J Nonlinear Sciences and Numerical Simulation* 10 (3), 273-290.
[61] Y. Lasne, *et al.* (2007). Effect of salinity on the dielectric properties of geological materials: Implication for soil moisture detection by means of remote sensing. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3689–3693.
[62] A. Colliander, *et al.* (2017). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.*, 191, 215–231.
[63] T. Jackson *et al.* (1999). Soil moisture mapping at regional scales using microwave radiometry: the Southern great plains hydrology experiment. *IEEE Trans Geoscience and Remote Sensing,* 37(5pt.1), 2136–2151.
[64] L. Munoz, *et al.* (2021) Digital Agriculture Profile: Argentina. Rome (Italy): FAO. 13 p.

Frederike Vincent obtained a M.S. degree in Bioscience Engineering from KU Leuven in 2020, Belgium.

Michiel Maertens obtained a M.S. degree in Geography and Ph.D. degree in Bioscience Engineering (2021) from KU Leuven, Belgium.

Michel Bechtold obtained a M.S. degree ("Diplom") in geosciences from the University of Göttingen, Germany, and the Ph.D. degree in geosciences (2012) from University of Bonn, Germany. He is currently holding a research expert position on water and carbon cycle processes at KU Leuven, Belgium.

Esteban Jobbagy received a M.S. in Argonomy from UBA Argentina, and a Ph.D. in Biology (2002) from Duke University, US. He is currently a senior scientists at the Universidad Nacional de San Luis, Argentina.

Rolf H. Reichle received the M.S. degree ("Diplom") in physics from the University of Heidelberg, Heidelberg, Germany, and the Ph.D. degree in environmental engineering (2000) from the Massachusetts Institute of Technology, Cambridge, MA, USA. He is currently a Research Physical Scientist with the Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, MD, USA. His research interests include land data assimilation, satellite-based remote sensing, and applications related to land–atmosphere interactions, weather prediction, and seasonal climate forecasting.

Veerle Vanacker received a M.S. degree and Ph.D. degree from KU Leuven, and is currently at UC Louvain, Belgium

Jasper Vrugt received his M.S. and Ph.D. (2004) from University of Amsterdam. He is currently an associate professor at the University of California, Irvine, USA.

Jean-Pierre Wigneron received the M.S./Engineering degree from SupAéro, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace (ENSAE), Toulouse, and the Ph.D. degree from the University of Toulouse, France. He is currently a senior research scientists at the Institut National de Recherche Agronomiques (INRA) and Head of the INRA/ISPA remote sensing team, Bordeaux, France.

Gabriëlle J. M. De Lannoy received M.S. degrees in Bioscience Engineering, Industrial Automation, and Academic Education, and a Ph.D. degree in Bioscience Engineering (2006) from UGent, Belgium. She is currently a full professor at the KU Leuven, Belgium, leading the research group on Land Surface Remote Sensing, Modeling and Data Assimilation. Her research interests include land data assimilation, satellite-based remote sensing, and applications related to soil moisture, snow, vegetation and their interactions with the atmosphere and human activities.