

On the Use of Native Resolution Backscatter Intensity Data for Optimal Soil Moisture Retrieval

Theresa C. van Hateren^{ID}, Marco Chini^{ID}, *Senior Member, IEEE*, Patrick Matgen, *Member, IEEE*, Luca Pulvirenti^{ID}, *Member, IEEE*, Nazzareno Pierdicca^{ID}, *Senior Member, IEEE*, and Adriaan J. Teuling

Abstract—The accuracy of soil moisture estimated from synthetic aperture radar (SAR) backscatter data at high resolutions is limited by speckle. A common practice to mitigate speckle is to multilook the data prior to retrieving soil moisture. While multilooking indeed reduces speckle, it also decreases the spatial resolution and removes possibly useful high resolution information from the data. We, therefore, hypothesized that using higher resolution backscatter data for soil moisture retrieval would lead to higher retrieval accuracies. A high resolution field study combined with a synthetic experiment showed that calculating soil moisture prior to multilooking to the final target resolution calculate-then-average (CtA) has substantial advantages over the average-then-calculate (AtC) approach. Currently, the AtC strategy is most often applied in soil moisture studies, mainly due to its computational advantage compared to the CtA approach. We show that by making use of a higher source resolution backscatter data than the target resolution, we could improve soil moisture retrieval over an agricultural field.

Index Terms—High resolution, soil moisture, synthetic experiment.

I. INTRODUCTION

SOIL moisture strongly influences meteorology, climate, and hydrology via its direct impact on evapotranspiration [1]. Soil moisture conditions influence drought events [2] and flash flood magnitude [3]. Observations of soil moisture are, therefore, imperative to monitoring and predicting extreme hydrological events. Unfortunately, large-scale in situ soil moisture observations are scarce because they are labor-intensive and made at a point scale. Large-scale soil moisture observations are thus best made using satellites [4], [5] equipped with an instrument like an optical sensor, radiometer, scatterometer, or synthetic aperture radar (SAR).

Manuscript received 21 December 2022; revised 21 March 2023; accepted 28 March 2023. Date of publication 5 April 2023; date of current version 17 April 2023. This work was supported by the Fonds National de la Recherche Luxembourg (FNR) under Grant PRIDE15/10623093–HYDRO-CSI. (Corresponding author: Theresa C. van Hateren.)

Theresa C. van Hateren is with the Department of Environmental Research and Innovation, Luxembourg Institute of Science and Technology, 4362 Esch-sur-Alzette, Luxembourg, and also with the Hydrology and Quantitative Water Management Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands (e-mail: tessa.vanhateren@list.lu).

Marco Chini and Patrick Matgen are with the Department of Environmental Research and Innovation, Luxembourg Institute of Science and Technology, 4362 Esch-sur-Alzette, Luxembourg.

Luca Pulvirenti is with the CIMA Research Foundation, 17100 Savona, Italy.

Nazzareno Pierdicca is with the Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, 00185 Rome, Italy.

Adriaan J. Teuling is with the Hydrology and Quantitative Water Management Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands.

Digital Object Identifier 10.1109/LGRS.2023.3264732

Space-borne SAR sensors emit microwaves and measure the fraction of radiation scattered back by the Earth's surface. This backscatter intensity can be used to retrieve soil moisture with a forward model. European Space Agency's (ESA) Sentinel-1 constellation currently provides high spatiotemporal resolution SAR data with a ground-range resolution of 20 m and a six-day revisit time until the failure of the B satellite in December 2021. At these high spatial resolutions, SAR backscatter accuracy is limited by the speckle that is inherent to the data. Speckle causes variations in the backscatter intensity that do not necessarily relate to variations in soil moisture. Multilooking the data reduces the speckle and for that reason, soil moisture products are often presented at relatively low resolutions (500–1000 m) [6], [7], [8].

Multilooking is generally performed on the backscatter data (e.g., [9]), rather than multilooking soil moisture data after its inversion (e.g., [10]). This choice has the advantage of higher computational efficiency. However, applying a multi-look means assuming that the mean of the surrounding pixels is equal to the central pixel of interest [11]. This assumption does not hold when there is a significant spatial variability in soil moisture or other soil parameters that influence backscatter intensity [12]. Over such heterogeneous surfaces, multilooking of the backscatter not only averages speckle and soil moisture conditions, but also other types of information that are contained in the backscatter signal (land cover, vegetation water content, and roughness), and in turn, they cannot be easily removed from the signal as the information on individual pixels is now lost.

Ma et al. [13] have briefly compared the use of high-resolution backscatter data to the use of multilooked backscatter data for soil moisture estimation. Two strategies were applied and compared (Fig. 1): the average-then-calculate (AtC) strategy and the calculate-then-average (CtA) strategy. The AtC strategy consists of multilooking backscatter data to the target resolution and then converting it to soil moisture, whereas the CtA strategy consists of computing soil moisture from high-resolution backscatter data, and then multilooking it to the target resolution. Ma et al. [13] found that the CtA strategy showed significantly better results than the AtC strategy, at a spatial resolution of about 125 000 m. In a related study, Satalino et al. [14] used a synthetic dataset to show that increased accuracy of the CtA strategy occurs, especially when model errors are expected to be large. To the best of our knowledge, the difference between the two aggregation strategies has not been tested on a sub-field scale, even though

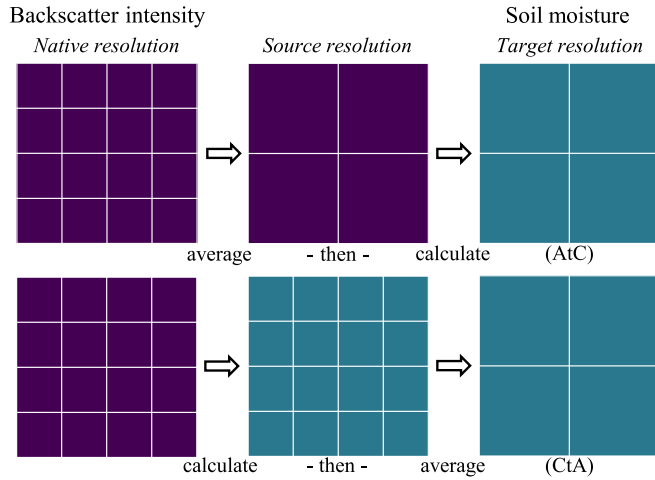


Fig. 1. Illustration of the terms used in this manuscript. (Top) AtC and (bottom) CtA strategies, and native, source, and target resolution. Purple represents backscatter intensity, and blue represents soil moisture.

the processes underlying the effects of multilooking (e.g., heterogeneity) change with changing spatial resolution. The strategies have also not yet been tested with in situ soil moisture data.

We hypothesize that a substantial loss of information occurs in the AtC strategy compared to the CtA strategy. Hence, reducing the spatial source resolution (i.e., the resolution of data going into the retrieval algorithm) to mitigate the speckle, especially at high target resolutions, can negatively impact soil moisture retrieval accuracy. To test this hypothesis, we performed a synthetic experiment with increasing variation in soil moisture content (SMC) and roughness and applied the two strategies to these data. The experiment was duplicated using in situ data to determine whether the results from the synthetic experiment could be confirmed in a field in southeastern Luxembourg [15].

II. DATA AND METHODS

A. In Situ Soil Moisture Data

A field experiment was performed on a 2.5 ha non-irrigated agricultural field in southeastern Luxembourg [15]. The field was sampled on 20 m resolution under different moisture and vegetation conditions. Data were collected in 38 field visits between March 2020 and June 2021, except during the peak of the growing season, because soil moisture retrieval would be unreliable. The observed soil moisture was taken to be the average of five soil moisture measurements taken at each sampling location with a time domain reflectivity (TDR) probe with 5 cm pins. These measurements were calibrated using volumetric soil samples and linear regression.

B. Field Experiment

For the field experiment, satellite soil moisture was retrieved from descending Sentinel-1 (S1) IW ground range detected (GRD) backscatter data with a 33.5° incidence angle over the field. Soil moisture was inverted from these data using the

MULESME algorithm [6], [15]. MULESME is a multitemporal physical pixel-based algorithm that infers soil roughness and soil moisture from backscatter intensity by inverting the Oh forward model [16]. MULESME assumes that SMC changes considerably faster than surface roughness, and any change in backscatter intensity can therefore be attributed to changes in SMC. Since the algorithm is pixel-based, the soil moisture retrieval in a pixel is independent of that in its neighboring pixels.

SMC maps were retrieved from backscatter intensity data at six different resolutions (20, 40, 60, 80, 100, and 120 m) following the two strategies shown in Fig. 1. For both strategies, backscatter images were preprocessed to a square 20 m pixel. For the AtC strategy, these images were then multilooked into the five lower resolutions, and finally, the SMC was retrieved for all six resolutions. Alternatively, for the CtA strategy, the SMC was retrieved at a 20m resolution, and subsequently multilooked into the different target resolutions.

C. Synthetic Experiments

Three synthetic experiments with increasing data variability were set up to study how the computation strategy impacts SMC retrieval accuracy. A square region of 18×18 km (or 360 000 pixels at a 20 m resolution) was used for the synthetic experiments. Synthetic data for the experiments consisted of 15 days of soil moisture, roughness, normalized difference vegetation index (NDVI), land cover, local incidence angle (LIA), and backscatter intensity data. In all three experiments, speckle was added to the synthetic backscatter data, since speckle is inherent to SAR data and does not contain relevant information for soil moisture retrieval. SMC and roughness were varied in time and/or space as shown in Table I. In experiment 1, SMC varied only in time, and roughness was constant in both time and space. Any multilooking, therefore, only removed speckle, and no information on SMC or roughness was lost. In experiment 2, spatial SMC variation was added, so that during multilooking some information on SMC was lost since the resulting aggregated backscatter is an average of different values of SMC. In experiment 3, roughness also varied in space, thereby losing even more information during the multilooking. The synthetic data were used as input data for the MULESME algorithm and processed according to the two different strategies as outlined in Section II-B.

Synthetic data were sampled from either a normal or a uniform distribution, with their parameters derived from in situ (Section II-A) and satellite (Section II-B) field data. Both sets of experiments (i.e., “normal” and “uniform”) were performed, because using a uniform distribution leads to increased spatial variability in soil moisture compared to the normal distribution. However, the uniform distribution is less comparable to a ground truth scenario, where neighboring soil moisture pixel values are usually similar. Hence, the results from the “normal” analysis were compared to the in situ data, and the results from the “uniform” analysis were used to demonstrate a more extreme case. Since random sampling was used, the entire experiment was carried out 15 times with different random values to reduce the chance that the presented results are merely the result of a stochastic artifact.

TABLE I
DATA VARIABILITY IN THREE SYNTHETIC EXPERIMENTS

	Exp.	1	2	3
SMC	Space	constant	varying	varying
	Time	varying	varying	varying
Roughness	Space	constant	constant	varying
	Time	constant	constant	constant
Speckle	Space	varying	varying	varying
	Time	varying	varying	varying

The random sampling is described for every variable in Sections II-C1–II-C4.

1) *Soil Moisture*: For the synthetic SMC dataset based on normal distributions, data were sampled from two truncated normal distributions, one for the temporal variation and one for the spatial variation. The mean of both distributions was set to the spatiotemporal average of the S1 retrieved SMC data over the studied field (i.e., $0.19 \text{ m}^3\text{m}^{-3}$) and the distribution was truncated at the fifth and 95th percentiles of the S1 retrieved SMC data (i.e., 0.08 and $0.29 \text{ m}^3\text{m}^{-3}$). The standard deviation for the spatial variation was taken to be the same as the spatial standard deviation in the in situ SMC data, and for the temporal variation it was taken to be the same as the temporal standard deviation in the in situ SMC data. For the uniform distributions, the lower and upper limits were set to the same as for the normal distribution: 0.08 and $0.29 \text{ m}^3\text{m}^{-3}$.

2) *Roughness*: In the experiments with a constant value for roughness, the spatiotemporal field average roughness was taken from the MULESME runs that were performed for the field study (1.2 cm). In the case of spatially variable roughness, the data were sampled from a normal distribution, based on the same mean, and supplemented with the standard deviation (0.48 cm) and fifth and 95th percentiles (0.50 and 3.8 cm) of the MULESME output. For the uniform distribution, the same percentiles were used as lower and upper limits, respectively.

3) *NDVI, Land Cover, and LIA*: The synthetic NDVI, land cover, and LIA data were set to be constant in both space and time. NDVI was set at 0.15 to mimic near-bare soil conditions, optimal for soil moisture retrieval. The land cover was set to class 211 [non-irrigated arable land in Corine Land Cover (CLC)], which is the same classification as the field in the in situ study. LIA was set to the field average S1 incidence angle (33.5°).

4) *Backscatter Intensity*: The Oh forward model was then used to infer synthetic vertical-vertical (VV) and vertical-horizontal (VH) backscatter from the sampled values of SMC, roughness, and LIA. The noise was added to all three experiments, based on the noise in the real S1 backscatter data over the field. By inferring the noise from real data, we accounted not only for speckle multiplicative noise but also for thermal noise [17].

The noise was added to the synthetic data by multiplying the synthetic linear backscatter at 20 m resolution by a random sample from a truncated normal distribution with a mean of one and a minimum of zero. The standard deviation of the distribution was derived from backscatter data over pixels that showed homogeneous soil moisture conditions: in that case, any remaining variation in backscatter is most likely caused by

speckle. Homogeneous pixels were identified by first selecting days where in situ SMC data had a spatial standard deviation lower than $0.03 \text{ m}^3\text{m}^{-3}$. Second, the 30% of pixels where the in situ SMC was closest to the field average were selected. Satellite backscatter was then extracted over the selected pixels and their spatial standard deviation was computed. Finally, the median of these standard deviations was used as the standard deviation of the speckle distribution.

D. Analysis

For both the synthetic and the field experiment, two performance metrics were used to compare observed and retrieved soil moisture: the Pearson correlation (R), and the unbiased root mean square error (ubRMSE) [18]. The analysis was performed for both retrieval strategies.

In Section III, the Pearson correlation is not shown for the “uniform” synthetic data. In those cases, when moving to lower spatial resolutions, the soil moisture values in the field converge toward the mean of the uniform distribution. Hence, variation in the data decreases with coarser resolutions and the resulting Pearson correlation cannot be fairly compared between different resolutions.

III. RESULTS AND DISCUSSION

The analysis performed on the “normal” synthetic data [Fig. 2(a), top two rows] shows that for all source resolutions and experiments, correlation increases and ubRMSE decreases with a coarser target resolution. Differences in performance were found in the results of the two computation strategies CtA (circles) and AtC (triangles). For every target resolution, the CtA strategy outperforms the AtC strategy. This indicates that retrieving soil moisture at fine resolutions prior to multilooking results in higher retrieval performance at both fine and coarse target resolutions. The CtA performance is especially good when the difference between source resolution and target resolution increases, and peaks at the lowest target resolution.

When moving from experiments 1 to 3, i.e., with increasing spatial variability in the data, the performance variations between the different source resolutions grow (Fig. 3). A general decrease in performance is visible, especially in the correlation. The difference between the performance of the two strategies also increases. In experiment 1, no spatial variation in soil moisture and roughness was simulated. We hypothesized that under these conditions, the difference between the performance of the two strategies would be minimal as any spatial variation in backscatter is only due to speckle, and multilooking to mitigate speckle thus indeed only reduces speckle without losing any other type of information. The results confirm our hypothesis, with small but consistent differences in performance between different resolutions and between the AtC and CtA strategies. In experiment 2, where spatial variation in SMC was added, the performance metrics deteriorated compared to experiment 1. Furthermore, the difference between the two strategies as well as the difference between the six source resolutions increased. In this second experiment, multilooking not only averages the speckle but

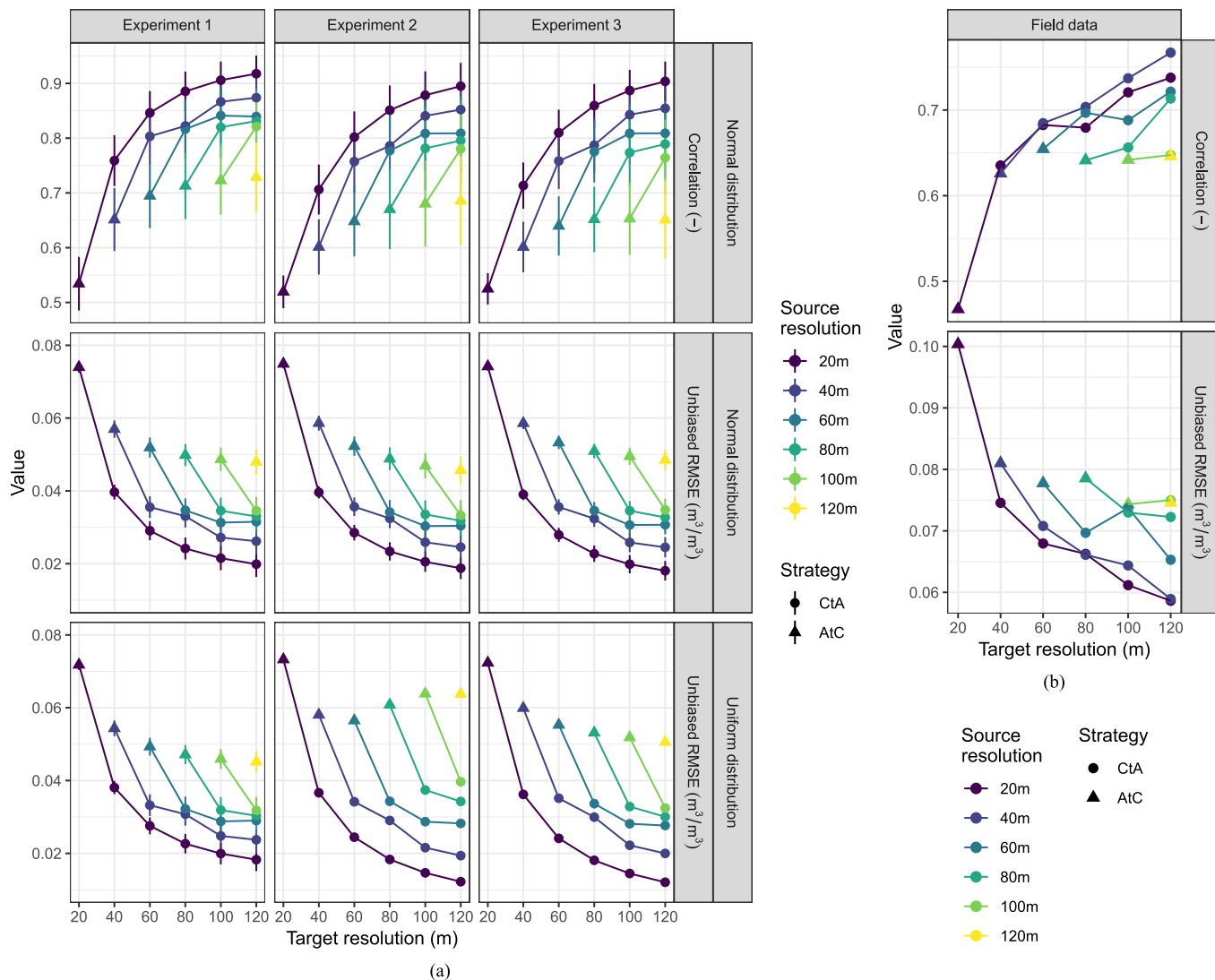


Fig. 2. Resolution- and strategy-dependent accuracy of soil moisture retrieval performance. (a) Mean results and their 95% confidence interval of the 15 synthetic experiments with a normal distribution in the top two rows and the 15 synthetic experiments with a uniform distribution in the bottom row. (b) Results of the two different strategies applied to the field data. In panels where confidence intervals are not visible, their lines are smaller than the diameter of the circles.

also different values of SMC. This trend continued when spatial variation in soil roughness was added to the simulation in experiment 3: the difference in performance further increased, as hypothesized.

Even stronger patterns were found in the “uniform” synthetic data [Fig. 2(a), bottom rows]. Again, for all source resolutions and experiments, ubRMSE decreases with a coarser target resolution, and for each target resolution, the best performance is found for data with high source resolutions. In comparison with the normal distribution, the performance differences between the two strategies are larger, as well as the performance differences between the first two experiments. Interestingly though, the difference between the first and third experiments is smaller than the difference between the first and second experiments, contrary to the experiments based on synthetic data with a normal distribution (Fig. 3).

In this idealized synthetic scenario, it is possible that speckle has a smaller effect on the results than in a field experiment.

An analysis based on the field experiment [Fig. 2(b)] confirms this. However, the patterns found in the synthetic data are still visible in the field experiment data. The best SMC accuracy on any target resolution is obtained when the calculation is performed with backscatter data with a finer source resolution. At the same time, this does not mean that at any target resolution, the highest possible source resolution should be used at resolutions of 80 m or lower, the data with a 40-m source resolution shows better performance than data with a 20 m resolution. This could indicate that at 20 m resolution, the soil moisture signal in the backscatter data was still too weak compared to the speckle to produce consistent soil moisture estimates over the field, and a performance improvement at lower target resolutions was possible by aggregating the source data to 40 m before calculating soil moisture. At high target resolutions (20 or 40 m), using the highest possible source resolution remains the best choice.

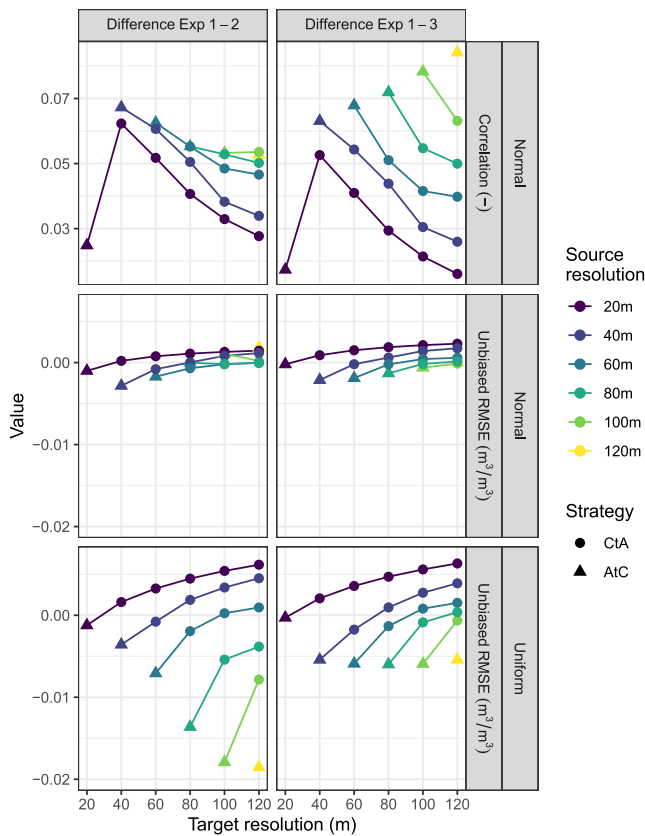


Fig. 3. Impact of increasing spatial variation in the different synthetic experiments. Differences between results from synthetic experiments 1 and 2 (left) and 1 and 3 (right), with a normal distribution in the top two rows and a uniform distribution in the bottom row.

In summary, the results from the synthetic and field experiments confirm our hypothesis: the CtA strategy leads to better retrieval performance than the AtC strategy. Moreover, we saw that in general, when more information is contained in the backscatter data, more information is lost in multilooking and the resulting performance difference between the two computation strategies increases. The exception to this was found in data derived from uniform distributions. In those data the CtA–AtC performance difference was more pronounced in experiment 2 than in experiment 3, indicating that the added spatial variation in roughness does not have as big of an impact on the results as the added spatial variation in SMC, although it could also have been caused by the numerical setup of the synthetic experiment.

IV. CONCLUSION

We performed a synthetic experiment using Sentinel-1 C-band SAR data to test the performance of two retrieval strategies: CtA and AtC. We hypothesized that the AtC strategy leads to the loss of important information on soil moisture conditions, which would mean that applying the CtA strategy would lead to higher accuracies over the area of interest. Our results showed that, indeed, applying the CtA strategy to native resolution (20 m) Sentinel-1 data led to a smaller ubRMSE and a higher correlation on all tested target resolutions (20–120 m).

The results from the synthetic experiment were confirmed in a 2.5 ha field in southeastern Luxembourg that was intensively sampled for in situ soil moisture conditions. Since in both the synthetic and the field experiment an increase in performance was found even at small resolution gaps between source and target data, we expect that the presented results are also relevant for coarser resolutions and for soil moisture applications on larger scales.

REFERENCES

- [1] S. I. Seneviratne et al., “Investigating soil moisture-climate interactions in a changing climate: A review,” *Earth-Sci. Rev.*, vol. 99, nos. 3–4, pp. 125–161, 2010.
- [2] A. F. Van Loon, “Hydrological drought explained,” *WIREs Water*, vol. 2, no. 4, pp. 359–392, Jul. 2015.
- [3] M. G. Grillakis, A. G. Koutroulis, J. Komma, I. K. Tsanis, W. Wagner, and G. Blöschl, “Initial soil moisture effects on flash flood generation—A comparison between basins of contrasting hydro-climatic conditions,” *J. Hydrol.*, vol. 541, pp. 206–217, Oct. 2016.
- [4] L. Brocca, L. Ciabatta, C. Massari, S. Camici, and A. Tarpanelli, “Soil moisture for hydrological applications: Open questions and new opportunities,” *Water*, vol. 9, no. 2, p. 140, Feb. 2017.
- [5] Z.-L. Li, P. Leng, C. Zhou, K.-S. Chen, F.-C. Zhou, and G.-F. Shang, “Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future,” *Earth-Sci. Rev.*, vol. 218, Jul. 2021, Art. no. 103673.
- [6] L. Pulvirenti et al., “A surface soil moisture mapping service at national (Italian) scale based on Sentinel-1 data,” *Environ. Model. Softw.*, vol. 102, pp. 13–28, Apr. 2018.
- [7] B. Bauer-Marschallinger et al., “Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 520–539, Jan. 2019.
- [8] L. Zappa, S. Schlaffer, B. Bauer-Marschallinger, C. Nendel, B. Zimmerman, and W. Dorigo, “Detection and quantification of irrigation water amounts at 500 m using Sentinel-1 surface soil moisture,” *Remote Sens.*, vol. 13, no. 9, p. 1727, Apr. 2021.
- [9] N. N. Das et al., “The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product,” *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111380.
- [10] A. Balenzano et al., “Sentinel-1 soil moisture at 1 km resolution: A validation study,” *Remote Sens. Environ.*, vol. 263, Sep. 2021, Art. no. 112554.
- [11] M. Mansourpour, M. Rajabi, and J. Blais, “Effects and performance of speckle noise reduction filters on active radar and SAR images,” in *Proc. ISPRS*, vol. 36, Ankara, Turkey, 2006, p. W41.
- [12] A. J. Teuling, “Improved understanding of soil moisture variability dynamics,” *Geophys. Res. Lett.*, vol. 32, no. 5, 2005, Art. no. L05404.
- [13] C. Ma, X. Li, and M. F. McCabe, “Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data,” *Remote Sens.*, vol. 12, no. 14, p. 2303, Jul. 2020.
- [14] G. Satalino, G. Pasquariello, F. Mattia, and L. Dente, “On the accuracy of soil moisture content retrieved, at pixel, segment or field scale, from advanced-SAR data—A simulation study,” in *Proc. IEEE Int. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2004, pp. 3532–3535.
- [15] T. C. Van Hateren, M. Chini, P. Matgen, L. Pulvirenti, N. Pierdicca, and A. J. Teuling, “Optimal spatial resolution of Sentinel-1 surface soil moisture evaluated using intensive in situ observations,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 6311–6314.
- [16] Y. Oh, “Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 596–601, Mar. 2004.
- [17] D. Palmisano, F. Mattia, A. Balenzano, G. Satalino, N. Pierdicca, and A. V. M. Guarnieri, “Sentinel-1 sensitivity to soil moisture at high incidence angle and the impact on retrieval over seasonal crops,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7308–7321, Sep. 2021.
- [18] D. Entekhabi, R. Reichle, R. D. Koster, and W. Crow, “Performance metrics for soil moisture retrievals and application requirements,” *J. Hydrometeorol.*, vol. 11, pp. 832–840, Jun. 2010.