International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105011

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

AI in soil moisture remote sensing

Carsten Montzka[a,*], Luca Brocca[b], Hao Chen[c,a], Narendra N. Das[d], Antara Dasgupta[e], Mehdi Rahmati[a], Thomas Jagdhuber[f,g]

[a] Institute of Bio- and Geosciences: Agrosphere (IBG-3), Forschungszentrum Jülich, 52425 Jülich, Germany
[b] Research Institute for Geo-Hydrological Protection, National Research Council of Italy, 06128 Perugia, Italy
[c] Faculty of Geography, Tianjin Normal University, 300387 Tianjin, China
[d] Biosystems & Agricultural Engineering and Civil & Environmental Engineering, Michigan State University, 48824 East Lansing, USA
[e] Institute of Hydraulic Engineering and Water Resources Management, RWTH Aachen University, 52062 Aachen, Germany
[f] Microwaves and Radar Institute (HR), German Aerospace Center (DLR), 82234 Wessling, Germany
[g] Institute of Geography, University of Augsburg, 86159 Augsburg, Germany

ARTICLE INFO

Keywords:
Soil moisture
Artificial intelligence
Machine learning
Deep learning
Remote sensing
Microwave

ABSTRACT

Soil moisture, a pivotal component of the hydrological cycle, exerts a profound influence on land surface exchange processes, but its spatial variability poses challenges for large-scale field observations, increasing reliance on satellite-based retrievals. However, spaceborne estimates face limitations due to model uncertainties and sensor-related constraints. Recent advances in artificial intelligence (AI) offer promising alternatives to traditional methods by enabling data-driven estimation of soil moisture without strong physical assumptions. Thus, a critical review of emerging AI-based soil moisture retrieval methods with respect to their advantages and disadvantages is vital to ensure the best utilization of such tools for soil moisture sensing, especially with novel sensors and data constantly being generated.

In this comprehensive review, we furnish the first structured overview of AI methods and their applications in soil moisture retrievals from remote sensing. AI is able to enhance soil moisture retrieval by learning complex (highly nonlinear) relationships between satellite observations and ground reference data, to support time series reconstruction by filling gaps in data sets, to estimate subsurface soil moisture conditions from surface signals and auxiliary inputs, to enable spatial scaling by translating soil moisture estimates across different resolutions using multi-resolution data, to predict temporal dynamics as a soil moisture forecast, and to contribute to broader assessments of the water cycle and beyond by integrating soil moisture with further hydrological variables. Future directions for each method are also identified to address the scientific challenges of soil moisture retrieval and help focus the research community on the key open questions in the new era of rapidly expanding AI applications.

1. Introduction

Soil moisture (SM) refers to the volumetric or gravimetric moisture contained in the unsaturated soil zone, which plays a critical role in land–atmosphere feedbacks at both weather and climate scales (Li et al., 2021). Due to its pivotal role in partitioning the incoming radiation flux over land into latent and sensible heat fluxes to the atmosphere, as well as dividing precipitation into surface and subsurface flows, soil moisture is recognized by the World Meteorological Organisation as an Essential Climate Variable (GCOS, 2022). SM also strongly influences the climate-vegetation feedback, and thereby has substantial impacts on agricultural productivity, forestry, and ecosystem health as well as on extreme climate events (Daly and Porporato, 2005; Green et al., 2019; Humphrey et al., 2021). Despite its critical importance for the water, energy, carbon or nutrient cycles, the inherently large spatial and temporal variability in SM, inhibits its estimation from traditional field-based measurements in a scalable fashion. Satellites provide a transformative alternative to ground-monitoring techniques and are able to monitor SM spatiotemporal heterogeneity from local to global scales, and from short- to long-term dynamics. Indeed, spaceborne SM remote sensing has a long history dating back to the late 1970 s (Dorigo et al., 2017), offering an important record to investigate changes in the climate system across

* Corresponding author.
E-mail address: c.montzka@fz-juelich.de (C. Montzka).

https://doi.org/10.1016/j.jag.2025.105011
Received 2 October 2024; Received in revised form 14 November 2025; Accepted 6 December 2025
Available online 11 December 2025

different space–time scales and the occurrence of extreme events (Berg and Sheffield, 2018; Seneviratne et al., 2010). Recent advances in SM retrieval and processing from satellite data have led to notable gains in accuracy, spatial representation, and thus possible applications, but several challenges remain (Babaeian et al., 2019).

Active and passive microwave methods have been utilized for SM remote sensing (Mohanty et al., 2017). Besides few longer wavelength microwave sensors (Etminan et al., 2020; Fluhrer et al., 2022; Shen et al., 2021; Yueh et al., 2020), SM remote sensing is typically related to the top few centimeters (~0–5 cm) of soil (Escorihuela et al., 2010). However, most applications require information about the state of the whole water in the plant available soil region, i.e. in the root zone. This is necessary, for example, to identify agricultural drought conditions, determine when plants begin to experience stress, and inform irrigation decisions (Bolten et al., 2010). As a first challenge, root zone SM can be estimated from surface SM by an exponential filter, considering an empirical parameter accounting for infiltration and temporal storage capacity (Kornelsen and Coulibaly, 2014; Stefan et al., 2021; Wagner et al., 1999). This can be achieved under the assumption of a hydrologic equilibrium within the soil profile and the availability of soil physical properties (Mishra et al., 2020). Pasik et al. (2023) extend the exponential filter by structural uncertainty estimations to provide a global root zone SM data set. With further assumptions and the Green-Ampt approach (widely used in hydrology for estimating the rate of water infiltration into soil, see also Rawls et al. (1983)), Manfreda et al. (2014) developed a SM analytical relationship to retrieve root zone SM. As such, data assimilation is able to provide root zone SM as an unobserved variable in the water cycle by merging the observations with hydrological models (Das and Mohanty, 2006; Mladenova et al., 2020; Qiu et al., 2021; Tobin et al., 2019).

Although passive microwave sensors allow almost daily high accuracy global SM products at tens of kilometers, the coarse spatial resolution is not adequate for many potential applications such as agricultural management (Peng et al., 2021; Sabaghy et al., 2018). To address this second challenge of spatial resolution, downscaling methods have been developed to closer match the user's needs, making use of additional observations at higher resolution mainly in the optical, thermal, and active microwave domain (Peng et al., 2017). In the thermal domain, studies make use of the time-dependent response of an object to temperature variation. The volumetric heat capacity of a soil increases with SM, and the higher water content corresponds to a smaller temperature variation (Fang et al., 2018). Most approaches nowadays utilize data from both optical and thermal spectral regions to account for soil temperatures at dry and wet conditions and the fractional vegetation cover to consider soil evaporation or vegetation transpiration dominance (Fang et al., 2022a; Merlin et al., 2008; Ojha et al., 2021; Zhao et al., 2021; Zheng et al., 2021). Passive-active downscaling considers high resolution backscatter from Synthetic Aperture Radar (SAR) systems (Bauer-Marschallinger et al., 2018), where challenges due to different microwave frequencies (e.g., passive L-band and active C-band) occur (Das et al., 2019; He et al., 2018; Li et al., 2018; Meyer et al., 2022). Data assimilation is another method to downscale SM, but requires a hydrological model typically implemented in a sequential prediction and analysis framework (Lievens et al., 2017; Naz et al., 2020; Naz et al., 2019).

For SM forecasting for the next days, weeks or even seasons, observations need to be accompanied by methods able to move the current state (e.g., SM estimates) towards future conditions. To address this third challenge, Earth system models are commonly employed, often in combination with data assimilation schemes. These models predict soil moisture dynamics based on physical land surface process formulations and are driven by factors such as precipitation and evapotranspiration (Oleson et al., 2008). Esit et al. (2021) utilize the Community Earth System Modeling (CESM) platform to perform seasonal to decadal SM predictions. Boas et al. (2023) were able to perform a seasonal SM and crop yield forecast with the Community Land Model (CLM), but found evidence to specifically account for drought and other stressors in the crop yield simulations. Vogel et al. (2021) use the gridded water balance model AWRA-L to generate not only SM forecasts but also evapotranspiration and runoff predictions for time periods of one to three months in future. They stress their importance for agricultural and reservoir management as well as to provide a bushfire risk assessment.

As shown above, these three exemplified challenges have been addressed by statistical approaches and physical models with impressive results. Further challenges exist, e.g. identifying interferences or bad data records, the initial SM retrieval, filling gaps in the time series or spatial domain, or estimating further Earth system variables. However, the applied methods may have drawbacks like missing important processes in their formulation, the inability to account for extreme events, or high computational demand. Here, fully data-driven artificial intelligence (AI) methods are able to provide another perspective on the retrieval and processing of SM data.

AI methods are being used in Earth sciences with increasing frequency. They use a bottom-up approach in which algorithms learn relationships between input data and output results as an important step in a model-building effort (Maskey et al., 2020). By doing so, irrespective of physical laws and simplifying assumptions, AI can discover patterns and trends buried within vast volumes of data that are not apparent to human analysts, who look at the data through the lens of their physical understanding. The huge amount of data recorded by remote sensors suitable to estimate and improve SM representations in combination with the emergence of advanced AI methods and the easy access to computing resources triggered the flood of related studies in the last decade. We need to discuss to which extent AI shall accompany or even substitute physical or statistical approaches and which benefits and drawbacks we can expect. Moreover, we need to investigate if a methodological transformation is possible in which AI and remote sensing observations are reassembled as a learning, self-validating and interpretable hybrid system (Irrgang et al., 2021).

With this review we aim at an informed overview of AI applications in the field of SM remote sensing. First, we provide the general background necessary to understand the traditional SM remote sensing and processing methods, as well as a brief introduction into AI methods (Section 2). This is followed by a structured overview over AI applications in SM remote sensing in Section 3, specifically in time series reconstruction, root zone SM estimation, spatial scaling, and forecasting. Moreover, we assess literature and identify the potential for improving the understanding of further water cycle components and linked processes such as in the energy and carbon cycles. Future perspectives are provided to identify avenues for further promising research topics in Section 4.

2. Background

2.1. Basic principles of microwave-based soil moisture remote sensing

SM is observed from space mainly in the microwave region of the electromagnetic spectrum (Mohanty et al., 2017). It affects the dielectric properties of the soil as a mixture of solid components, water and air, which, in turn, influence the emission and scattering of microwave radiation (Dobson et al., 1985). The relationship between dielectric properties and SM is described by so-called dielectric mixing models, where further information about soil texture or organic carbon content may be needed to constrain the predictions (Mironov et al., 2017; Park et al., 2019; Topp et al., 1980; Wang and Schmugge, 1980).

Microwave radiometry (passive) records the natural thermal emission from the land surface in terms of brightness temperature in the frequency range of 0.3 to 300 GHz, which is related to the surface emissivity and the effective microwave surface temperature (close to thermal infrared measurements). The emissivity is further influenced by the incidence angle, the soil reflectivity, the surface roughness, and multiple scattering within the soil and covering vegetation. At low frequencies (0.3 – 2 GHz), the vegetation impact can be well approximated by simple radiative transfer models (Wigneron et al., 2007). Here, the vegetation opacity for that microwave frequency needs to be accounted for (Wegmuller and C.M., E.G. Njoku, 1995). As the emission from land surface is very low, radiometers need to integrate over large regions to get adequate signal-to-noise ratios, leading to coarse spatial resolutions of tens of kilometers. For many applications this is too coarse and requires downscaling (Peng et al., 2017). Common satellite missions utilizing this observation concept are AMSR (Advanced Microwave Scanning Radiometer, two satellites AMSR-E and AMSR2, JAXA (2013); Njoku et al. (2003)), SMOS (Soil Moisture and Ocean Salinity, Kerr et al. (2010)), and SMAP (Soil Moisture Active and Passive, Entekhabi et al. (2010)).

Microwave scatterometry (active) makes use of radar systems at low spatial resolution, initially designed for ocean observations. Backscatter is the intensity of the returned microwave signal, derived from reflected microwave energy and often expressed relative to a baseline or calibration reference. This measure is influenced by soil surface roughness, vegetation and the observation angle. The main representative for this sensor group is ASCAT (Advanced Scatterometer, Wagner et al. (2013)) on the MetOp (Meteorological Operational Satellite) missions, implementing a change detection method for the retrieval of SM. ASCAT's three observation incidence angles on each side of the azimuth direction (six beams in total) are normalized to a 40° reference angle. Here, also the seasonal variation in vegetation is accounted for (Wagner et al., 1999). Long-term minimum (dry) and maximum (saturated) scattering values define the SM expressed as a saturation degree, which can be transferred to volumetric SM by multiplying with the soil porosity. However, in several ASCAT products the Soil Water Index or Soil Wetness Index (SWI) is provided as a result of the already mentioned exponential filter and denotes the profile soil saturation, see, e.g., Brocca et al. (2010) or Paulik et al. (2014). Brocca et al. (2017) reviewed the applications of ASCAT SM and provided a roadmap for future research activities.

Active radar systems using the Synthetic Aperture (SAR) principle are based on moving platforms like satellites and employ the distance they travel between sending and receiving the microwave signal as a large (synthetic) antenna. This allows high resolution data at the meter scale. External influences generally have a stronger impact on SAR than on scatterometry signals. Due to its high spatial resolution, SAR is more sensitive to viewing geometry effects (e.g., layover, foreshortening, shadowing) and fine-scale surface heterogeneity such as land cover, vegetation, and agricultural practices, whereas scatterometry averages these effects over much larger footprints. Methods to gain SM from SAR backscatter are various, ranging from full physical modeling of the scattering components towards empirical relationships or time series analyses. However, complex physical models require the input of a large number of parameters not easy to obtain at larger scales, and, as such, semi-empirical and change analysis models have generally been the most favored. Bauer-Marschallinger et al. (2019) used the ASCAT change detection method to retrieve the SWI at 1 km from Sentinel-1. The method to obtain surface SM from Sentinel-1 has been recently improved by Quast et al. (2023) by using a new bistatic radiative transfer modeling framework. Balenzano et al. (2021) developed an alternative approach calculating the ratio of two consecutive co-polarized backscattering signals to solve an underdetermined linear equation system for volumetric SM. Mengen et al. (2023) extended the co-polarized backscatter approach to account for incidence angle and vegetation changes. Vegetation can be considered explicitly as a cloud of water droplets because vegetation water is the main reason for scattering impact. Therefore, the Water Cloud Model has widely been implemented, e.g., Ayari et al. (2022), Weiss et al. (2020), Zribi et al. (2019). To explicitly simulate the soil scattering components, the Oh (Oh, 2004) and Dubois (Dubois et al., 1995) models have been developed, fitting functions according to the polarization (ratios) to retrieve SM. The Integral Equation Model (IEM, Fung and Chen (2004)) has been developed for a range of soil surface characteristics.

A special type of SM retrieval is Global Navigation Satellite System (GNSS) Reflectometry (GNSS-R). It makes use of signals of opportunity from navigation satellites, utilizing them as transmitters to capture reflected signals from the Earth's surface (Yang et al., 2024). This innovative technique requires only a single, small, and low-cost receiver to detect these reflections, making it a cost-effective and versatile tool for Earth observation. By analyzing the reflected signals, valuable information about surface properties, including SM, can be extracted (Rahmani et al., 2022), e.g. from NASA's Cyclone GNSS (CYGNSS) constellation (Yan et al., 2020). As GNSS operate mainly in L-band, the same physical principles as stated before apply, in addition making use of the different scattering mechanisms between a direct measurement and a ground reflected measurement. Also here, challenges remain, e.g. regarding observation angles and coherence scattering, see Wu et al. (2021a) and Edokossi et al. (2020) for further details.

Common frequency bands for SM retrieval include L-band (1–2 GHz), C-band (4–8 GHz), and X-band (8–12.5 GHz). In recent years also P-band (0.23–1 GHz) gains interest (Alemohammad et al., 2019; Fluhrer et al., 2022). The different wavelengths have diverse advantages and drawbacks, for end users important is mainly the penetration depth into the soil. The longer the wavelength, the deeper the microwaves can enter, and from deeper regions the emissions evolve, respectively. As the sensing depth is dependent on SM, soil texture, organic carbon content etc., often rule of thumb estimates for typical soils with low vegetation cover are used. Where at X-band the surface soil is sensed only, at C-band the upper 1 cm, at L-band the upper 5 cm, and at P-band the upper 20–30 cm are observed (El Hajj et al., 2019; Escorihuela et al., 2010; Fluhrer et al., 2022; Nolan and Fatland, 2003). Penetration through covering vegetation or sensitivity to soil surface roughness, also increase and decrease, respectively, with increasing wavelength.

A large variety of microwave-based satellite SM products have been released, spanning a time period from late 1970 s to today with global coverage. Please refer to Liu and Yang (2022); Mohanty et al. (2017); Petropoulos et al. (2015), a list of products is also provided in Montzka et al. (2020). Also, methods exist utilizing optical methods, e.g., spanning a two-dimensional (triangular or trapezoid) feature space between land surface temperature (LST) and vegetation indices as SM proxies. However, due to brevity these are not covered in this section, please see Zhang and Zhou (2016) for further reading.

From these general SM retrieval concepts based on different types of microwave observations utilizing various radiative transfer, dielectric mixing, and SAR scattering models with partly requiring auxiliary data such as soil texture, vegetation opacity or vegetation water content estimates, it has to be stressed that the relationships between satellite observations and SM can be very complex. Here, AI is able to mimic these links, to identify shortcuts in the data structure, and to speed-up the processing of the retrievals once a model is trained. Assumptions and simplifications in the physical or conceptual observation-SM modeling may be obsolete, but, however, new assumptions related to the AI procedure may substitute the original ones. In addition, including further features related to the affecting factors in the AI system may better constrain the relationships. Physically meaningful covariates may improve the generalizability and mitigate ambiguities (e.g., separating vegetation effects from SM). Moreover, they may enable hybrid modeling that utilizes both physical and data-driven insights. Table 1 provides a list of features affecting SM spatio-temporal patterns as well as having an impact on microwave remote sensing of SM, potentially useful for AI approaches to estimate SM.

2.2. *Artificial intelligence methods and concepts*

Recent concurrent advances in the availability and accessibility of remote sensing big data and cloud computing resources, have massively accelerated the pace of developments in AI. AI generally refers to machines performing tasks which typically require human intelligence. For

Table 1
Covariates altering SM and microwave sensor characteristics.

Domain	Property	Co-variate	Explanation
SM Dynamics	Soil	Soil Texture	Sand, Silt, Clay Content determine infiltration rate, retention capacity, hydraulic conductivity
		Bulk Density	Influence water holding capacity and movement through soil
		Soil Organic Carbon Content	Increases water retention and influences hydraulic conductivity
	Vegetation	Type	Control transpiration and canopy interception
		Density	Control transpiration and canopy interception
		Structure	Controlling canopy interception, shading, root water uptake, together regulate soil water retention and evapotranspiration
		Photosynthetic Activity	Driving transpiration, extracts water from the soil and influences its temporal and spatial distribution
	Terrain	Slope	Regulate runoff, drainage
		Aspect	Solar radiation exposure
		Elevation	Control on climate, hydrology, and soil formation processes
	Meteorology	Precipitation	Primary input to soil moisture
		Temperature, Solar radiation, Wind speed	Drive evapotranspiration and drying processes
		Relative Humidity, Vapor Pressure Deficit	Influences evaporation rates
	Hydrology	Groundwater Table Depth	Affects capillary rise and soil moisture persistence
		Spatial Connectivity	Affects 3D soil moisture structure, e.g. interflow
	Human Interaction	Irrigation	Directly affecting input
		Impervious Surfaces	Hinders water exchange with surface
		Land Management	Impact infiltration and soil compaction
Remote Sensing	Soil	Surface Roughness	Modifies microwave scattering; rougher surfaces increase backscatter and reduce sensitivity to soil moisture; Rough surfaces have larger area than smooth ones with larger emissivity
		Dielectric properties	Strongly affect microwave emissivity and reflectivity
		Temperature	Influences thermal emission in passive microwave measurements
		Salinity	Alters dielectric constants and therefore signal response
	Vegetation	Structure	Causes attenuation and scattering of microwave signals
		Water Content	Attenuates and scatters microwave signals, reducing their sensitivity to underlying soil moisture
		Temperature	Affecting thermal radiation component, altering the observed brightness temperature in passive microwave
	Sensor	Incidence Angle	Affects signal sensitivity and backscatter intensity, larger angles generally reducing soil moisture sensitivity and increasing surface scattering effects
		Azimuth Angle	Directional dependence of surface roughness and vegetation structure, leading to variations in signal with viewing geometry, e.g. in periodic structures
		Microwave Frequency	Determines the penetration depth and sensitivity of the signal to soil moisture
		Polarization	Determining sensitivity to surface roughness, soil moisture, vegetation structure
		NESZ (Noise-Equivalent Sigma-Zero) Level	Controls the noise floor relative to signal strength, with lower NESZ enabling clearer detection of weak scatterers and improved overall image quality
		Radiometric Sensitivity	Influencing the ability to distinguish targets or subtle surface changes
		Geometric Sensitivity	Determines how precisely features are located and shaped on the ground, impacting positional correctness and distortion levels
	Atmosphere	Water Vapor	May attenuate signals, particularly at higher microwave frequencies

instance, if a machine, i.e. an automated algorithm is used to interpret satellite data previously visually interpreted by a human being, this would be a use of AI. Further, if the input features are manually engineered prior to the application of AI, i.e. the step of expert feature engineering is included, the application falls under machine learning (ML). ML algorithms such as *Artificial Neural Networks (ANNs)* are not new to hydrology or geoscience and have been used for decades in a variety of applications (Dramsch, 2020; Govindaraju and Artific, 2000; Sun et al., 2022; Zhang et al., 2022a), e.g., to invert SM from remote sensing data (Rodríguez-Fernández et al., 2015).

If the relevant features for a given classification/regression are auto derived i.e. learnt from the data internally within the AI algorithm using a deep neural network (>2 hidden layers), the approach is then known as deep learning (DL). DL has emerged as a breakthrough technology in remote sensing (Ma et al., 2019) with the rise of big data in Earth Observation and has quickly become an extremely powerful tool in solving the grand challenges (Zhu et al., 2017).

It is important to note that AI is a broad field, and not all AI applications rely on ML or DL. Examples of non-ML AI methods are expert systems, where retrieval rules are explicitly encoded in if-then or thresholding approaches, e.g., if backscatter is below threshold X and land cover is Y, then SM is low. A similar domain is fuzzy logic systems to handle uncertainty in noisy remote sensing data with vague rules. Evolutionary or genetic algorithms can be applied to tune parameters of physical models without training on data in the ML sense. These are AI search strategies through possible parameter combinations to minimize (or maximize) an objective function, rather than statistical learning. Although inherently not an AI method, data assimilation, i.e. the combination of predictive models with real-world observations to estimate the true state of a system, shows non-ML AI characteristics by the underlying Bayes theorem. Starting with a prior estimate from the model, assess the likelihood of the observed data given that state, and update to a posterior estimate that reflects both sources of information. In this way, data assimilation performs intelligent reasoning under uncertainty without depending on pattern-learning methods. While not ML in the modern sense, these are forms of algorithmic intelligence for combining uncertain data with prior knowledge. An introduction to AI which is not ML or DL is given by Russell and Norvig (2016). Most current SM remote sensing applications in AI are dominated by ML methods, primarily since DL methods require substantially more data for training given the huge parameter space (to the order of millions of parameters). In this limited data context, ML approaches typically generalize much better, due to their relatively simpler functional forms with a substantially lower volume of parameters. Additionally, mixed approaches are also evolving where non-ML AI methods like genetic algorithms are used to optimize ML models to refine the parameter estimation process further, see e.g., Srivastava et al. (2021) or Kara et al. (2024). However, with the growing volume of satellite and ground observations of soil moisture, the shift towards DL is imminent because they can model the complex, nonlinear relationships between the variables.

Both ML and DL algorithms can be implemented as supervised (learning from labeled targets) or unsupervised (pattern recognition without labels) methods, depending on the availability of labeled training data. A labeled dataset in this context refers to a collection of data instances where each example is paired with a corresponding label or annotation that indicates the reference or the desired output. The label is essentially the correct answer or the category to which the data point belongs. Specifically in the context of remote sensing applications, a majority of the applications have so far been supervised, thus, fueling a huge demand recently for labeled training datasets for a variety of geophysical applications. Data sets are typically divided into three subsets: training, validation, and testing. The training set is used to teach the model by adjusting its internal parameters based on input–output patterns. The validation set is used during model development to fine-tune hyperparameters and prevent overfitting, providing an unbiased evaluation of model performance during training. Finally, the test set is held back until the end and used to assess how well the trained model generalizes to completely unseen data. A common split is 70 % for training, 15 % for validation, and 15 % for testing, though this can vary depending on the dataset size and problem. This section presents the main ML and DL algorithms used in the context of SM retrievals from satellites, and Table 2 summarizes the advantages and drawbacks of the main AI methods.

2.2.1. *Machine learning algorithms*

ANNs are simple feedforward neural networks in which sets of neurons are organized in layers, with each computing a weighted sum of its inputs (Fig. 1). The intermediate layers disconnected from the environment (i.e. not accepting inputs or delivering outputs), are known as hidden layers. However, with enough hidden layers such neural networks can theoretically estimate any function given adequate amounts of data, the lack of which can easily lead to overfitting to the training data and delivering biased estimators which do not generalize across cases. Thus, ANNs can learn complex, non-linear, spatial relationships between datasets given enough input and labeled output pairs, and are typically robust to noisy data. *Support Vector Machines (SVMs),* which are kernel-based classification and regression algorithms, which have also been relatively successful in soil moisture estimation from remote sensing data (Ahmad et al., 2010). The goal of SVMs is to fit a hyperplane in an N-dimensional feature space which maximizes the margin between the data points and the decision boundary. SVMs are known to generalize better due to their reliance on the support vectors, or the data points nearest to the decision boundary which influence the position and orientation of the hyperplane, which helps reduce the risk of overfitting to the training data (Fig. 1). More recently, *Decision Tree (DT)* algorithms with its nodes representing a decision based on a specific feature, have been extended to satellite-based soil moisture retrievals. DT algorithms learn or develop threshold-based soil moisture condition rules about the behavior of each input feature towards the outcome, and are thus more robust in certain applications. However, their simplicity makes them prone to overfitting and less effective in capturing complex spatial or temporal relationships in SM dynamics. The *Random Forest algorithms (RF),* on the other hand, create an ensemble of DTs, each trained on a random subset of the dataset and using a random subset of features for each split. This randomness helps in reducing overfitting and improving the model's generalization performance (Fig. 1). RFs are widely used in SM remote sensing for their robustness and ability to handle nonlinear relationships and high-dimensional satellite data. Despite not modeling spatial or temporal dependence explicitly, they often perform well with engineered features (e.g. vegetation indices and surface temperature). Similarly, *Extreme Gradient Boosting (XGBoost)* is a powerful and widely utilized ML algorithm known for its exceptional predictive performance. Based on the ensemble learning concept of boosting, XGBoost combines the strengths of multiple DTs to create a robust and accurate predictive model, through regularization techniques which control the complexity of individual trees and prevent overfitting. XGBoost has been applied in SM retrieval due to its high accuracy and efficiency on structured remote sensing data. It works well with tabular inputs like vegetation indices and soil properties but lacks spatiotemporal context unless features are manually crafted.

Table 2
Most popular machine learning methods and their advantages and drawbacks including adequacy for spatiotemporal SM applications.

Model	Pros	Cons	Applicability in SM field
Decision Trees (DTs)	Simple to interpret, robust to noise	Can overfit the training data, can be inaccurate for complex problems	Limited for spatiotemporal SM data unless spatial and/or temporal features are explicitly engineered
Random Forest (RF)	Accurate, robust, flexible, interpretable	Can be computationally expensive to train	Moderate if spatiotemporal SM features are well-crafted, otherwise limited
Support Vector Machines (SVMs)	Able to handle nonlinear relationships and high-dimensional data	Can be computationally expensive to train	Poor for raw spatiotemporal SM data, features need to be engineered explicitly
Artificial Neural Networks (ANNs)	Able to learn complex nonlinear relationships between input and output data	Can be difficult to train and interpret, can be overfitting-prone	Weak for raw spatiotemporal SM data unless structured inputs are manually included
Convolutional Neural Networks (CNNs)	Able to learn multi-scale spatial relationships in data, closest to human interpretation	Require large amounts of data for training, do not generalize well on unseen distributions	Good for spatial SM data, moderate for spatiotemporal if extended properly
Generative Adversarial Networks (GANs)	Able to generate realistic super-resolution, could help in SM downscaling	Computationally expensive to train	Strong for spatial SM data, suitable for spatiotemporal with extensions
Graph Neural Networks (GNNs)	Able to retain geometric interrelationships in the data, could help in reducing data needs by encoding known inductive biases	Relatively time consuming also in the inference phase depending on the complexity of the mesh/graph	Very good for spatial and spatiotemporal SM data when used with temporal layers
Recurrent Neural Networks (RNNs)	Suitable for short-term predictive modeling	Results in vanishing gradients for time series longer than ten timesteps	Good for temporal SM data, limited for spatiotemporal unless combined with spatial methods
Long Short-term Memory (LSTM) networks	Well-suited for modeling sequential data and long time series, effective for predicting future SM	Can be computationally expensive to train	Strong for temporal SM data, effective for spatiotemporal when hybridized
Extreme Gradient Boosting (XGBoost)	Accurate, efficient, scalable	Can be difficult to tune hyperparameters	Moderate with good feature engineering, not ideal for raw spatiotemporal SM structure
Transformers	Allow representing input importances through attention mechanisms	Computationally expensive and data hungry	Excellent for temporal and increasingly strong for spatiotemporal SM tasks
Physics-Informed Neural Networks (PINNs)	Allow embedding causality and physical laws	Can struggle to converge effectively under input/target uncertainties	Highly effective for spatiotemporal SM dynamics with physical constraints

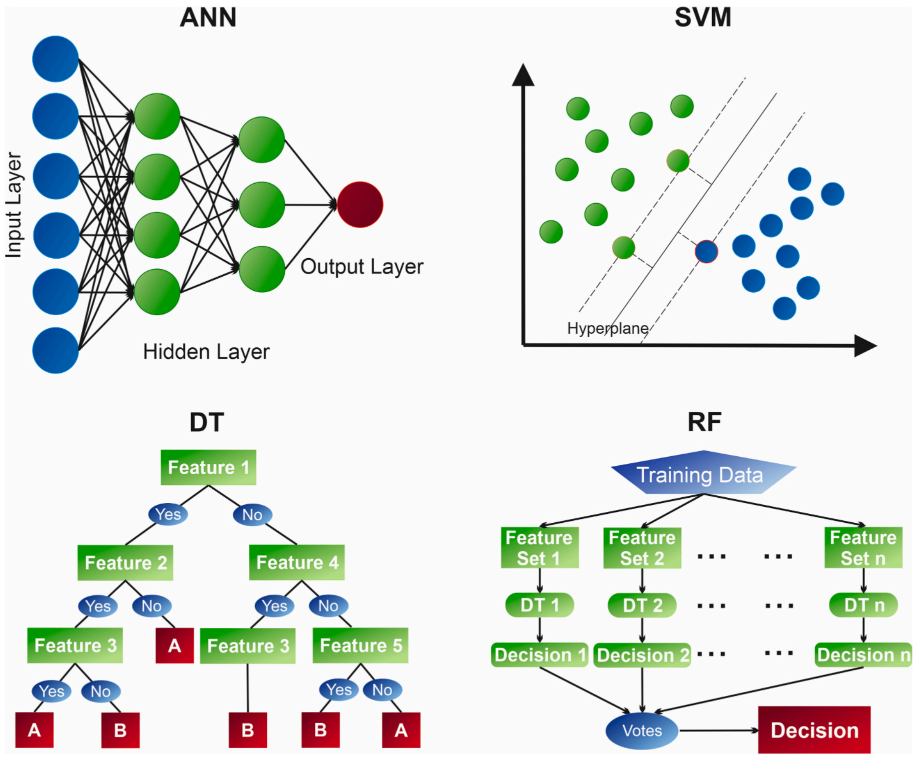

Fig. 1. Basic machine learning methods: Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF).

2.2.2. Deep learning algorithms

One of the most popular DL algorithms in remote sensing are the *Convolutional Neural Networks (CNNs)* (Le Cun, 1990), which can be viewed as adaptations of ANNs for 2-dimensional image data (Fig. 2). CNNs typically comprise three types of layers, including convolutional layers, pooling layers, and fully-connected layers (O'Shea and Nash, 2015). The convolutional layers typically "convolve" the inputs using a variety of kernels which slide across the datasets reducing dimensionality and deriving scale relevant patterns. As the kernel glides over the input, the scalar product is calculated and the network learns where and why the kernel activates, thus intrinsically recognizing patterns in the data. Afterwards, the pooling layers reduce data dimensionality through

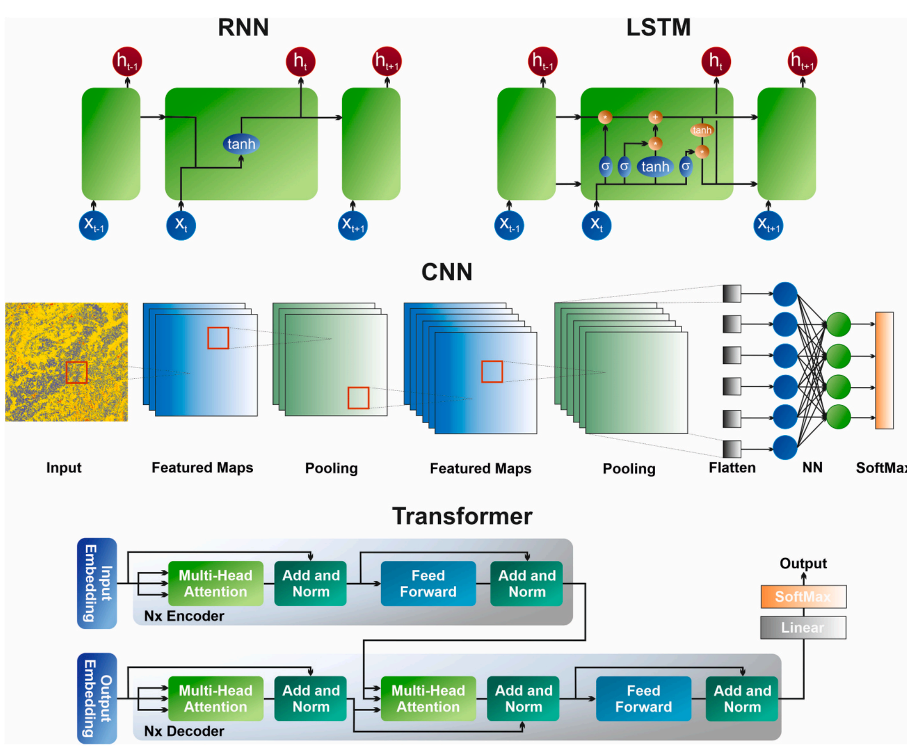

Fig. 2. Advanced DL architectures: The principle of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Convolutional Networks (CNNs), and Transformers.

a variety of user-defined operations (e.g. maximum, minimum, average, L1/L2 Norm, etc.), and feed into final fully connected layers which eventually predict the likely labels per pixel. Significant improvements have been achieved in the representational capacity of deep CNN through architectural innovations (Khan et al., 2020). CNNs have been successful in many remote sensing applications including for vegetation (Kattenborn et al., 2021), image classification (Maggiori et al., 2017; Xu et al., 2018), and object detection (Fu et al., 2020), because they are well-suited for capturing spatial patterns. They can be combined with temporal models or used in 3D form to handle spatiotemporal inputs for SM applications, see also Wang et al. (2019) and Hegazi et al. (2021).

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) enable generative modeling using CNNs, where the model learns to generate new samples that plausibly could have been drawn from the original dataset. GANs consist of two sub-models: the generator model that learns to generate new examples, and the discriminator model that tries to classify them as either real (from the domain) or fake (generated). The two (typically CNN) models are adversarially trained simultaneously, until the discriminator model is wrong $\sim$ 50 % of the time. GANs attempt to represent the complex latent space (the part of the observational distribution not observable) of the observed data, using a (relatively simple) multi-dimensional approximation (Creswell et al., 2018). Once trained, GANs can sample from this learned latent space to generate plausible examples of the observed data distribution, which can be useful for downscaling applications in SM (Jiang et al., 2022).

Another emerging field is that of geometric DL, which moves away from the need for regular grids to the possibility of predictions at unstructured mesh nodes in non-Euclidean spaces (Bronstein et al., 2017). Geometric deep learning allows enforcing known systemic symmetries resulting in inductive biases, which decrease the required training data and allow the use of different data types (Bentivoglio et al., 2022). Of these, the *Graph Neural Networks (GNNs)* are the most well developed so far, which use different types of graphs defined by a set of nodes and edges and process the information encoded therein with neural networks. GNN-based models also enable domain weighting (Wu et al., 2021b), but only consider pairwise geometrical properties between nodes. Thus, mesh CNNs have been proposed to generalize GNNs to higher order geometrical structures, which can even incorporate information on triangular and polyhedral elements. GNNs can model SM over irregular sensor networks or regions by capturing spatial dependencies in a graph structure, such as topography or catchment connectivity. When combined with temporal dynamics, they offer a powerful framework for spatiotemporal SM prediction.

One key limitation of CNNs or indeed any ANN inspired DL structures, is the inability to map or model dynamic phenomena. Feedforward neural networks where information is mapped linearly forward from input to output are typically limited to static classification tasks. In order to enable such networks for dynamic tasks, the concept of memory and time need to be introduced. Feeding signals from previous timesteps back into the network, through so-called recurrent connections then give *Recurrent Neural Networks (RNN)* which are better suited for complex and dynamic geophysical tasks (Williams and Zipser, 1989), see Fig. 2. RNNs can only look back in time for a maximum of ten timesteps, due to the feedback signal vanishing or exploding (Staudemeyer and Morris, 2019). RNNs are suitable for modeling the temporal evolution of SM using satellite time series or meteorological inputs. However, they are less effective at capturing spatial variability unless integrated with other models. A special case of RNNs, the *Long Short-term Memory Networks (LSTMs)* (Hochreiter and Schmidhuber, 1997) deal with the issue of vanishing gradients by enforcing constant error flows for particular cells, such that short-term memory storage is achieved for longer time series (Fig. 2). Fang and Shen (2020) demonstrate the possibility of using LSTMs for nowcasts of SM with the real time assimilation of satellite SM data to yield unprecedented accuracies as compared to subsequent SM retrievals.

With the publication of ChatGPT (Generative Pre-trained Transformer) for generating coherent and contextually relevant responses in a conversation by answering questions, providing information, and engaging in open-ended discussions, the transformer as an AI method gained public interest. The use of ChatGPT in Earth (Foroumandi et al., 2023a) and environmental sciences (Zhu et al., 2023), has also been widely discussed, with critical implications to produce scientific text, codes, and indeed teaching/evaluation concepts. Transformers overcome the sequential processing limitations of RNNs and LSTMs by an encoder-decoder structure, with attention mechanisms enabling the model to weigh the importance of different parts of the input sequence when making predictions (Fig. 2). By doing this, transformers can not only predict whether a particular feature appears in the image, but also exactly where it should appear (Jamali et al., 2024). Potential transformer applications in SM remote sensing may be successful with high performance due to effectively handling long-range dependencies and understanding global context, so that it is an alternative to applications by LSTMs and their derivatives. Examples for transformer applications in SM remote sensing are provided by Liu et al. (2025), Wang and Zha (2024) and Madhukumar et al. (2024).

2.2.3. Hybrid explainable AI techniques

Integrating prior physical knowledge into AI algorithms is not a new concept. However, the emergence of *Physics-Informed Neural Networks (PINNs)* represents a novel scientific AI technique tailored for solving problems associated with Partial Differential Equations (PDEs). PINNs function as DL networks, generating estimated solutions at specific points within the integration domain of a differential equation after training (Cuomo et al., 2022). A notable innovation in PINNs is the inclusion of a residual network that encapsulates the governing physics equations, marking a significant advancement in this approach. PINNs are promising for SM applications where physical processes like infiltration, evapotranspiration, and runoff are governed by known equations. By embedding these laws into the network, PINNs offer physically consistent SM estimates that generalize across conditions and locations. However, the presence and ubiquity of uncertainties in the input data used for predictive modeling, measurements, as well as process understanding, often lead to a degradation of retrievals when physics-based constraints are too strongly applied (Beven, 2020). For example, Frame et al. (2023) show that a DL-based LSTM network without mass conservation constraints, was able to learn the non-uniform patterns of biases in input–output data, better than traditional physics-based hydrological models.

While PINNs focus on incorporating physical principles into AI, explainable AI (XAI) techniques complement them by providing tools to interpret and explain the model's predictions and the importance of different physics-based constraints (Mamalakis et al., 2022). For instance, *LIME* (Local Interpretable Model-agnostic Explanations, Ribeiro et al. (2016)) generates perturbations around a specific instance in a dataset, obtaining predictions for these perturbed instances, and then fitting an interpretable model (such as linear regression) to explain the local behavior of the complex model. Others are *SHAP* (SHapley Additive exPlanations, Lundberg and Lee (2017)), and *P/RFI* (Permutation/Relative Feature Importance, Breiman (2001)). These techniques offer various ways to interpret and understand ML models, providing insights into feature importance, model behavior, and the factors influencing individual predictions, which help in enhancing trust in AI products.

2.2.4. Transfer learning

ML relies on having labeled environmental data for effective training, yet this is frequently unavailable or insufficient for estimating the number of trainable parameters. Additionally, disparities in data distributions, stemming from variations in data collection dates, spectral ranges, and geospatial environments (referred to as domain shift), pose challenges. When DL models are trained on labeled data from a domain rich in labeled information (the "source domain"), they often exhibit subpar performance when directly applied to tasks in the domain of interest (the "target domain") (Ma et al., 2024). This common occurrence of domain shift hampers the generalizability and transferability of ML models. To address this limitation, the concept of Transfer Learning (TL) has been introduced. TL involves transferring knowledge acquired from the source domain to perform related tasks in the target domain (Ma et al., 2024). In the realm of SM remote sensing, this concept holds the potential to facilitate knowledge transfer across scales or to train a ML model on one dataset and apply its insights to another, for example.

3. AI applications in soil moisture remote sensing

3.1. Ai-based soil moisture retrieval

Advancements in AI have opened up possibilities for estimating SM across diverse locations, conditions, climates, and land surfaces. The AI techniques proposed in Section 2 offer a means for retrieving SM. Initial applications emerged between 2000 and 2010, focusing on non-linear machine learning methods like multi-layer perceptron NNs and support vector regression techniques (Ahmad et al., 2010; Gill et al., 2006; Notarnicola et al., 2008; Pasolli et al., 2009). In the early stages, these AI techniques were employed without tailored or generative aspects, and lacked combined or intricate retrieval architectures, resulting in SM estimations that were not yet highly performant or versatile at required resolutions.

In the intermediate years from 2010 to 2020 the applied ML methods proliferated in techniques and specified for the variety of input data (from optical to microwave remote sensing). In an evaluation of different AI-based SM retrieval methods Amer et al. (2011) showed in 2011 that SVMs still outperformed ANNs and multivariate linear regression (MLR) models. The study was conducted over ten sites of the Lower Colorado River Basin in the western United States. The SVM was trained for five years (1998–2002) and tested for the following three years, emphasizing the need for a sufficiently long time series of input data. Jahan and Gan (2014) compared SM estimation with, IEM, SVM, ANN and an Adaptive Neuro-Fuzzy Interference System (ANFIS) on Radarsat-2 images from 9 sites (agriculture, meadows and pasture) within the Paddle River Basin in Alberta, Canada, on 10 days between 2009–2011. SVM performed best ($r = 0.85$), whereas for all AI approaches vegetation indices and land surface temperature from MODIS were included as auxiliary data. The estimation of SM under vegetation cover reveals the potential of data fusion from microwaves and optical wavelengths. Even for steep meadows in the Tyrol Alps, support vector regression (SVR) techniques Pasolli et al. (2012) proved their potential to deal with small scale spatial heterogeneities and to retrieve small-scale SM in 2010 and 2011 from RADARSAT-2 and ENVISAT ASAR with a Root Mean Squared Error (RMSE) from 0.045-0.07 m^3/m^3 (Pasolli et al., 2015). Santi et al. (2019) also used compact polarimetric RADARSAT-2 data and trained an ANN for two agricultural sites in Canada to estimate SM from the C-band data. In order to map multi-layer SM in three depths (5 cm, 25 cm & 60 cm) across the entire state of Oklahoma, U.S., from 2010 to 2014, Zeng et al. (2019) used multi-source data (e.g. SMOS, MODIS, TRMM, SRTM & in situ) and the RF technique to acquire daily and 8-day SM maps. The achieved accuracies of 0.038 m^3/m^3 to 0.05 m^3/m^3 for a year-to-year experiment and 0.044 m^3/m^3 to 0.057 m^3/m^3 for a station-to-station experiment. This is remarkable given the heterogeneity in land cover types, topography and soil variation across Oklahoma. Finally, a comprehensive review on SM (and biomass) assessment with ML techniques for the earlier years (until 2015) is provided by Ali et al. (2015).

Recently (from 2020 onwards), the appearance of ML or AI techniques for SM retrieval and estimation is multifaceted and ubiquitous. Rani et al. (2022) provided a very comprehensive summary on machine learning techniques, including linear regression, ANN, DNN, SVM, CART & RF, for SM estimation. Furthermore, three techniques (SVM, K-Nearest Neighbors (KNN) & RF regression models) were trained for agricultural areas around Midnapore (West Bengal, India) from Datta et al. (2021) to estimate SM from bare soils using time series of Sentinel-1 (for SM dynamics) and Sentinel-2 (for mapping bare soil fields) data. The RF approach outperformed SVM and KNN leading to an RMSE of 0.03 m^3/m^3 and an $r^2 = 0.93$ for the dry season from January to March 2019. Nguyen et al. (2022) also based their approach on XGBoost regression together with a genetic algorithm (GA) optimizer for SM estimation in agricultural areas in West Australia. They feed Sentinel-1, Sentinel-2 and ALOS Global Digital Surface model data in the form of 52 predictor variables into the ML algorithm to retrieve SM at 10 m spatial resolution. The developed algorithm was also compared against SVM, RF and CatBoost regression methods and outperformed these with an R^2 of 0.89. Greifeneder et al. (2021) used Gradient Boosted Regression Trees with Sentinel-1 to train on International Soil Moisture Network (ISMN) SM data, see Fig. 3.

Wang et al. (2024) developed a stacked learning approach combining four regression approaches (classified regression tree, RF, gradient boosting decision tree (GBDT), and extreme random tree) and two stacking ensemble models (least absolute shrinkage and selection operator (LASSO) & the generalized boosted regression model (GBM)). They found that enhanced vegetation index, land surface temperature, radar backscatter, temperature, vegetation drought index, as well as solar local incident angle qualified as the most valuable predictors out of 19 candidates for SM estimation. Interestingly, the stacking models were more accurate in prediction of SM than the single optimized regression models showing the complexity as well as the potential of combined model architectures. Zhang et al. (2023a) developed an enhanced generalized regression neural network (EGRNN) with backward sequential feature selection (EBSFS) method for SM estimation over the Qinghai-Tibet Plateau from April 2015 to March 2016. This iterative method optimizes the variable search to find the best feature subset starting with 19 features and aiming at minimum feature redundancy. Best performance ($r = 0.95$ & $RMSE = 0.03\ m^3/m^3$) is achieved for 13 input features. External factors such as groundwater influence can be integrated into the AI system, as proposed by Wang and Zha (2024). They found that in-situ groundwater level data improves SM time series predictions in shallow groundwater regions. Moreover, in their study the Transformer architecture better supports SM time series prediction capabilities compared to LSTM, while coupling the LSTM structure to the Transformer enhances the stability of the system when dealing with long time lags.

The GNSS-R SM retrieval field of research is heavily involved in AI. The introduction of AI-based retrieval methods has addressed the limitations of traditional fitting equations by eliminating the need for approximations and complex parameters in the computation of empirical and physical models (Yang et al., 2024). E.g., Roberts et al. (2022) trained a CNN using CYGNSS delay-Doppler maps and ancillary datasets as inputs, with SMAP SM as target, generating a SM product from 2017 to 2019 that compares well with existing global products. The CNN predictions showed strong correlation with SMAP targets and in-situ measurements, offering potential advantages in spatial resolution and coverage, especially in regions where SMAP underperforms (high porosity, dense vegetation). Munoz-Martin et al. (2021) uses GNSS-R data from Flexible Microwave Payload – 2 (FMPL-2) onboard FSSCat with a similar strategy and SMAP targets.

Since ML techniques are fully established as retrieval option for SM estimation by remote sensing, the above discussion of approaches must be seen as exemplary showcasing the potential and variety of options (Adab et al., 2020; Babaeian et al., 2021; Batchu et al., 2023; Chaudhary et al., 2022; Cui et al., 2020; Santi et al., 2019; Senyurek et al., 2020; Singh and Gaurav, 2023). Lamichhane et al. (2025) performed a structured literature analysis about machine learning approaches in SM remote sensing and provided in table 2 a comprehensive overview about studies in that field including the region and scale as well as the method applied and the data sets utilized. AI-based SM retrieval has made significant strides, but several challenges remain. The accuracy of these

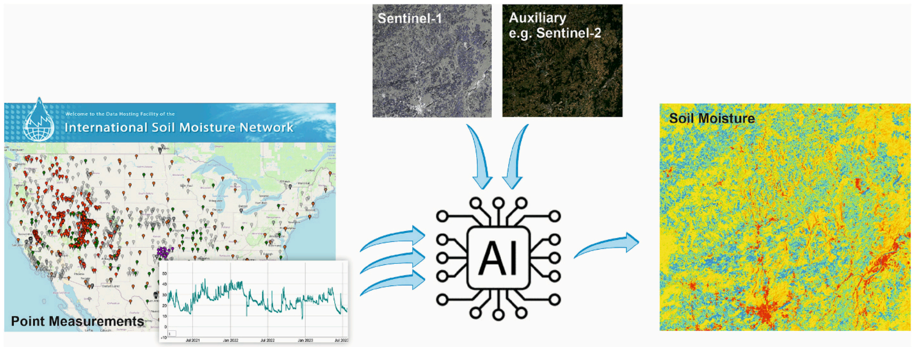

Fig. 3. Example of SM retrieval by AI, modified after Greifeneder et al. (2021).

models is highly dependent on the availability and quality of input data, which can be inconsistent across different regions. AI models trained on specific environmental conditions may not generalize well to diverse locations, limiting their broader applicability. Here, support of the ISMN to extend the in-situ observation networks to non-covered environments, is necessary.

3.2. Time series reconstruction

SM is a dynamic parameter that can exhibit rapid changes due to weather events, land use, and seasonal variations (Seneviratne et al., 2010). Time series data provide valuable insights into these temporal dynamics, helping us understand SM patterns over different timescales (Vereecken et al., 2014). As the field of SM remote sensing continues to advance and climate change related threats to food and water security grow, the need for accurate and continuous monitoring of SM levels has become increasingly apparent (Ochsner et al., 2013). However, inherent challenges such as spatial gaps and temporal inconsistencies in the collected data due to issues like snow cover, high canopy density, frozen soil, or extremely dry soil, sensor malfunctions, or limitations in data acquisition frequency, capabilities of satellite orbit/swath and retrieval algorithms (Dorigo et al., 2017), making it challenging to monitor SM dynamics continuously and detect short-term changes. Historically, filling temporal gaps in SM time series data has relied on traditional statistical methods like linear interpolation, moving averages, or seasonal decomposition (Llamas et al., 2020; Sandholt et al., 2002; Wang et al., 2012). While these methods can provide reasonable estimations, they have limitations in capturing complex, nonlinear relationships and addressing irregular gaps effectively (Mi et al., 2023). AI, particularly ML techniques, has emerged as a powerful tool to overcome these issues, and has revolutionized our ability to reconstruct high-frequency, reliable time series data from various remote sensing platforms (Chen et al., 2021). These techniques can adapt to the specific characteristics of SM data and capture complex spatiotemporal dependencies.

ML models have been used for time series reconstruction of remotely sensed SM using various approaches from simple models to hybrid methods combining ML with process-based models. For example, classical ML methods have been used for monthly to daily gap-filling of satellite-retrieved SM at region to global scales, and include back propagation neural network (BPNN; Yao et al. (2017)); DT-based regression models such as RF, XGBoost, and their variants (Ahmad et al., 2010; Liu et al., 2023; Liu et al., 2018; Mi et al., 2023; Nadeem et al., 2023; Sun and Xu, 2021). This includes also the detection of (and partly filling the gaps caused by) Radio Frequency Interferences for passive microwave sensors like SMOS or SMAP (Alam et al., 2022; Mohammed and Piepmeier, 2021; Nazar and Aksoy, 2023).

In recent years, spatiotemporal DL networks have become an increasingly popular choice for SM time series reconstruction due to its ability to encode prior system states in the cell memory. For example, Fang et al. (2017) used a LSTM to reconstruct SM of the United States. ElSaadani et al. (2021) assessed the spatiotemporal DL method for filling the gaps in SM observations, and Li et al. (2022a); Li et al. (2022b) further improved satellite SM prediction using the DL model. Zhang et al. (2021b) developed a spatiotemporal partial CNN framework to implement gap-filling for the AMSR2 SM product for the years 2013–2019. Chen et al. (2021) conducted data calibration and data fusion of eleven well-acknowledged microwave remote sensing SM products since 2003 through an ANN approach, with SMAP SM data applied as the primary training target. The training efficiency was high ($R^2 = 0.95$) due to the selection of nine quality impact factors of microwave SM products and the complicated organizational structure of multiple neural networks (five rounds of iterative simulations, eight substeps, 67 independent neural networks, and more than 1 million localized subnetworks). Then, they developed a global remote sensing-based surface SM dataset (RSSSM) covering 2003–2018 at 0.1° spatial sampling. Shangguan et al. (2023) implemented a Spatiotemporal Attention-Based Residual Deep Network to fill gaps in the ESA-CCI data. As a DL method, it used transfer learning by ERA5-Land (ECMWF Reanalysis) SM to circumvent limited original training data. As the combined AMSR-E/2 data set provides the longest SM time series from a single mission (with two satellites), Sivaprasad et al. (2025) combined convolutional LSTM (ConvLSTM2D) layers and CNN layers to reconstruct a gap-free SM time series for Europe spanning from 2003 to 2023. As it was intended to be utilized in data assimilation in a later step, ASCAT without any further auxiliary data set was used as a predictor to establish a model with AMSR2 as the target, while using transfer learning for the AMSR-E period.

Moreover, hybrid models with ensembles of standalone ML models using methods such as bagging, stacking, and random subspace (Das et al., 2022) or combining ML with process-based models (Lee et al., 2022) have been used to surface SM mapping and gap-filling. More recent advances in SM time series reconstruction are targeting high-resolution (1 km), gap-filled (2000 to present), global daily products by integrating satellite data, reanalysis, and ML techniques (Han et al., 2023; Zhang et al., 2023c; Zheng et al., 2023). In all cases, the ML approaches were able to improve prediction accuracies compared to basic statistical and process-based models (or reanalysis).

3.3. Root zone soil moisture

Several studies have used AI techniques to retrieve root zone SM (RZSM), as it is critical for various applications, see also Lamichhane et al. (2025). Kornelsen and Coulibaly (2014) were pioneers in testing an ensemble of ANNs for providing RZSM at depths of 10, 20 and 50 cm using land surface observations. The study incorporated modeled SM observations from the well-established HYDRUS process-based model, which was calibrated with in situ data. The ANNs were found to be highly effective in replicating the behavior of the model. Pan et al. (2017) implemented the same ANN approach to estimate continental scale (United States) RZSM from SMOS. The resulting spatial patterns of

RZSM were consistent with those produced by the Noah model of the Global Land Data Assimilation System. Zhang et al. (2017) evaluated three distinct approaches – the exponential filter, ANN, and linear regression – in estimating RZSM through in situ measurements in Oklahoma. The results demonstrated that the exponential filter was superior to its counterparts in performance. Three methods, namely the exponential filter, ANN, and cumulative distribution function (CDF), were evaluated also by Tian et al. (2020) in a mountainous region of China by using in situ and SMAP SM data. The findings suggest that the exponential filter performed better than the other techniques and demonstrated a good performance when implemented at regional scale (with SMAP data) compared to in situ data.

Karthikeyan and Mishra (2021) took another route and estimated SM in different depth (5 cm, 10 cm, 20 cm, 50 cm & 100 cm) down to the root zone with the XGBoost algorithm at 1 km spatial resolution for the contiguous United States of America (CONUS) based on the SMAP L4 root zone SM product (a model-data fusion result) together with climate and landscape datasets. A validation on 79 SM in situ measurement stations across CONUS led to an unbiased RMSE of less than 0.04 m^3/m^3. Especially the sub-grid heterogeneity is captured well by the ML approach compared to the original SMAP L4 product. Babaeian et al. (2021) implemented an automated ML approach to estimate RZSM from surface measurements obtained from an Unmanned Aerial System at the field scale in Arizona. Very good estimates of RZSM were obtained when compared with in situ observations. Carranza et al. (2021) assessed a RF methodology for interpolating and extrapolating RZSM in a small Dutch catchment. They found that the technique was similarly accurate compared to process-based models (HYDRUS) and emphasized the importance of land cover and soil physical properties for achieving improved outcomes. More recently, A et al. (2022) tested the convolutional LSTM (ConvLSTM) model for its capability to capture spatio-temporal patterns as compared with ANN techniques. The model was trained using simulated SM data from HYDRUS and was found to be more effective than the Global Land Data Assimilation System RZSM product in northeastern China. Sungmin and Orth (2021) developed a global SM dataset by training a LSTM with in-situ measurements and forced with meteorological reanalysis data (no satellite data) and static data. The dataset named SoMo.ml provides multi-layer SM data (0–10 cm, 10–30 cm, and 30–50 cm) at 0.25° spatial and daily temporal resolution over the period 2000–2019. The dataset was found outperforming reanalysis (ERA5) and satellite datasets (GLEAM, ESA-CCI) in terms of correlation versus independent in situ observations.

The analysis of scientific literature indicates that AI techniques possess significant potential for estimating RZSM from satellite measurements. Earlier studies have found lower performance to capture the temporal SM variations of ANN techniques with respect to statistical and analytical approaches such as the exponential filter (Tian et al., 2020), but more recent approaches (e.g., LSTM, ConvLSTM) seem to provide improved performance (A et al., 2022). However, for these recent approaches, comparison studies have not been carried out so far. Indeed, the exponential filter approach is currently the most commonly employed method for continental and global-scale RZSM estimation when satellite data are used as input (e.g. the Copernicus Soil Water Index, https://land.copernicus.eu/global/products/swi and the SMOS Level 4 RZSM, Al Bitar and Mahmoodi (2020)).

The study by Sungmin and Orth (2021) showed an interesting application on a global scale, but using only reanalysis and ancillary data as input. The application of the LSTM approach with satellite data would be an interesting way forward. The use of ML techniques for large-scale applications necessitates the incorporation of precise ancillary data for land use and soil physical attributes, and by extension requires the participation of diverse communities.

3.4. Spatial scaling

Spatial scaling of SM refers to the process of changing the spatial resolution of SM data, either by increasing it (downscaling) or decreasing it (upscaling), while preserving its inherent properties. This is often necessary because satellite- and model-based SM products typically have coarse spatial resolutions, which may not be sufficient to capture fine-scale variations in different hydro-climatic conditions. The mechanisms of spatial scaling can be used to generate higher-resolution SM data that is more relevant for local- and regional-scale decision-making and research. For example, downscaled SM data can be used to improve the accuracy of irrigation scheduling or to predict flood risk (Dari et al., 2020; Kwon et al., 2022). Upscaled SM data can be used to assess the impact of climate change at a regional/global scale (Lal et al., 2023a) or to monitor large-scale agricultural drought (Chatterjee et al., 2022; Martínez-Fernández et al., 2016).

The process of downscaling SM in spatial scaling is more intricate, as it involves employing complex algorithms and multiple ancillary datasets to maintain the inherent properties of the data while introducing fine-scale variability. In contrast, upscaling SM employs a simpler approach, such as kriging or spatial interpolation, primarily aimed at representing average SM conditions over a broader area.

Spatial downscaling of SM can also be achieved through conventional methods, including data-driven approaches (Das et al., 2011; Lal et al., 2023b), empirical techniques (Piles et al., 2016), physically based models (Merlin et al., 2008), data assimilation methods (Kaheil et al., 2008; Reichle et al., 2001), and land surface modeling (Vergopolan et al., 2020). These approaches have been developed and applied to downscale SM using optical, thermal, and active microwave data, see Fig. 4. Nevertheless, implementing these methods can pose challenges, including the need for multiple datasets, difficulties in data filling, a requirement for synchronized overpass times, computational inefficiencies, and issues with the relationship between SM and backscatter when using SAR data, particularly in regions with high vegetation water content. These challenges may introduce biases in the retrieved high-resolution SM data.

To address the limitations of the aforementioned spatial scaling methods, AI has emerged as a promising approach for spatial downscaling of SM. AI models have the capability to learn complex nonlinear relationships between SM and various land surface variables, including vegetation cover, topography, surface temperature, and climate. This enables them to downscale coarse-resolution SM to finer spatial resolutions while preserving its spatial variability (Fig. 4). In this context we need to stress that the evaluation of the downscaling result needs to be performed in the spatial domain. Examples can be found in Montzka et al. (2018) and Merlin et al. (2015), a comprehensive discussion about validation methods is provided by Gruber et al. (2020) and Montzka et al. (2020). Instead, most studies evaluate the spatial scaling performance in the temporal domain with few reference sites only, without any relevance for the scaling procedure.

In general machine learning algorithm approaches such as ANNs, RF, SVM and XGBoost are widely used and considered as powerful technique for spatial scaling i.e., downscaling of SM (Luo et al., 2023; Mohseni et al., 2023; Moosavi et al., 2016; Sun and Cui, 2021). This is true also for complex terrain (Chen et al., 2023). Recent research has underscored the promise of ensemble learning models, notably XGBoost and RF, in achieving high-resolution SM retrieval. For instance, Zhang et al. (2022b) successfully integrated Landsat 8 optical and thermal observations with diverse datasets, employing XGBoost and RF models to attain high-resolution SM data at a 30-meter scale. Interestingly, the XGBoost model demonstrated a slight edge over RF, producing results comparable to SMAP SM but with more pronounced spatial variability. However, this does not necessarily indicate a better performance in terms of reproduction of spatial patterns as the evaluation was conducted in the temporal domain only. Moreover, Karthikeyan and Mishra (2021) utilized the XGBoost model to downscale SMAP Level-4 multi-layer SM from 9 km to a high spatial resolution of 1 km. Their research incorporated surface and rootzone SM data alongside climate and landscape datasets as predictor variables. The multi-depth SM

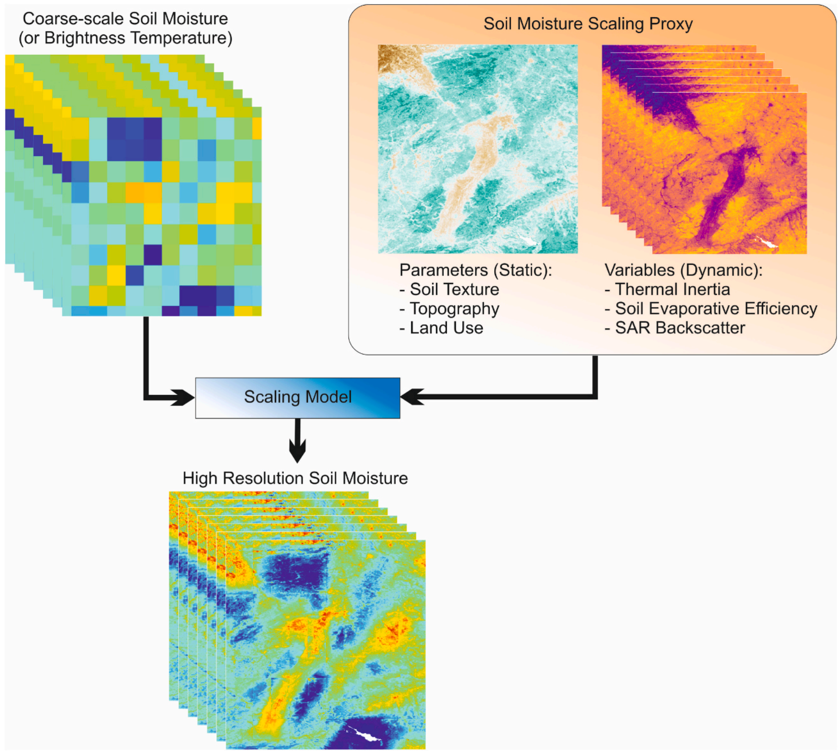

Fig. 4. Example of a SM downscaling procedure.

information is a result of land surface model simulations and assimilated SMAP brightness temperature observations, so that 3-hourly SM at 5, 10, 20, 50, and 100 cm depth is available as input for the XGBoost model. The findings shed light on the varying degrees of influence that predictor variables have on SM, contingent upon soil depth, with meteorological variables assuming the least importance. In a complementary approach, Xu et al. (2021) introduced the use of CNNs to downscale Enhanced SMAP Level 3 Passive (L3_SM_P_E) and Original SMAP Level 3 Passive (L3_SM_P) SM to 1-kilometer high-resolution data. Their method entailed the design of a weight layer for input and considered the residual SM layer as the model's output to enhance accuracy, with validation studies showcasing an impressive 95.81 % similarity with SMAP SM products. Furthermore, in a recent study, Mohseni et al. (2023) developed a 1-kilometer SM product for Africa by combining SMAP L3_SM_P_E and MODIS data through an RF model hosted on a cloud platform. The results indicated consistent findings with SMAP SM, accompanied by enhanced spatial variability (also here the validation was performed in the temporal domain). Remarkably, these four distinct studies employed various ML algorithms, all achieving nearly identical accuracy levels coupled with substantial spatial variability. This underscores the flexibility of ML in generating high-resolution SM data from coarser-resolution inputs, contingent upon the unique hydroclimatic conditions. Cui et al. (2020) scaled FY-3B (National Remote Sensing Center of China and China Meteorological Administration Feng-Yun-3B satellite mission) SM from 0.25° to 0.05°, and filled gaps to improve areal coverage from 23.7 % to 78.7 %. They trained a General Regression Neural Network at low resolution and applied it on higher resolution land surface temperature and vegetation indices.

Additionally, Mao et al. (2022) proposed a spatial downscaling method for high-resolution SM retrieval, leveraging the RF method, while considering memory and mass conservation principles for SM. Their experiment selected three-day, and seven-day lagged SM data for memory conservation, which yielded improvements in the retrieved SM. Simultaneously, Wei et al. (2019) introduced the innovative Downscaling based on Gradient Boosting Decision Tree (DENSE) method to compensate for the limitations of SMAP radar data. The Gradient Boosting Decision Tree (GBDT) emerged as the pivotal ensemble learning technique, iteratively constructing an ensemble of weak DT learners through boosting. The DENSE approach, further enriched with 26 SM indices derived from MODIS and DEM data, selected seven proxy variables to establish a downscaling model. These models successfully retrieved high-resolution SM at a 1-kilometer spatial scale, preserving the accuracy of SMAP data and enhancing coarse-resolution results.

Continuing in a similar vein, Liu et al. (2018) explored six distinct ML algorithms, including ANN, Bayesian, Classification and Regression Trees (CART), K-Nearest Neighbor (KNN), RF, and SVM, to retrieve and compare high-resolution SM data. Their study incorporated ESA CCI SM at 0.25-degree spatial sampling and utilized MODIS data for retrieving 1-kilometer high-resolution SM across various global regions. Validation studies underscored the superior performance of the RF algorithm in terms of accuracy and temporal trend variations across all sites. Likewise, in a study by Im et al. (2016), different ML techniques were employed to downscale AMSR-E 25-kilometer SM to 1 km resolution in two regions (South Korea and Australia), using MODIS data. Among the diverse ML algorithms tested, the RF method consistently emerged as the superior choice, yielding low RMSE and high correlation. Continuing the drive for advancing ML models for downscaling SM products, Jin et al. (2021) proposed the Support Vector Area-to-Area Regression Kriging (SVATARK) model, which skillfully combines SVM and Area-to-Area Kriging techniques. This research effectively downscaled ESA-CCI SM to 1-kilometer resolution, employing ancillary datasets, such as land surface temperature, vegetation indices, land cover, blue sky albedo, digital elevation model, aspect, and slope. Notably, the SVATARK downscaling approach achieved the highest level of accuracy in validation statistics. The dynamic analysis of 1-kilometer SM at various stations corroborated the model's superiority.

Furthermore, a unique approach was introduced by Ming et al. (2022) through the Hybrid Triple Collocation-Deep Learning approach, which facilitated the merging and downscaling of three different SM datasets (SMAP, GLDAS-Noah, and ERA5-Land) from a 0.36-degree resolution to 0.01-degree resolution. This novel downscaling process drew on the relationship between SM data and auxiliary environmental variables, including elevation, land surface temperature, vegetation index, surface albedo, and soil texture. The Hybrid Triple Collocation-Deep Learning method cleverly combined the Triple Collocation (TC) approach with the Long Short-Term Memory (LSTM) network to effectively downscale SM from the merged products, producing superior results compared to the original products. Sungmin et al. (2022) harnessed LSTM for developing a high-resolution SM product (SoMo.ml-EU) for Europe over the period from 2003 to 2020. Their methodology involved scaling in-situ data using the long-term mean and standard deviation of ERA-Land gridded SM. This scaling aimed to eliminate systematic biases in SM means and variabilities arising from diverse sensor types and calibration techniques, as identified in the work of Sungmin et al. (2022).

In addition to the aforementioned studies, it is worth noting that the RF method stands out as a highly utilized model for spatial scaling through AI. RF models have gained prominence due to their superior attributes, including accuracy, robustness, flexibility, and interpretability. They have consistently demonstrated their capability to achieve correlation coefficients of 0.9 or higher when compared to in-situ measurements, and they exhibit robustness in the presence of noise and outliers in the data (Chen et al., 2020; Cheng et al., 2018; Hu et al., 2020; Inoubli et al., 2022; Qu et al., 2021; Sishah et al., 2023; Yan and Bai, 2020; Zhao et al., 2018). Furthermore, RF models offer a degree of interpretability, enabling a deeper understanding of the factors that play a pivotal role in predicting SM. This interpretability can prove invaluable in identifying the underlying physical processes that govern SM variability and in the development of more effective downscaling models.

3.5. Soil moisture forecast

Forecasting SM for future conditions holds significant importance in various domains, such as flood and streamflow prediction (Koster et al., 2010; Yin et al., 2022), irrigation scheduling (Togneri et al., 2022), landslide early warning (Wicki et al., 2020), and managing pests affected by moisture conditions, like desert locusts (Piou et al., 2019). SM predictions can draw from historical data or rely on atmospheric forecasts from numerical weather prediction models. Notably, these forecasts may not necessarily incorporate remotely sensed SM data. Traditional weather prediction models often provide forecasts with coarse spatial resolutions. AI methods show promise in predicting both seasonal and short-term SM states, offering the advantage of obtaining data at management scales ranging from tens to hundreds of meters. As exemplified by the study conducted by Karandish and Šimůnek (2016), ML can serve as a valuable alternative to physics-based models, particularly when dealing with limited input data. This suggests that computationally lightweight ML models can be developed for SM forecasting.

In the first scenario, the focus is on seasonal forecasting. For instance, Joshi et al. (2023) employed a Gradient Boosting ML model to predict Australian summer monsoon SM at varying depths, from the root-zone (0–1 m) to deeper layers (1–6 m). Their approach incorporated site-specific attributes such as soil properties and topography, alongside meteorological data and vegetation information spanning 19 years. This study revealed that summer SM could be accurately anticipated as early as the end of winter, with potential enhancements achievable through the integration of remotely sensed vegetation data as indicators of plant response. Prasad et al. (2018) developed an innovative approach known as the Artificial Neural Net Committee of Models (ANN-CoM) system. This method aggregates outputs from four individual models, including the second-order Volterra, M5 model tree (method combining decision trees with linear regression models to make accurate predictions, particularly effective for problems involving continuous target variables), random forest, and extreme learning machine models, which then feed into another ANN to generate collective forecasts. Their model incorporated a wealth of predictors, including hydro-meteorological data, large-scale climate indices, and atmospheric parameters derived from ERA Interim reanalysis, amounting to a total of 60 potential predictors for SM forecasting. Utilizing Neighborhood Component Analysis feature selection, the model selected the optimal features. Benchmarking against the four primary standalone models demonstrated the system's ability to provide reliable seasonal forecasts. Additionally, Zeynoddin and Bonakdari (2022) employed a simple LSTM with a single hidden state. They used stationary SM data, removing seasonality and negligible trends, sourced from the SMAP mission. Their results showcased high performance, emphasizing the benefits of adequate preprocessing. Notably, this performance was surpassed when LSTM optimization was achieved through a combination of genetic and teacher-learner algorithms. In another approach, Ahmad et al. (2010) harnessed remote sensing data, which included backscatter and incidence angle information from the Tropical Rainfall Measuring Mission (TRMM) and the Normalized Difference Vegetation Index (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR). Their prediction model utilized SVM to estimate SM. The estimates compared well against simulations by the Variable Infiltration Capacity Three Layer model, demonstrating the effectiveness of this remote sensing-based approach in SM estimation.

The second category focuses on short to medium-term forecasts, spanning days or weeks. In an early study, Gill et al. (2006) employed SVM to predict future SM levels using historical SM and contemporaneous meteorological data. Dubois et al. (2021) aimed to predict soil water potential for a potato field in the upcoming week to prevent water stress through efficient irrigation scheduling. They harnessed data from tensiometer measurements, rain gauges, as well as mean air temperature and rainfall records from previous and subsequent days. They employed NN, RF, and SVM systems, optimizing their performance with feature selection capabilities. Feature selection is a preprocessing method for dimensionality reduction aimed at producing a more efficient (and therefore, faster) model by excluding features that contribute least to explaining the target phenomenon, while minimizing the impact on model performance. Implementing these near real-time (NRT) capabilities, while promising, poses technical challenges. To address the NRT challenge, Fang and Shen (2020) extended a LSTM with a data integration kernel. This allowed the integration of the most recent SMAP observations as soon as they became available. Similarly, Zhang et al. (2023b) utilized SMAP data with a data integration procedure for NRT SM prediction, employing Gated Recurrent Unit (GRU) and LSTM as machine learning tools.

Li et al. (2022b) developed a causality-structured LSTM model (CLSTM) for forecasting on both short and medium timescales (7–15 days). They used meteorological, energy balance, and land surface variables from FLUXNET sites. Through Granger causality tests, they constructed a causal structure for the input variables, which improved LSTM performance. While they noted that incorporating numerical weather predictions could enhance results, it could introduce system errors from the numerical model. Tandon et al. (2022) designed an ANN to predict daily SM up to 5 days in advance, along with associated forecast uncertainty. They re-estimated ANN parameters within a probabilistic framework using a Particle Filter, with the ANN mimicking the state space model, where ANN parameters represented the system's states.

At very short timescales (5 to 24 h), Basak et al. (2023) introduced novel methods grounded in deterministic and physically based hydrology, accompanied by optimization procedures. These methods, known as the Naive Accumulative Representation and Additive Exponential Accumulative Representation, employed single and dual exponential functions to describe SM responses to rainfall. Compared to classical ML methods such as LSTM and RF, these approaches demonstrated superior performance. Cai et al. (2019) observed that a deep learning neural network (DLNN) offered more accurate daily SM predictions based on a variety of meteorological factors, including daily precipitation, mean surface temperature, wind speed, relative humidity, air pressure, and temperature, in comparison to the Multi-Layer Perceptron (MLP) model.

In contrast to studies utilizing a multitude of input variables, Heddam (2021) conducted successful testing of four ML models, namely MARS (Multivariate Adaptive Regression Splines), M5Tree, RF, and MLP-NN, to estimate SM while considering only hourly soil temperature as the input. All AI methods outperformed a multiple linear regression model, with the MARS approach performing best. Utilizing a single data set can speed-up the forecast. Similarly, in 2023, Granata and colleagues employed an ensemble of ML methods, incorporating precipitation, air temperature, humidity, and wind speed as inputs to predict short-term SM. Their investigation centered on whether precipitation alone, as an exogenous input variable, could yield sufficiently accurate predictions for practical applications, and the results were indeed promising.

It's noteworthy that many studies in the realm of SM forecasting do not rely solely on conventional ML methods but rather augment them with data integration, advanced preprocessing, dimensionality reduction techniques, or employ them in conjunction with Bayesian frameworks to further refine the predictive accuracy. Moreover, studies employing remotely sensed SM data in SM forecast are rare, this field would benefit from further research. This could encompass the application of TL approaches when trained on a model and applied to observation data.

3.6. Assessing the water cycle and beyond

SM assumes a pivotal role within the nexus of terrestrial and ecological processes, functioning as a essential climate variable integral to the Earth system. Its influence extends to cloud formation, modulating precipitation occurrences, and consequently, the onset and duration of drought events, contingent upon the availability or scarcity of water evaporating from the soil and assimilated by vegetation (Robock, 2003; Seneviratne et al., 2010; Tuller et al., 2023). Regulating surface soil temperature, SM wields influence over the prevalence of wildfires and heat waves through its modulation of the distribution of available surface energy into sensible and latent heat components. It also governs the likelihood of flooding episodes by determining the extent to which precipitation or snowmelt directly contributes to river and stream flows. In its role as a predictor of monthly to seasonal climate variability, SM stands out as the paramount component of meteorological memory within the terrestrial climatic framework. Notably, it serves as a storage for retaining historical meteorological events, such as droughts or rainy periods, and facilitating the transference of such information into the future through the mechanism of SM memory (Rahmati et al., 2023). Within this discourse, we present a brief overview delineating the utilization of spaceborne SM as an input to AI models, predicting a spectrum of ecosystem variables encompassing the water cycle, carbon cycle, and various biogeochemical cycles, along with its impact on land surface processes, depicted also in Fig. 5.

Due to the importance of SM, there is a body of research on the use of space-based SM measurements and AI modeling to study the state of water resources and thus the occurrence and severity of droughts, subsurface drainage, water requirements for irrigation, etc., at high spatial resolution. For example, Cho et al. (2019) used satellite big data and ML via Google Earth Engine to evaluate subsurface drainage. To this end, they used spaceborne SM data from SMOS together with other remotely sensed proxy indices to predict subsurface drainage using the RF classification method. Although not directly spaceborne SM is used, Deines et al. (2017) introduced the Water-Adjusted Green Index (WGI), which integrates moisture information with vegetation greenness. The index is calculated from Landsat data as the product of the Normalized Difference Water Index (NDWI) and the Green Index (GI). WGI, along with other indices, was used as input to a RF classifier to generate annual, high-resolution (30 m) irrigation maps for the period 1999–2016, achieving an accuracy of over 92 %. Similarly, Gökkaya et al. (2017) fed shortwave infrared (SWIR) reflectance as indicator of SM (as they are inversely but highly related) along with other proxy variables to DT classifier to detect subsurface tile drained area in an agricultural watershed, Shatto Ditch Watershed (SDW) in Indiana, USA with accuracy higher than 94 %. Brust et al. (2021) introduced a ML framework, named DroughtCast, which uses spaceborne SM data from SMOS along with other variables as inputs to forecast the United States Drought Monitor (USDM). Their results show that DroughtCast was able to forecast 2017 Northern Plains Flash Drought, as a very extreme drought event, up to 12 weeks before its onset.

There are also a variety of publications using AI techniques to downscale space-based SM data and then monitor land surface processes, particularly, but not limited to, drought monitoring. For example, Ma et al. (2021) and Ma et al. (2022) used LSTM to estimate water table depth anomalies from SM and precipitation, the latter implemented a TL scheme while training on a hydrological model and implementing the trained system on remote sensing data. Foroumandi et al. (2023b) employed the Gravity Recovery and Climate Experiment

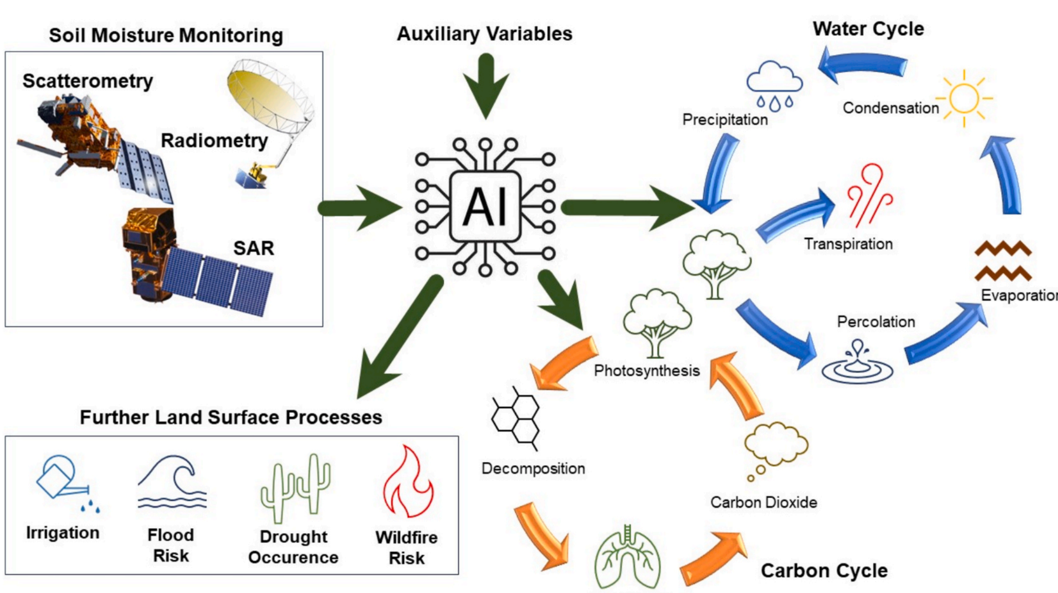

Fig. 5. AI in SM remote sensing and assessing further land surface processes.

(GRACE)-derived Terrestrial Water Storage Anomaly (TWSA) data, as an indication of SM anomaly plus groundwater storage and canopy water storage anomalies, to map drought in Iran for a period of 2002 to 2016. To do this, they employed DL Convolutional LSTM (ConvLSTM), shallow learning Feed Forward Neural Networks (FFNN), and RF to downscale GRACE data to 10 km spatial resolution and then computed drought frequency maps by applying annual standardized precipitation index (SPI) method. Similar studies by Rahaman et al. (2019), Verma and Katpatal (2020), Zhang et al. (2021a), Majumdar et al. (2020), and Sahour et al. (2020) used SM data to downscale GRACE data to answer different environmental questions.

There are also a number of publications that aim to integrate low-resolution SM data from space observations with commonly available meteorological data or else using AI techniques and then monitor land surface processes. Liu et al. (2017) applied SVMs to combine in-situ meteorological data and remotely sensed products including leaf area index (LAI), AMSR-E and SMAP SM retrievals for near-real time agricultural drought forecasting. Their results showed that assimilating AMSR-E and SMAP SM retrievals with in-situ meteorological data through SVMs significantly improved the accuracy of drought monitoring. Similarly, Park et al. (2017) downscaled the AMSR-E SM data over the Korean peninsula using seven products from the Moderate Resolution Imaging Spectroradiometer (MODIS) and Tropical Rainfall Measuring Mission (TRMM) satellite sensors through the RF machine learning approach and then provided a high resolution SM Drought Index (HSMDI) to monitor meteorological and agricultural droughts with high accuracy.

A number of publications have been also released indicating that in-situ measured SM data alone or together with modeled SM data are used as input to AI models predicting various land surface responses such as evapotranspiration (ET), carbon cycling, nutrients, and wildfires, which are mainly confined to small areas or regions. However, SM data derived from space observations can be used instead of in-situ SM data in those models to improve the spatial extent of these applications for large-scale decision making. For example, Babaeian et al. (2022) used LSTM and ConvLSTM using the main meteorological and ground measured (i.e. SM) input variables to calculate the actual ET in seven different climate zones in the contiguous United States in real time and in advance (forecast). Similarly, Fernández-López et al. (2020) applied ANNs, support vector regression, and other methods such as randomizable filtered classifiers to estimate reference ET using in-situ SM data as input. Trained on FLUXNET in-situ SM data, Zhao et al. (2019) also implemented a physically constrained ML method to estimate ET with the aim of obtaining the surface energy balance. In-situ SM data is also widely used as input for AI models when it comes to irrigation scheduling (Gu et al., 2021) or yield estimation (Chakraborty et al., 2023), which can be combined with remotely sensed SM data for large-scale applications.

There are also studies showing the impact of SM regimes on carbon mineralization rates and consequently on CO_2 emissions (Schlüter et al., 2022). Combining these studies with remotely sensed SM data and AI models can contribute to a global assessment of CO_2 emissions from different sections and provide strategies for better control of these emissions. In this sense, a recent review article by Grunwald (2022), which examines the opportunities, potentials and threats of AI in soil carbon modeling, states that remotely sensed SM data are among the environmental covariates that serve as input for AI-based soil carbon models. Keskin et al. (2019) also employed several AI models along with 327 geospatial soil-environmental variables (including SM) to explain the variation in soil C pools and total carbon content and showed that biotic and hydro-pedological factors (which also includes SM) explained most of the variation in C pools and total carbon content.

The assessment of soil nutrient status based on in situ or modeled SM together with other data applying AI is also of interest for various research projects. However, the potential of SM remote sensing has not yet been fully explored and could be a possible field of research for the future. The same is true for the application of remotely sensed SM and AI for fire hazard models. In this context, Sharma and Dhakal (2021) argue that the dependencies of grassland drying and curing dynamics on SM are poorly represented in fire hazard models. Therefore, we believe that large-scale grassland data obtained from space observations together with the high capacity of AI in modeling nonlinear relationships between SM and wildfire is a promising combination for this goal. Although AI is not applied, Lu and Wei (2021) used in-situ SM data along with ESA ECV and SMAP SM data as a predictor of Live fuel moisture content which is a key variable in fire modeling. The same is also conducted by Jia et al. (2019) where they predicted live fuel moisture content applying a regression model along with SMAP L-band radiometer SM as input.

4. Future perspectives

The preceding sections have highlighted the well-established presence of AI in SM remote sensing, showing new avenues for exploration as big data is available and analysis ready on cloud computing environments. Example platforms are the public services ESA DIAS (Data and Information Access Services) for Copernicus data access (https://www.copernicus.eu/en/access-data/dias), NASA Earthdata Cloud (https://www.earthdata.nasa.gov/), the German CODE-DE (https://code-de.org) and EO-Lab (https://eo-lab.org), and the private company services like Google Earth Engine (https://earthengine.google.com) and Open Data on Amazon Web Services (https://registry.opendata.aws). It has often been reported that the performance of AI methods is superior to that of traditional (physical) methods, particularly where a full understanding of the underlying processes is lacking. The higher comparative performance of AI methods suggests that existing models, while valuable, may not fully exploit the available data, highlighting the potential benefits of harnessing the power of AI. However, the applications discussed often remain closely tied to individual research challenges, which have previously been addressed by (often simplified) physical models. In the broader context of geosciences, a concern highlighted by Tuia et al. (2021) is the tendency to stay within disciplinary comfort zones and apply recent AI advances primarily to well-resolved remote sensing problems. They emphasize the need to embrace AI technology in order to tackle new, challenging problems, ultimately generating added value from these data. To achieve this objective, Tuia et al. (2021) outline six research avenues, which we summarize in Fig. 6. These are: (i) developing AI models capable of deeper reasoning going beyond just recognizing objects and patterns which requires understanding and predicting relationships and processes occurring at different temporal and spatial scales, (ii) integrating multimodal data sources into AI workflows which can create models capable of combining traditional spatial data with new forms of information like real-time social media or satellite data, (iii) building AI models capable of active data query, hypothesis formulation, and insight generation in response to specific problems or changes in the environment, (iv) combining ML models with established domain principles (e.g., physics-based models) to ensure that AI-driven predictions align with natural laws or known theories, (v) focusing on developing interpretable models or methods that explain the AI's decisions in ways that are accessible to non-experts which can lead to improved trust in AI applications, and (vi) using advanced techniques like causal inference and counterfactual reasoning in AI models that can lead to identifying causal mechanisms, rather than just correlating variables.

Some of these research directions are already being researched by few studies in our field of research. Examples are hybrid AI-physical approaches considering the energy balance in ET estimation (Zhao et al., 2019) or the mass balance conservation in spatial scaling (Mao et al., 2022). The directions may still provide general guidance for further experiments.

However, there are still risks. The huge amount of freely available data is one of the driving forces of AI applications in SM remote sensing.

Fig. 6. Research avenues for AI in geosciences after Tuia et al. (2021).

However, implementing AI here bears the risk that this engineering, modeling and environmental issue is treated as a data science task only, neglecting the physical relationships of the microwave observation and their relationship to SM, as well as the Earth system processes where SM is only one compartment. We recognized few studies from AI specialists without knowledge in Earth system processes misinterpreting their results, or even not discussing and interpreting the physical-environmental origins and implications (Batchu et al., 2023). Here, better exchange and collaboration between data science and remote sensing experts could help. This issue has gained rising awareness that innovative approaches, such as fusing scientific knowledge and AI methods, require deep knowledge integration across disciplinary boundaries (Ebert-Uphoff et al., 2019). Even step-by-step guidelines have been developed for such collaborations (Ebert-Uphoff and Deng, 2017), and proposals for formal transdisciplinary educational opportunities (Pennington et al., 2020).

Moreover, conventional ML methods often incorporate built-in assumptions, such as the assumption that samples are fully independent and identically distributed. However, these assumptions can result in inaccurate estimates when they do not align with the characteristics of certain datasets (Varadharajan et al., 2022). Notably, the selection of a loss function can significantly impact the outcomes. In the case of environmental extreme events, with SM being fully saturated after rain events or dry after longer drought conditions, outliers can be equally important as normal conditions. The results may be affected if these conditions have happened before and are included in the training stage, or the event is extreme and a new record in the changing climate system. ML approaches need to account for this outlier fragility. For example, one can train exclusively on extreme events in historical time series, or adjust ML prediction bias to improve performance on the tails of the distribution (Varadharajan et al., 2022). Xie et al. (2021) implemented physical constraints into DL systems by the exterior terms imposed on the loss function.

In certain situations, AI models generate outputs or predictions that deviate from factual or realistic representations of environmental phenomena (Rawte et al., 2023). This can occur when AI models, particularly DL models, are trained on vast and diverse datasets, leading them to extrapolate or interpolate information in ways that may not align with actual physical processes. Reasons can also be noise or bias (data imperfection), or applications for circumstances the AI model was not trained for, such as extreme events. This is currently discussed under the term "AI hallucination" for the large language models such as ChatGPT. In Earth sciences, where accurate and reliable predictions are crucial for environmental understanding, this issue has already been addressed (Annau et al., 2023). Here, the occurrence of AI hallucination raises concerns about the model's interpretability and the potential for misleading insights. Addressing and mitigating AI hallucination in Earth sciences necessitate a careful examination of model training practices, dataset quality, and the incorporation of domain-specific knowledge (e.g., physical boundaries) to ensure that AI-generated outputs align with the known principles and behaviors of Earth systems.

Based on the discussion above, new challenges and opportunities arise, from which we draw the following recommendations for further development and research in the field of AI in SM remote sensing:

Recommendation 1: Utilize new data sources

A primal driving force for future AI applications in SM remote sensing are new missions. New spaceborne satellites and also new sensor development on other platforms (e.g., Unmanned Aerial Systems) will provide data at alternative spatial and temporal resolutions, space orbits, microwave frequencies, mission concepts, etc. Here, AI can be implemented in the same way as discussed in this review. Recent (ALOS-4, Advanced Land Observing Satellite-4, launched 1st July 2024) and soon upcoming missions like BIOMASS (launched April 2025), NiSAR (NASA-ISRO SAR, launched July 2025), CIMR (Copernicus Imaging Microwave Radiometer, 2025 +), and further planned satellites like ROSE-L (Radar Observing System for Europe in L-band, launch in 2028) extend the database available for SM studies. Similarly, future GNSS-R missions such as HydroGNSS (Unwin et al., 2021) (launch in 2025) will enable the enhancement of the data basis available for SM studies.

Recommendation 2: Provide more and better training data

For AI to work efficiently as a data driven method, the quantity and quality of input and learning data sets needs to be high, which is a problem in many geoscientific applications where variables are measured at limited sites (McDonnell and Beven, 2014). AI is susceptible to challenges such as data errors, incompleteness, out-of-sample predictions, and bias in inputs or training data. The quality of AI models is inherently constrained by the quantity, diversity, and quality of observations (Fang et al., 2022b; Read et al., 2019). Of course, this is also true for physical models, where input variables and model parameters require high quality observed data to ensure adequate accuracy and effective calibration. Purely data-driven AI models can at best replicate patterns in the training data, inheriting biases, inadequate spatiotemporal resolutions, and an inability to address non-stationarity or unforeseen extremes in time series. It is also possible to generate synthetic data, e.g. by utilizing hydrological models, probabilistic models, or GANs. The ISMN (Dorigo et al., 2011), known as the fundamental source of in situ SM data and widely accepted by the SM community (Montzka et al., 2020), shows uneven global coverage with focus on the USA and parts of Europe. Further measurements, especially in South America, Africa and Asia need to be performed and included in this database, to serve as an AI target variable for all climates, land cover and vegetation types, as well as soil characteristics. For example, using North American measurements to train for South American conditions may introduce large inaccuracies in the final SM products. ISMN data originate from various methods to measure SM at ground, e.g. capacitance sensors, time domain reflectometry, or cosmic ray neutron sensors, which have different characteristics, measurement volumes, etc. to be accounted for (Bogena et al., 2015; Montzka et al., 2017; Montzka et al., 2020). Moreover, the ISMN is currently targeting higher quality standards by characterizing the contributing network measurements as Fiducial Reference Measurements (FRMs) (Dorigo et al., 2021). This includes independent, fully characterized, and traceable ground measurements that follow specific guidelines, e.g. those of the GEO/CEOS Quality Assurance framework for Earth Observation (QA4EO), see also the Global Climate Observing System Implementation Plan (GCOS, 2022), as well as (Montzka et al., 2020) and (Gruber et al., 2020). This helps not only during validation of SM products, but also to provide accuracy information to the AI method for adequate predictions. Once the accuracy (of features and targets) is known and data is too noisy or erroneous, advanced pre-processing methods may be used for improvements. These can be generative modeling for data augmentation or training methods/models which emphasize more accurate training data.

Recommendation 3: Provide results faster

As SM is an important source of information for management actions, timely data delivery also to end-users is requested. Current workflows provide coarse global SM data within 1–2 days (e.g. SMAP and SMOS), but higher resolution data includes higher latency (e.g. $>$ 2 months for Copernicus Land Monitoring Service Sentinel-1 1 km). These data are addressed for specialists, and the hurdles for end-users are too high. Here, special technological solutions need to be developed to provide data in NRT, e.g. cloud-based online data integration methods as well as data dashboards and mobile phone apps, respectively. AI techniques can be used for such a purpose thanks to their low computational requirements (fast delivery of results) and easy-established adaptations to any information broadcasting environments. To generate real societal impact, standard data repositories of the space agencies need to be extended by NRT processing facilities and adequate interfaces for end-user applications.

Recommendation 4: Utilize hybrid approaches

Purely data-driven AI models are unable to predict untrained variables since they are designed to output only the training targets, while physics-based models are full of a multitude of simplifying assumptions to make the problem mathematically tractable. Thus, relying solely on either AI or physics-based models may not meet the requirements of SM remote sensing, suggesting that the development and benchmarking of hybrid methods, leveraging the best of both approaches, would be the best way forward (Reichstein et al., 2019). Hybrid methods containing data-driven AI methods in combination with physical models or mathematical theories may ensure control and explainability of simulations. Physical approaches are typically implemented in a simplified way (e.g., zeroth order radiative transfer models). Here, AI may provide the chance to be trained on more complex or generalized approaches, with fast delivery of results, there might be even generative approaches to establish high complexity models due to the nexus of AI with physics principles. One future direction could thus be to combine the strengths of a physical model like the Dubois Model that describes the interaction between radar signals and soil dielectric constant with an AI model that can learn complex patterns from data. The physical relationship can be implemented directly into a neural network's loss function, i.e., a loss function that not only minimizes prediction errors but also ensures that the network's outputs satisfy the underlying physical equations. Another example could be the estimation of deeper layer SM, which can be simulated by a combination of physical soil hydraulic modeling and an observation-based AI approach, e.g. by integrating the Richards equation calculating the movement of water in unsaturated soils accounting for soil properties, boundary and initial conditions. This could also improve the estimation of further hydrological variables, such as evapotranspiration.

Recommendation 5: Move from correlation to causality

AI algorithms rely on correlations rather than causality, making them prone to producing accurate results for incorrect reasons, especially in the presence of confounding factors. While causal representation learning and explainable AI methods show promise in overcoming this limitation, challenges persist in learning causality and interpretability (Bach et al., 2015; Schölkopf, 2022; Toms et al., 2020). Approaches that can adeptly query an AI model, incorporate causality and prior information, and pinpoint any missing physics within the model chain that could prove to be valuable, should be further developed and explored.

Recommendation 6: Think in larger systems

Present AI applications tend to replicate individual physical relationships or models. While risky, prospective AI implementations have the potential to broaden their scope, aiming to identify relationships across the entire intricate technological-environmental system. This expansion would move beyond the confined boundaries of sensor characteristics, all the way towards delving into how SM impacts various other compartments within the Earth system. Remotely sensed SM data is heavily used to estimate other compartments of the water cycle, also by implementing AI. However, there are few studies investigating the impact of SM on other environmental processes (e.g., carbon, energy and nutrient cycles). Here, more research is needed to include SM data (together with further auxiliary data) in estimating adjacent states and fluxes.

Recommendation 7: Exploit synergies from multimodal data

Future missions with alternative observation concepts (such as GNSS-R or multi-static SAR (Mittermayer et al., 2020)) may open avenues for AI implementation, once the physical principles and relationships are clarified. Similarly, with emerging new satellite missions, synergies between different missions may be exploited by AI methods to produce SM data at higher spatio-temporal accuracy. Some research fields are not yet fully explored by AI methods, especially the estimation of indirectly SM-related variables such as live fuel moisture content (Rao et al., 2020) or wildfire risks (Chaparro et al., 2016). Here, we still see high potential.

Recommendation 8: Keep track of AI innovations

Super-resolution, Domain Shifts, and Transfer Learning applications of AI in SM remote sensing are also limited to a few studies so far. Here, the diverse satellite missions, open mission concepts, modeling architectures create opportunities for novel research questions. Recently, foundation models gained interest in remote sensing applications (Hong et al., 2024; Siqi et al., 2024). These are large-scale AI models that are pre-trained on vast amounts of data across a wide range of tasks and then fine-tuned for specific applications. A global SM foundation model can be trained on ISMN data, and then applied for specific regions with fast adaptation to local characteristics. As AI is in general a field with high innovative and creational power, methods will be developed with large benefits for the SM remote sensing field.

Recommendation 9: Become interdisciplinary

As a remote sensing scientist with a specific focus on SM, it might be hard to keep track of and understand all innovations in the AI field. This includes also the correct implementation of an AI method for a SM remote sensing task. Similarly, data scientists may not understand the origin of the remotely-sensed data and the technological-physical characteristics to tailor AI approaches to the specific tasks need. Therefore, build small, but heterogeneous teams, including data scientists, Earth scientists, sensor engineers, and end-users to best address the problem and to provide adequate interpretation of AI generated results. This includes also transdisciplinary education and training.

Current interactive AI capabilities enable responses to queries like “What will the weather be like tomorrow in city XY?” This functionality is supported by the AI system’s online access to operational numerical weather prediction model results, incorporating assimilated observation data. Extending this capability to the realm of SM remote sensing, future scenarios envision a prompt or spoken command to a generative AI, such as “Generate a root zone SM map for district YZ at 10 m resolution for today and provide a forecast for the next week.” The prerequisites include an interconnected weather prediction system with a land surface scheme and high-resolution SM observations, all freely and fully accessible to AI systems. While each of these elements exists independently, the missing link is in effectively coupling these components. Such functionality integration holds the potential to significantly benefit farmers’ decision-making processes, seamlessly incorporating into farm management systems for tasks such as planning planting, irrigation, fertilizer applications, and harvests. Foundation models can be used here, which are AI models trained on a broad set of unlabeled data. This can accelerate the analysis of tremendous amounts of data, e.g. with Landsat and Sentinel-2 in NASA’s first open-source geospatial AI foundation model for Earth Observation data. See also the discussion in Osco et al. (2023) for the integration of ChatGPT’s large language model capabilities with visual computation to enable effective image analysis in Visual ChatGPT.

In the introduction we opened the initial inquiries regarding the role of AI in conjunction with or as a potential substitute for physical or statistical approaches, and the necessity for research aimed at a methodological transformation wherein AI and remote sensing observations merge into a learning, self-validating, and interpretable hybrid system. The discourse above illustrates that initial strides in this direction have been taken. However, both the six overarching research avenues proposed by Tuia et al. (2021) and our more detailed recommendations for future research in the intersection of AI and SM remote sensing indicate the emergence of new opportunities for our already vibrant and innovative community. While this review on AI applications in SM remote sensing offers a valuable overview of current advances, it also faces several limitations. Comparisons across studies are often hindered by inconsistent datasets, metrics, and regional biases, limiting generalizability. Moreover, an overemphasis on performance metrics can overshadow concerns around interpretability, physical consistency, and real-world applicability. Many reviews underrepresent failed or inconclusive results, and few address interdisciplinary integration or the societal implications of AI-driven SM products.

5. Conclusions

The integration of AI into SM remote sensing heralds a transformative era in understanding and managing a critical component of the hydrological cycle. As a linchpin in land surface exchange processes, SM’s significance in climate change research cannot be overstated. This review article has provided a comprehensive exploration of AI applications in SM remote sensing, ranging from time series reconstruction, RZSM estimation, spatial scaling, forecasting, and the chance to get information about further elements of the hydrological cycle or related processes. The unique strength of AI, rooted in its capacity to discern data-driven relationships without predetermined assumptions, positions it as a formidable tool in overcoming challenges traditionally faced by statistical and physical models. The current strides made in addressing these challenges pave the way for an exciting future, offering promising research avenues that hold the potential to unlock new insights into the intricate dynamics of the hydrological cycle and associated processes.

To advance the field, future research must move beyond well-established use cases and embrace AI for solving unexplored, complex geoscientific challenges. This includes developing interpretable models that incorporate physical laws, expanding multimodal data integration, and ensuring models are robust to extreme events and data uncertainties. Notably, the need for better training data, quality control, and interdisciplinary collaboration is emphasized. We identify several key recommendations for future progress: addressing new satellite missions and sensors, improving the availability and representativeness of training data, ensuring near real-time data delivery, embracing hybrid AI-physical modeling, focusing on causality rather than correlation, and extending AI applications to system-level analyses. Keeping pace with rapid AI innovations, such as foundation models and generative approaches, will be crucial. Most importantly, fostering interdisciplinary teams across remote sensing, data science, and domain expertise is essential to harness AI’s full potential while preserving scientific rigor and interpretability. As we navigate this dynamic landscape, AI emerges as a key ally in advancing our comprehension and harnessing the full potential of SM data for meaningful environmental impact and hydrological process understanding.

Carsten Montzka reports financial support was provided by Forschungszentrum Jülich Institute of Bio- and Geosciences Agrosphere 3. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Carsten Montzka: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Luca Brocca:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Hao Chen:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Narendra N. Das:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Antara Dasgupta:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Mehdi Rahmati:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Thomas Jagdhuber:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Carsten Montzka reports financial support was provided by Forschungszentrum Jülich, Institute of Bio- and Geosciences: Agrosphere (IBG-3). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge funding by the German Ministry of Economic Affairs and Climate (BMWK) through the German Aerospace Center for the AssimEO project (50EE1914A), by the German Federal Ministry for Digital and Transport (BMDV) for the HABKIS2 project (19F2264B), by the European Space Agency for the SARWATER project (ESA Contract No. 4000148119/25/I-EB-bgh), by the National Natural Science Foundation of China (42101034), and by the European Union Open Earth Monitor Cyberinfrastructure (OEMC) project (grant agreement No. 101059548).

Data availability

No data was used for the research described in the article.

References

A, Y., Wang, G., Hu, P., Lai, X., Xue, B. and Fang, Q., 2022. Root-zone soil moisture estimation based on remote sensing data and deep learning. Environmental Research, 212: 113278.

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., Ghalhari, G.A., 2020. Machine learning to estimate surface soil moisture from remote sensing data. Water.

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: a machine learning approach. Adv. Water Resour. 33 (1), 69–80.

Al Bitar, A. and Mahmoodi, A., 2020. Algorithm Theoretical Basis Document (ATBD) for the SMOS Level 4 Root Zone Soil Moisture.

Alam, A.M., Kurum, M., Gurbuz, A.C., 2022. Radio frequency interference detection for SMAP radiometer using convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 10099–10112.

Alemohammad, S.H., Jagdhuber, T., Moghaddam, M., Entekhabi, D., 2019. Soil and vegetation scattering contributions in l-band and p-band polarimetric SAR observations. IEEE Trans. Geosci. Remote Sens. 57 (11), 8417–8429.

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. Remote Sens. (Basel) 16398–16421.

Amer, A., Yun, Z., Sue, N., 2011. Review and evaluation of remote sensing methods for soil-moisture estimation. SPIE Rev. 2 (1), 028001.

Annau, N.J., Cannon, A.J. and Monahan, A.H., 2023. Algorithmic Hallucinations of Near-Surface Winds: Statistical Downscaling with Generative Adversarial Networks to Convection-Permitting Scales.

Ayari, E., Kassouk, Z., Lili-Chabaane, Z., Baghdadi, N., Zribi, M., 2022. Investigation of multi-frequency SAR data to retrieve the soil moisture within a drip irrigation context using modified water cloud model. Sensors 22 (2).

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M., 2021. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. Remote Sens. Environ. 260, 112434.

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M., 2022. Short- and mid-term forecasts of actual evapotranspiration with deep learning. J. Hydrol. 612, 128078.

Babaeian, E., Sadeghi, M., Jones, S.B., Montzka, C., Vereecken, H., Tuller, M., 2019. Ground, proximal and satellite remote sensing of soil moisture. Rev. Geophys. 57, 530–616.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10 (7).

Balenzano, A., Mattia, F., Satalino, G., Lovergine, F.P., Palmisano, D., Peng, J., Marzahn, P., Wegmuller, U., Cartus, O., Dabrowska-Zielinska, K., Musial, J.P., Davidson, M.W.J., Pauwels, V.R.N., Cosh, M.H., McNairn, H., Johnson, J.T., Walker, J.P., Yueh, S.H., Entekhabi, D., Kerr, Y.H. and Jackson, T.J., 2021. Sentinel-1 soil moisture at 1 km resolution: a validation study. Remote Sensing of Environment, 263: ARTN 112554.

Basak, A., Schmidt, K.M., Mengshoel, O.J., 2023. From data to interpretable models: machine learning for soil moisture forecasting. Int. J. Data Sci. Anal. 15 (1), 9–32.

Batchu, V., Nearing, G., Gulshan, V., 2023. A deep learning data fusion model using sentinel-1/2, soilgrids, SMAP, and GLDAS for soil moisture retrieval. J. Hydrometeorol. 24 (10), 1789–1823.

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massario, C., Ciabatta, L., Brocca, L., Wagner, W., 2019. Toward global soil moisture monitoring with sentinel-1: harnessing assets and overcoming obstacles. IEEE Trans. Geosci. Remote Sens. 57 (1), 520–539.

Bauer-Marschallinger, B., Paulik, C., Hochstoger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., Wagner, W., 2018. Soil moisture from fusion of scatterometer and SAR: closing the scale gap with temporal filtering. Remote Sens. (Basel) 10 (7).

Bentivoglio, R., Isufi, E., Jonkman, S.N., Taormina, R., 2022. Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrol. Earth Syst. Sci. 26 (16), 4345–4378.

Berg, A., Sheffield, J., 2018. Climate change and drought: the soil moisture perspective. Curr. Clim. Change Rep. 4 (2), 180–191.

Beven, K., 2020. Deep learning, hydrological processes and the uniqueness of place. Hydrol. Process. 34 (16), 3608–3613.

Boas, T., Bogena, H.R., Ryu, D., Vereecken, H., Western, A., Hendricks Franssen, H.J., 2023. Seasonal soil moisture and crop yield prediction with fifth-generation seasonal forecasting system (SEAS5) long-range meteorological forecasts in a land surface modelling approach. Hydrol. Earth Syst. Sci. 27 (16), 3143–3167.

Bogena, H.R., Huisman, J.A., Hübner, C., Kusche, J., Jonard, F., Vey, S., Güntner, A., Vereecken, H., 2015. Emerging methods for non-invasive sensing of soil moisture dynamics from field to catchment scale: a review. WIREs Water 2 (6), 635–647.

Bolten, J.D., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A., 2010. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3 (1), 57–66.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brocca, L., Ciabatta, L., Massari, C., Camici, S., Tarpanelli, A., 2017. Soil moisture for hydrological applications: open questions and new opportunities. Water 9 (2), 140.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Hasenauer, S., 2010. ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy. Remote Sens. Environ. 114 (11), 2745–2755.

Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2017. Geometric deep learning: going beyond euclidean data. IEEE Signal Process Mag. 34 (4), 18–42.

Brust, C., Kimball, J.S., Maneta, M.P., Jencso, K., Reichle, R.H., 2021. DroughtCast: a machine learning forecast of the United States drought monitor. Front. Big Data 4.

Cai, Y., Zheng, W.G., Zhang, X., Zhangzhong, L.L., Xue, X.Z., 2019. Research on soil moisture prediction model based on deep learning. PLoS One 14 (4).

Carranza, C., Nolet, C., Pezij, M., van der Ploeg, M., 2021. Root zone soil moisture estimation with random forest. J. Hydrol. 593.

Chakraborty, D., Başağaoğlu, H., Alian, S., Mirchi, A., Moriasi, D.N., Starks, P.J., Verser, J.A., 2023. Multiscale extrapolative learning algorithm for predictive soil moisture modeling & applications. Expert Syst. Appl. 213, 119056.

Chaparro, D., Piles, M., Vall-llossera, M., 2016. Chapter 13 - remotely sensed soil moisture as a key variable in wildfires prevention services: towards new prediction tools using SMOS and SMAP data. In: Srivastava, P.K., Petropoulos, G.P., Kerr, Y.H. (Eds.), Satellite Soil Moisture Retrieval. Elsevier, pp. 249–269.

Chatterjee, S., Desai, A.R., Zhu, J., Townsend, P.A., Huang, J.Y., 2022. Soil moisture as an essential component for delineating and forecasting agricultural rather than meteorological drought. Remote Sens. Environ. 269.

Chaudhary, S.K., Srivastava, P.K., Gupta, D.K., Kumar, P., Prasad, R., Pandey, D.K., Das, A.K., Gupta, M., 2022. Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation. Adv. Space Res. 69 (4), 1799–1812.

Chen, Q., Tang, X., Li, B., Tang, Z., Miao, F., Song, G., Yang, L., Wang, H., Zeng, Q., 2023. Spatial downscaling of soil moisture based on fusion methods in complex terrains. Remote Sens. (Basel).

Chen, Q.Q., Miao, F., Wang, H., Xu, Z.X., Tang, Z.Y., Yang, L., Qi, S.X., 2020. Downscaling of satellite remote sensing soil moisture products over the tibetan plateau based on the random forest algorithm: preliminary results. Earth Space Sci. 7 (6).

Chen, Y.Z., Feng, X.M., Fu, B.J., 2021. An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003-2018. Earth Syst. Sci. Data 13 (1), 1–31.

Cheng, Y., Liu, Q.Y., Wang, J., Wan, S.H., Umer, T., 2018. Distributed fault detection for wireless sensor networks based on support vector regression. Wireless Commun. & Mobile Comput.

Cho, E., Jacobs, J.M., Jia, X., Kraatz, S., 2019. Identifying subsurface drainage using satellite big data and machine learning via google earth engine. Water Resour. Res. 55 (10), 8028–8045.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A., 2018. Generative adversarial networks: an overview. IEEE Signal Process Mag. 35 (1), 53–65.

Cui, Y., Chen, X., Xiong, W., He, L., Lv, F., Fan, W., Luo, Z., Hong, Y., 2020. A soil moisture spatial and temporal resolution improving algorithm based on multi-source remote sensing data and GRNN model. Remote Sens. (Basel).

Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F., 2022. Scientific machine learning through physics–informed neural networks: where we are and What's next. J. Sci. Comput. 92 (3), 88.

Daly, E., Porporato, A., 2005. A review of soil moisture dynamics: from rainfall infiltration to ecosystem response. Environ. Eng. Sci. 22 (1), 9–24.

Dari, J., Brocca, L., Quintana-Seguí, P., Escorihuela, M.J., Stefan, V., Morbidelli, R., 2020. exploiting high-resolution remote sensing soil moisture to estimate irrigation water amounts over a mediterranean region. Remote Sens. (Basel).

Das, B., Rathore, P., Roy, D., Chakraborty, D., Jatav, R.S., Sethi, D., Kumar, P., 2022. Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies. Catena 217.

Das, N.N., Entekhabi, D., Dunbar, R.S., Chaubell, M.J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P.E., Walker, J.P., Berg, A., Bosch, D.D., Caldwell, T., Cosh, M.H., Collins, C.H., Lopez-Baeza, E., Thibeault, M., 2019. The SMAP and copernicus sentinel 1A/B microwave active-passive high resolution surface soil moisture product. Remote Sens. Environ. 233, 111380.

Das, N.N., Entekhabi, D., Njoku, E.G., 2011. An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. IEEE Trans. Geosci. Remote Sens. 49 (5), 1504–1512.

Das, N.N., Mohanty, B.P., 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone J. 5 (1), 296–307.

Datta, S., Das, P., Dutta, D., Giri, R.K., 2021. Estimation of surface moisture content using sentinel-1 c-band SAR data through machine learning models. J. Indian Soc. Remote Sens. 49 (4), 887–896.

Deines, J.M., Kendall, A.D., Hyndman, D.W., 2017. Annual irrigation dynamics in the U. S. northern high plains derived from landsat satellite data. Geophys. Res. Lett. 44 (18), 9350–9360.

Dobson, M.C., Ulaby, F.T., Hallikainen, M.T., Elrayes, M.A., 1985. Microwave dielectric behavior of wet soil .2. dielectric mixing models. IEEE Trans. Geosci. Remote Sens. 23 (1), 35–46.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.C., Camarero, J.J., Capello, G., Choi, M., Cosh, M.C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K.H., Kanniah, K.D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M.A., Pellarin, T., Petropoulos, G.P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J.P.,

Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., Sabia, R., 2021. The international soil moisture network: serving earth system science for over a decade. Hydrol. Earth Syst. Sci. 25 (11), 5749–5804.
Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P.D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S.I., Smolander, T., Lecomte, P., 2017. ESA CCI soil moisture for improved earth system understanding: state-of-the art and future directions. Remote Sens. Environ. 203, 185–215.
Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., Jackson, T., 2011. The international soil moisture network: a data hosting facility for global in situ soil moisture measurements. Hydrol. Earth Syst. Sci. 15 (5), 1675–1698.
Dramsch, J.S., 2020. Chapter one - 70 years of machine learning in geoscience in review. In: Moseley, B., Krischer, L. (Eds.), Advances in Geophysics. Elsevier, pp. 1–55.
Dubois, A., Teytaud, F., Verel, S., 2021. Short term soil moisture forecasts for potato crop farming: a machine learning approach. Comput. Electron. Agric. 180.
Dubois, P.C., Vanzyl, J., Engman, T., 1995. Measuring soil-moisture with imaging radars. IEEE Trans. Geosci. Remote Sens. 33 (4), 915–926.
Ebert-Uphoff, I. and Deng, Y., 2017. Three steps to successful collaboration with data scientists. EOS, 89.
Ebert-Uphoff, I., Samarasinghe, S.M., Barnes, E.A., 2019. Thoughtfully using artificial intelligence in Earth science. Eos 100.
Edokossi, K., Calabia, A., Jin, S., Molina, I., 2020. GNSS-reflectometry and remote sensing of soil moisture: a review of measurement techniques, methods, and applications. Remote Sens. (Basel) 12 (4), 614.
El Hajj, M., Baghdadi, N., Bazzi, H., Zribi, M., 2019. Penetration analysis of SAR signals in the c and l bands for wheat, maize, and grasslands. Remote Sens. (Basel) 11 (1), 31.
ElSaadani, M., Habib, E., Abdelhameed, A.M., Bayoumi, M., 2021. Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. Front. Artif. Intell. 4.
Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98 (5), 704–716.
Escorihuela, M.J., Chanzy, A., Wigneron, J.P., Kerr, Y.H., 2010. Effective soil moisture sampling depth of L-band radiometry: a case study. Remote Sens. Environ. 114 (5), 995–1001.
Esit, M., Kumar, S., Pandey, A., Lawrence, D.M., Rangwala, I., Yeager, S., 2021. Seasonal to multi-year soil moisture drought forecasting. npj Clim. Atmos. Sci. 4 (1), 16.
Etminan, A., Tabatabaeenejad, A., Moghaddam, M., 2020. Retrieving root-zone soil moisture profile from p-band radar via hybrid global and local optimization. IEEE Trans. Geosci. Remote Sens. 58 (8), 5400–5408.
Fang, B., Lakshmi, V., Bindlish, R., Jackson, T.J., 2018. AMSR2 soil moisture downscaling using temperature and vegetation data. Remote Sens. (Basel) 10 (10).
Fang, B., Lakshmi, V., Cosh, M., Liu, P.-W., Bindlish, R., Jackson, T.J., 2022a. A global 1-km downscaled SMAP soil moisture product based on thermal inertia theory. Vadose Zone J. 21 (2), e20182.
Fang, K., Kifer, D., Lawson, K., Feng, D.P., Shen, C.P., 2022b. The data synergy effects of time-series deep learning models in hydrology. Water Resour. Res. 58 (4).
Fang, K., Shen, C., 2020. Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel. J. Hydrometeorol. 21 (3), 399–413.
Fang, K., Shen, C.P., Kifer, D., Yang, X., 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. Geophys. Res. Lett. 44 (21), 11030–11039.
Fernández-López, A., Marín-Sánchez, D., García-Mateos, G., Ruiz-Canales, A., Ferrández-Villena-García, M., Molina-Martínez, J.M., 2020. A machine learning method to estimate reference evapotranspiration using soil moisture sensors. Appl. Sci.
Fluhrer, A., Jagdhuber, T., Tabatabaeenejad, A., Alemohammad, H., Montzka, C., Friedl, P., Forootan, E., Kunstmann, H., 2022. Remote sensing of complex permittivity and penetration depth of soils using p-band SAR polarimetry. Remote Sens. (Basel) 14 (12), 2755.
Foroumandi, E., Moradkhani, H., Sanchez-Vila, X., Singha, K., Castelletti, A., Destouni, G., 2023a. ChatGPT in hydrology and earth sciences: opportunities, prospects, and concerns. Water Resour. Res. 59 (10), e2023WR036288.
Foroumandi, E., Nourani, V., Huang, J.J., Moradkhani, H., 2023b. Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: a deep learning approach. J. Hydrol. 616.
Frame, J.M., Kratzert, F., Gupta, H.V., Ullrich, P., Nearing, G.S., 2023. On strictly enforced mass conservation constraints for modelling the rainfall-runoff process. Hydrol. Process. 37 (3), e14847.
Fu, K., Chang, Z.H., Zhang, Y., Xu, G.L., Zhang, K.S., Sun, X., 2020. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. ISPRS J. Photogramm. Remote Sens. 161, 294–308.
Fung, A.K., Chen, K.S., 2004. An update on the IEM surface backscattering model. IEEE Geosci. Remote Sens. Lett. 1 (2), 75–77.
GCOS, 2022. The 2022 GCOS Implementation Plan, Geneva, Switzerland.
Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. Soil moisture prediction using support vector machines. J. Am. Water Resour. Assoc. 42 (4), 1033–1046.
Gökkaya, K., Budhathoki, M., Christopher, S.F., Hanrahan, B.R., Tank, J.L., 2017. Subsurface tile drained area detection using GIS and remote sensing in an agricultural watershed. Ecol. Eng. 108, 370–379.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative Adversarial Nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence and K.Q. Weinberger (Editors).
Govindaraju, R.S., Artific, A.T.C.A., 2000. Artificial neural networks in hydrology. Ii: hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.
Green, J.K., Seneviratne, S.I., Berg, A.M., Findell, K.L., Hagemann, S., Lawrence, D.M., Gentine, P., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. Nature 565 (7740), 476.
Greifeneder, F., Notarnicola, C., Wagner, W., 2021. A machine learning-based approach for surface soil moisture estimations with google earth engine. Remote Sens. (Basel).
Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., Schalie, R.v.d., Wigneron, J.-P. and Wagner, W., 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors? Remote Sensing of Environment, 244: 111806.
Grunwald, S., 2022. Artificial intelligence and soil carbon modeling demystified: power, potentials, and perils. Carbon Footprints 1 (1), 5.
Gu, Z., Zhu, T., Jiao, X., Xu, J., Qi, Z., 2021. Neural network soil moisture model for irrigation scheduling. Comput. Electron. Agric. 180, 105801.
Han, Q., Zeng, Y., Zhang, L., Wang, C., Prikaziuk, E., Niu, Z., Su, B., 2023. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. Sci. Data 10 (1), 101.
He, L., Hong, Y., Wu, X.L., Ye, N., Walker, J.P., Chen, X.N., 2018. Investigation of SMAP active-passive downscaling algorithms using combined sentinel-1 SAR and SMAP radiometer data. IEEE Trans. Geosci. Remote Sens. 56 (8), 4906–4918.
Heddam, S., 2021. New formulation for predicting soil moisture content using only soil temperature as predictor: Multivariate adaptive regression splines versus random forest, multilayer perceptron neural network, M5Tree, and multiple linear regression, Water Engineering Modeling and Mathematic Tools. Elsevier, Amsterdam, pp. 45-62.
Hegazi, E.H., Yang, L., Huang, J., 2021. A convolutional neural network algorithm for soil moisture prediction from sentinel-1 SAR images. Remote Sens. (Basel) 13 (24), 4964.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J., 2024. SpectralGPT: spectral remote sensing foundation model. IEEE Trans. Pattern Anal. Mach. Intell. 46 (8), 5227–5244.
Hu, F.M., Wei, Z.S., Zhang, W., Dorjee, D., Meng, L.K., 2020. A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. J. Hydrol. 590.
Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S.I., Frankenberg, C., 2021. Soil moisture-atmosphere feedback dominates land carbon uptake variability. Nature 592 (7852), 65.
Im, J., Park, S., Rhee, J., Baik, J., Choi, M., 2016. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. Environ. Earth Sci. 75 (15).
Inoubli, R., Bennaceur, L., Jarray, N., Ben Abbes, A., Farah, I.R., 2022. A comparison between the use of machine learning techniques and the water cloud model for the retrieval of soil moisture from Sentinel-1A and Sentinel-2A products. Remote Sens. Lett. 13 (10), 980–990.
Irrgang, C., Boers, N., Sonnewald, M., Barnes, E.A., Kadow, C., Staneva, J., Saynisch-Wagner, J., 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. Nat. Mach. Intell. 3 (8), 667–674.
Jahan, N. and Gan, T.Y., 2014. Soil Moisture Retrieval from Microwave (RADARSAT-2) and Optical Remote Sensing (MODIS) Data Using Artificial Intelligence Techniques. In: V. Lakshmi, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. van Oevelen, T. Painter, J. Parajka, M. Rodell and C. Rüdiger (Editors), Remote Sensing of the Terrestrial Water Cycle. John Wiley & Sons, Inc;, Hoboken, NJ, pp. 255-275.
Jamali, A., Roy, S.K., Hashemi Beni, L., Pradhan, B., Li, J., Ghamisi, P., 2024. Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. Int. J. Appl. Earth Obs. Geoinf. 127, 103662.
JAXA, 2013. Descriptions of GCOM-W1 AMSR2 Level 1R and Level 2 Algorithms.
Jia, S., Kim, S.H., Nghiem, S.V. and Kafatos, M., 2019. Estimating Live Fuel Moisture Using SMAP L-Band Radiometer Soil Moisture for Southern California, USA, Remote Sensing.
Jiang, M., Shen, H., Li, J., 2022. Cycle GAN based heterogeneous spatial-spectral fusion for soil moisture downscaling. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 4819–4822.
Jin, Y., Ge, Y., Liu, Y.J., Chen, Y.H., Zhang, H.T., Heuvelink, G.B.M., 2021. A machine learning-based geostatistical downscaling method for coarse-resolution soil moisture products. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 1025–1037.
Joshi, R.C., Ryu, D., Lane, P.N.J., Sheridan, G.J., 2023. Seasonal forecast of soil moisture over Mediterranean-climate forest catchments using a machine learning approach. J. Hydrol. 619.
Kaheil, Y.H., Gill, M.K., Mckee, M., Bastidas, L.A., Rosero, E., 2008. Downscaling and assimilation of surface soil moisture using ground truth measurements. IEEE Trans. Geosci. Remote Sens. 46 (5), 1375–1384.

Kara, A., Pekel, E., Ozcetin, E., Yıldız, G.B., 2024. Genetic algorithm optimized a deep learning method with attention mechanism for soil moisture prediction. Neural Comput. & Applic. 36 (4), 1761–1772.
Karandish, F., Šimůnek, J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. J. Hydrol. 543, 892–909.
Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266.
Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49.
Kerr, Y.H., Waldteufel, P., Richaume, P., Davenport, I., Ferrazzoli, P. and Wigneron, J.-P., 2010. SMOS Level 2 Processor Soil Moisture ATBD. Toulouse SO-TN-ESL-SM-GS-0001, 24/10/2010, 2010.
Keskin, H., Grunwald, S., Harris, W.G., 2019. Digital mapping of soil carbon fractions with machine learning. Geoderma 339, 40–58.
Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53 (8), 5455–5516.
Kornelsen, K.C., Coulibaly, P., 2014. Root- zone soil moisture estimation using data-driven methods. Water Resour. Res. 50 (4), 2946–2962.
Koster, R.D., Mahanama, S.P.P., Livneh, B., Lettenmaier, D.P., Reichle, R.H., 2010. Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. Nat. Geosci. 3 (9), 613–616.
Kwon, Y., Kumar, S.V., Navari, M., Mocko, D.M., Kemp, E.M., Wegiel, J.W., Geiger, J.V., Bindlish, R., 2022. Irrigation characterization improved by the direct use of SMAP soil moisture anomalies within a data assimilation system. Environ. Res. Lett. 17 (8).
Lal, P., Shekhar, A., Gharun, M., Das, N.N., 2023a. Spatiotemporal evolution of global long-term patterns of soil moisture. In: Science of the Total Environment, p. 867.
Lal, P., Singh, G., Das, N.N., Entekhabi, D., Lohman, R., Colliander, A., Pandey, D.K., Setia, R.K., 2023b. A multi-scale algorithm for the NISAR mission high-resolution soil moisture product. Remote Sens. Environ. 295.
Lamichhane, M., Mehan, S., Mankin, K.R., 2025. Soil moisture prediction using remote sensing and machine learning algorithms: a review on progress, challenges, and opportunities. Remote Sens. (Basel) 17 (14).
Le Cun, Y.a.B., B. and Denker, J. S. and Howard, R. E. and Habbard, W. and Jackel, L. D. and Henderson, D., 1990. Handwritten Digit Recognition with a Back-Propagation Network. Advances in Neural Information Processing Systems: 396–404.
Lee, J., Park, S., Im, J., Yoo, C., Seo, E., 2022. Improved soil moisture estimation: synergistic use of satellite observations and land surface models over CONUS based on machine learning. J. Hydrol. 609.
Li, J.H., Wang, S.S., Gunn, G., Joosse, P., Russell, H.A.J., 2018. A model for downscaling SMOS soil moisture using Sentinel-1 SAR data. Int. J. Appl. Earth Obs. Geoinf. 72, 109–121.
Li, L., Dai, Y.J., Shangguan, W., Wei, N., Wei, Z.W., Gupta, S., 2022a. Multistep forecasting of soil moisture using spatiotemporal deep encoder-decoder networks. J. Hydrometeorol. 23 (3), 337–350.
Li, L., Dai, Y.J., Shangguan, W., Wei, Z.W., Wei, N., Li, Q.L., 2022b. Causality-structured deep learning for soil moisture predictions. J. Hydrometeorol. 23 (8), 1315–1331.
Li, Z.-L., Leng, P., Zhou, C., Chen, K.-S., Zhou, F.-C., Shang, G.-F., 2021. Soil moisture retrieval from remote sensing measurements: current knowledge and directions for the future. Earth Sci. Rev. 218, 103673.
Lievens, H., Reichle, R.H., Liu, Q., De Lannoy, G.J.M., Dunbar, R.S., Kim, S.B., Das, N.N., Cosh, M., Walker, J.P., Wagner, W., 2017. Joint sentinel-1 and SMAP data assimilation to improve soil moisture estimates. Geophys. Res. Lett. 44 (12), 6145–6153.
Liu, D., Mishra, A.K., Yu, Z.B., Yang, C.G., Konapala, G., Vu, T., 2017. Performance of SMAP, AMSR-E and LAI for weekly agricultural drought forecasting over continental United States. J. Hydrol. 553, 88–104.
Liu, K., Li, X., Wang, S., Zhang, H., 2023. A robust gap-filling approach for European space agency climate change initiative (ESA CCI) soil moisture integrating satellite observations, model-driven knowledge, and spatiotemporal machine learning. Hydrol. Earth Syst. Sci. 27 (2), 577–598.
Liu, Y., Xin, Y., Yin, C., 2025. A Transformer-based method to simulate multi-scale soil moisture. J. Hydrol. 655, 132900.
Liu, Y., Yang, Y., 2022. Advances in the quality of global soil moisture products: a review. Remote Sens. (Basel) 14 (15), 3741.
Liu, Y., Yang, Y., Jing, W., Yue, X., 2018. Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over northeast China. Remote Sens. (Basel).
Llamas, R.M., Guevara, M., Rorabaugh, D., Taufer, M., Vargas, R., 2020. Spatial gap-filling of ESA CCI satellite-derived soil moisture based on geostatistical techniques and multiple regression. Remote Sens. (Basel).
Lu, Y., Wei, C., 2021. Evaluation of microwave soil moisture data for monitoring live fuel moisture content (LFMC) over the coterminous United States. Sci. Total Environ. 771, 145410.
Lundberg, S.M. and Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (Editors).
Luo, Q., Liang, Y., Guo, Y., Liang, X., Ren, C., Yue, W., Zhu, B., Jiang, X., 2023. Enhancing spatial resolution of GNSS-R soil moisture retrieval through XGBoost algorithm-based downscaling approach: a case study in the southern United States. Remote Sens. (Basel).
Ma, Y., Chen, S., Ermon, S., Lobell, D.B., 2024. Transfer learning in environmental remote sensing. Remote Sens. Environ. 301, 113924.
Ma, Y.L., Montzka, C., Bayat, B., Kollet, S., 2021. An indirect Approach based on long short-term memory networks to estimate groundwater table depth anomalies across Europe with an application for drought analysis. Front. Water 3.
Ma, Y.L., Montzka, C., Naz, B.S. and Kollet, S., 2022. Advancing AI-based pan-European groundwater monitoring. Environmental Research Letters, 17(11): 114037.
Madhukumar, N., Wang, E., Everingham, Y., Xiang, W., 2024. Hybrid transformer network for soil moisture estimation in precision irrigation. IEEE Access 12, 48898–48909.
Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans. Geosci. Remote Sens. 55 (2), 645–657.
Majumdar, S., Smith, R., Butler, J.J., Lakshmi, V., 2020. Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning. Water Resour. Res. 56 (11).
Mamalakis, A., Ebert-Uphoff, I., Barnes, E.A., 2022. Explainable artificial intelligence in meteorology and climate science: model fine-tuning, calibrating trust and learning new science. In: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K.-.-R., Samek, W. (Eds.), xxAI - beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers. Springer International Publishing, Cham, pp. 315–339.
Manfreda, S., Brocca, L., Moramarco, T., Melone, F., Sheffield, J., 2014. A physically based approach for the estimation of root-zone soil moisture from surface measurements. Hydrol. Earth Syst. Sci. 18 (3), 1199–1212.
Mao, T.N., Shangguan, W., Li, Q.L., Li, L., Zhang, Y., Huang, F.N., Li, J.D., Liu, W., Zhang, R.Q., 2022. A spatial downscaling method for remote sensing soil moisture based on random forest considering soil moisture memory and mass conservation. Remote Sens. (Basel) 14 (16).
Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A. and Herrero-Jiménez, C.M., 2016. Satellite soil moisture for agricultural drought monitoring: Assessment of the SMOS derived Soil Water Deficit Index (vol 177, pg 277, 2016). Remote Sensing of Environment, 183: 368-368.
Maskey, M., Alemohammad, H., Murphy, K.J., Ramachandran, R., 2020. Advancing AI for Earth science: a data systems perspective. Eos 101.
McDonnell, J.J., Beven, K., 2014. Debates-the future of hydrological sciences: a (common) path forward? a call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. Water Resour. Res. 50 (6), 5342–5350.
Mengen, D., Jagdhuber, T., Balenzano, A., Mattia, F., Vereecken, H., Montzka, C., 2023. High Spatial and temporal soil moisture retrieval in agricultural areas using multi-orbit and vegetation adapted sentinel-1 SAR time series. Remote Sens. (Basel) 15 (9), 2282.
Merlin, O., Malbéteau, Y., Notfi, Y., Bacon, S., Khabba, S.-E.-R.-S., Jarlan, L., 2015. Performance metrics for soil moisture downscaling methods: application to dispatch data in central Morocco. Remote Sens. (Basel) 7 (4), 3783–3807.
Merlin, O., Walker, J.P., Chehbouni, A., Kerr, Y., 2008. Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. Remote Sens. Environ. 112 (10), 3935–3946.
Meyer, R., Zhang, W., Kragh, S.J., Andreasen, M., Jensen, K.H., Fensholt, R., Stisen, S., Looms, M.C., 2022. Exploring the combined use of SMAP and Sentinel-1 data for downscaling soil moisture beyond the 1 km scale. Hydrol. Earth Syst. Sci. 26 (13), 3337–3357.
Mi, P., Zheng, C., Jia, L., Bai, Y., 2023. Reconstruction of global long-term gap-free daily surface soil moisture from 2002 to 2020 based on a pixel-wise machine learning method. Remote Sens. (Basel).
Ming, W., Ji, X., Zhang, M., Li, Y., Liu, C., Wang, Y., Li, J., 2022. A hybrid triple collocation-deep learning approach for improving soil moisture estimation from satellite and model-based data. Remote Sens. (Basel).
Mironov, V.L., Kosolapova, L.G., Lukin, Y.I., Karavaysky, A.Y., Molostov, I.P., 2017. Temperature- and texture-dependent dielectric model for frozen and thawed mineral soils at a frequency of 1.4 GHz. Remote Sens. Environ. 200, 240–249.
Mishra, V., Ellenburg, W.L., Markert, K.N., Limaye, A.S., 2020. Performance evaluation of soil moisture profile estimation through entropy-based and exponential filter models. Hydrol. Sci. J. 65 (6), 1036–1048.
Mittermayer, J., Krieger, G., Moreira, A., 2020. Concepts and applications of multi-static MirrorSAR systems. In: 2020 IEEE Radar Conference (RadarConf20), pp. 1–6.
Mladenova, I.E., Bolten, J.D., Crow, W., Sazib, N., Reynolds, C., 2020. Agricultural drought monitoring via the assimilation of SMAP soil moisture retrievals into a global soil water balance model. Front. Big Data 3.
Mohammed, P.N., Piepmeier, J.R., 2021. Microwave radiometer RFI detection using deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6398–6405.
Mohanty, B.P., Cosh, M.H., Lakshmi, V., Montzka, C., 2017. Soil moisture remote sensing: state-of-the-science. Vadose Zone J. 16 (1).
Mohseni, F., Ahrari, A., Haunert, J.H., Montzka, C., 2023. The synergies of SMAP enhanced and MODIS products in a random forest regression for estimating 1 km soil moisture over Africa using Google Earth Engine. Big Earth Data.
Montzka, C., Bogena, H.R., Zreda, M., Monerris, A., Morrison, R., Muddu, S., Vereecken, H., 2017. Validation of spaceborne and modelled surface soil moisture products with cosmic-ray neutron probes. Remote Sens. (Basel) 9 (2), 103.
Montzka, C., M. Cosh, B. Bayat, A. Al Bitar, A. Berg, R. Bindlish, H. R. Bogena, J. D. Bolten, F. Cabot, T. Caldwell, S. Chan, A. Colliander, W. Crow, N. Das, G. De Lannoy, W. Dorigo, S. R. Evett, A. Gruber, S. Hahn, T. Jagdhuber, S. Jones, Y. Kerr, S. Kim, C. Koyama, M. Kurum, E. Lopez-Baeza, F. Mattia, K. McColl, S. Mecklenburg, B. Mohanty, P. ÓNeill, D. Or, T. Pellarin, G. P. Petropoulos, M. Piles, R. H. Reichle, N. Rodriguez-Fernandez, C. Rüdiger, T. Scanlon, R. C. Schwartz, D. Spengler, P. Srivastava, S. Suman, R. van der Schalie, W. Wagner, U. Wegmüller, J.-P. Wigneron,

F. Camacho, J. Nickeson, 2020. Soil Moisture Product Validation Good Practices Protocol Version 1.0, Land Product Validation Subgroup (WGCV/CEOS).
Montzka, C., Rötzer, K., Bogena, H.R., Sanchez, N., Vereecken, H., 2018. A new soil moisture downscaling approach for SMAP, SMOS, and ASCAT by predicting sub-grid variability. Remote Sens. (Basel) 10 (3).
Moosavi, V., Talebi, A., Mokhtari, M.H., Hadian, M.R., 2016. Estimation of spatially enhanced soil moisture combining remote sensing and artificial intelligence approaches. Int. J. Remote Sens. 37 (23), 5605–5631.
Munoz-Martin, J.F., Llavería, D., Herbert, C., Pablos, M., Camps, A., 2021. Soil moisture retrieval using the FMPL-2/FSSCat GNSS-R and microwave radiometry data. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 7638–7641.
Nadeem, A.A., Zha, Y., Shi, L., Ali, S., Wang, X., Zafar, Z., Afzal, Z. and Tariq, M.A., 2023. Spatial Downscaling and Gap-Filling of SMAP Soil Moisture to High Resolution Using MODIS Surface Variables and Machine Learning Approaches over ShanDian River Basin, China, Remote Sensing.
Naz, B.S., Kollet, S., Franssen, H.J.H., Montzka, C., Kurtz, W., 2020. A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015. Sci. Data 7 (1).
Naz, B.S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., Gao, H., Springer, A., Franssen, H.J.H., Kollet, S., 2019. Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation. Hydrol. Earth Syst. Sci. 23 (1), 277–301.
Nazar, I.M., Aksoy, M., 2023. Radio frequency interference detection in passive microwave remote sensing using one-class support vector machines. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 6682–6692.
Nguyen, T.T., Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Nguyen, C.T., Zhang, J., Liang, S., Bui, X.T., Hoang, N.B., 2022. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Sci. Total Environ. 833, 155066.
Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V., 2003. Soil moisture retrieval from AMSR-E. IEEE Trans. Geosci. Remote Sens. 41 (2), 215–229.
Nolan, M., Fatland, D.R., 2003. Penetration depth as a DInSAR observable and proxy for soil moisture. IEEE Trans. Geosci. Remote Sens. 41 (3), 532–537.
Notarnicola, C., Angiulli, M., Posa, F., 2008. Soil moisture retrieval from remotely sensed data: neural network approach versus Bayesian method. IEEE Trans. Geosci. Remote Sens. 46 (2), 547–557.
O'Shea, K. and Nash, R., 2015. An Introduction to Convolutional Neural Networks. arXiv preprint.
Ochsner, T.E., Cosh, M.H., Cuenca, R.H., Dorigo, W.A., Draper, C.S., Hagimoto, Y., Kerr, Y.H., Larson, K.M., Njoku, E.G., Small, E.E., Zreda, M., 2013. State of the art in large-scale soil moisture monitoring. Soil Sci. Soc. Am. J. 77 (6), 1888–1919.
Oh, Y., 2004. Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 42 (3), 596–601.
Ojha, N., Merlin, O., Suere, C., Escorihuela, M.J., 2021. Extending the spatio-temporal applicability of dispatch soil moisture downscaling algorithm: a study case using SMAP, MODIS and sentinel-3 data. Front. Environ. Sci. 9.
Oleson, K.W., Niu, G.Y., Yang, Z.L., Lawrence, D.M., Thornton, P.E., Lawrence, P.J., Stöckli, R., Dickinson, R.E., Bonan, G.B., Levis, S., Dai, A., Qian, T., 2008. Improvements to the community land model and their impact on the hydrological cycle. J. Geophys. Res. 113 (G1), G01021.
Osco, L.P., Lemos, E.L., Gonçalves, W.N., Ramos, A.P., Marcato Junior, J., 2023. The potential of visual ChatGPT for remote sensing. Remote Sens. (Basel).
Pan, X., Kornelsen, K.C. and Coulibaly, P., 2017. Estimating Root Zone Soil Moisture at Continental Scale Using Neural Networks. JAWRA Journal of the American Water Resources Association, 53(1): 220-237.
Park, C.H., Montzka, C., Jagdhuber, T., Jonard, F., De Lannoy, G., Hong, J., Jackson, T.J., Wulfmeyer, V., 2019. A dielectric mixing model accounting for soil organic matter. Vadose Zone J. 18 (1).
Park, S., Im, J., Park, S., Rhee, J., 2017. Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. Agric. For. Meteorol. 237, 257–269.
Pasik, A., Gruber, A., Preimesberger, W., De Santis, D., Dorigo, W., 2023. Uncertainty estimation for a new exponential-filter-based long-term root-zone soil moisture dataset from Copernicus climate Change Service (C3S) surface observations. Geosci. Model Dev. 16 (17), 4957–4976.
Pasolli, L., Notarnicola, C., Bertoldi, G., Bruzzone, L., Remelgado, R., Greifeneder, F., Niedrist, G., Chiesa, S.D., Tappeiner, U., Zebisch, M., 2015. Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-Band. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (1), 262–283.
Pasolli, L., Notarnicola, C., Bruzzone, L., 2012. Multi-objective parameter optimization in support vector regression: general formulation and application to the retrieval of soil moisture from remote sensing data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (5), 1495–1508.
Pasolli, L., Notarnicola, C., & Bruzzone, L., 2009. Soil moisture estimation from microwave remote sensing data with nonlinear machine learning techniques. SPIE Image and Signal Processing for Remote Sensing XV, 7477: 382-393.
Paulik, C., Dorigo, W., Wagner, W., Kidd, R., 2014. Validation of the ASCAT soil water index using in situ data from the international soil moisture network. Int. J. Appl. Earth Obs. Geoinf. 30, 1–8.
Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W.J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y.H., Lovergine, F., Mahecha, M.D., Marzahn, P., Mattia, F., Musial, J.P., Preuschmann, S., Reichle, R.H., Satalino, G., Silgram, M., Van Bodegom, P.M., Verhoest, N.E.C., Wagner, W., Walker, J.P., Wegmuller, U., Loew, A., 2021. A roadmap for high-resolution satellite soil moisture applications - confronting product characteristics with user requirements. Remote Sens. Environ. 252.
Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 55 (2), 341–366.
Pennington, D., Ebert-Uphoff, I., Freed, N., Martin, J., Pierce, S.A., 2020. Bridging sustainability science, earth science, and data science through interdisciplinary education. Sustain. Sci. 15 (2), 647–661.
Petropoulos, G.P., Ireland, G. and Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. Physics and Chemistry of the Earth, Parts A/B/C, 83-84: 36-56.
Piles, M., Petropoulos, G.P., Sánchez, N., González-Zamora, Á., Ireland, G., 2016. Towards improved spatio-temporal resolution soil moisture retrievals from the synergy of SMOS and MSG SEVIRI spaceborne observations. Remote Sens. Environ. 180, 403–417.
Piou, C., Gay, P.E., Benahi, A.S., Ebbe, M.A.O.B., Chihrane, J., Ghaout, S., Cisse, S., Diakite, F., Lazar, M., Cressman, K., Merlin, O., Escorihuela, M.J., 2019. Soil moisture from remote sensing to forecast desert locust presence. J. Appl. Ecol. 56 (4), 966–975.
Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018. Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. Soil Tillage Res. 181, 63–81.
Qiu, J., Dong, J., Crow, W.T., Zhang, X., Reichle, R.H., De Lannoy, G.J.M., 2021. The benefit of brightness temperature assimilation for the SMAP Level-4 surface and root-zone soil moisture analysis. Hydrol. Earth Syst. Sci. 25 (3), 1569–1586.
Qu, Y., Zhu, Z., Montzka, C., Chai, L., Liu, S., Ge, Y., Liu, J., Lu, Z., He, X., Zheng, J., Han, T., 2021. Inter-comparison of several soil moisture downscaling methods over the Qinghai-Tibet Plateau China. J. Hydrol. 592, 125616.
Quast, R., Wagner, W., Bauer-Marschallinger, B., Vreugdenhil, M., 2023. Soil moisture retrieval from Sentinel-1 using a first-order radiative transfer model—a case-study over the Po-Valley. Remote Sens. Environ. 295, 113651.
Rahaman, M.M., Thakur, B., Kalra, A., Li, R.P., Maheshwari, P., 2019. Estimating high-resolution groundwater storage from GRACE: a random forest approach. Environments 6 (6).
Rahmani, M., Asgari, J., Asgarimehr, M., 2022. Soil moisture retrieval using space-borne GNSS reflectometry: a comprehensive review. Int. J. Remote Sens. 43 (14), 5173–5203.
Rahmati, M., Or, D., Amelung, W., Bauke, S.L., Bol, R., Hendricks Franssen, H.-J., Montzka, C., Vanderborght, J., Vereecken, H., 2023. Soil is a living archive of the Earth system. Nat. Rev. Earth Environ. 4 (7), 421–423.
Rani, A., Kumar, N., Kumar, J., Kumar, J., Sinha, N.K., 2022. Chapter 6 - machine learning for soil moisture assessment. In: Poonia, R.C., Singh, V., Nayak, S.R. (Eds.), Deep Learning for Sustainable Agriculture. Academic Press, pp. 143–168.
Rao, K., Williams, A.P., Flefil, J.F., Konings, A.G., 2020. SAR-enhanced mapping of live fuel moisture content. Remote Sens. Environ. 245, 111797.
Rawls, W.J., Brakensiek, D.L., Miller, N., 1983. Green-ampt infiltration parameters from soils data. Journal of Hydraulic Engineering-Asce 109 (1), 62–70.
Rawte, V., Sheth, A. and Das, A., 2023. A Survey of Hallucination in Large Foundation Models.
Read, J.S., Jia, X.W., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-guided deep learning predictions of lake water temperature. Water Resour. Res. 55 (11), 9173–9190.
Reichle, R.H., Entekhabi, D., McLaughlin, D.B., 2001. Downscaling of radio brightness measurements for soil moisture estimation: a four-dimensional variational data assimilation approach. Water Resour. Res. 37 (9), 2353–2364.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743): 195-204.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should I Trust you? : explaining the predictions of any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 1135–1144.
Roberts, T.M., Colwell, I., Chew, C., Lowe, S., Shah, R., 2022. A deep-learning approach to soil moisture estimation with GNSS-R. Remote Sens. (Basel) 14 (14), 3299.
Robock, A., 2003. HYDROLOGY | Soil Moisture. In: J.R. Holton (Editor), Encyclopedia of Atmospheric Sciences. Academic Press, Oxford, pp. 987-993.
Rodríguez-Fernández, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jiménez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: application to SMOS. IEEE Trans. Geosci. Remote Sens. 53 (11), 5991–6007.
Russell, S. and Norvig, P., 2016. Artificial Intelligence: A Modern Approach. Pearson.
Sabaghy, S., Walker, J.P., Renzullo, L.J., Jackson, T.J., 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. Remote Sens. Environ. 209, 551–580.
Sahour, H., Sultan, M., Vazifedan, M., Abdelmohsen, K., Karki, S., Yellich, J.A., Gebremichael, E., Alshehri, F., Elbayoumi, T.M., 2020. Statistical applications to downscale GRACE-derived terrestrial water storage data and to fill temporal gaps. Remote Sens. (Basel) 12 (3).
Sandholt, I., Rasmussen, K., Andersen, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sens. Environ. 79 (2–3), 213–224.
Santi, E., Dabboor, M., Pettinato, S., Paloscia, S., 2019. Combining machine learning and compact polarimetry for estimating soil moisture from C-Band SAR data. Remote Sens. (Basel).

Schlüter, S., Leuther, F., Albrecht, L., Hoeschen, C., Kilian, R., Surey, R., Mikutta, R., Kaiser, K., Mueller, C.W., Vogel, H.-J., 2022. Microscale carbon distribution around pores and particulate organic matter varies with soil moisture regime. Nat. Commun. 13 (1), 2098.
Schölkopf, B., 2022. Causality for Machine Learning, Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery, pp. 765–804.
Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture-climate interactions in a changing climate: a review. Earth Sci. Rev. 99 (3–4), 125–161.
Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A.C., Moorhead, R., 2020. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. Remote Sens. (Basel).
Shangguan, Y., Min, X., Shi, Z., 2023. Gap filling of the ESA CCI soil moisture data using a spatiotemporal attention-based residual deep network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 5344–5354.
Sharma, S. and Dhakal, K., 2021. Boots on the Ground and Eyes in the Sky: A Perspective on Estimating Fire Danger from Soil Moisture Content, Fire.
Shen, X., Walker, J.P., Ye, N., Wu, X., Boopathi, N., Yeo, I.Y., Zhang, L., Zhu, L., 2021. Soil moisture retrieval depth of p- and l-band radiometry: predictions and observations. IEEE Trans. Geosci. Remote Sens. 59 (8), 6814–6822.
Singh, A., Gaurav, K., 2023. Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images. Sci. Rep. 13 (1), 2251.
Siqi Lu, Junlin Guo, James R Zimmer-Dauphinee, Jordan M Nieusma, Xiao Wang, Parker VanValkenburgh, Steven A Wernke and Huo, Y., 2024. AI Foundation Models in Remote Sensing: A Survey. ArXiv: 2408.03464.
Sishah, S., Abrahem, T., Azene, G., Dessalew, A., Hundera, H., 2023. Downscaling and validating SMAP soil moisture using a machine learning algorithm over the Awash River basin. Ethiopia. Plos One 18 (1).
Sivaprasad, V., Rahmati, M., Springer, A., Vereecken, H., Montzka, C., 2025. Development of continuous AMSR-E/2 Soil moisture time series by hybrid deep learning model (ConvLSTM2D and Conv2D) and transfer learning for reanalyses. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18, 11169–11182.
Srivastava, P.K., Petropoulos, G.P., Prasad, R., Triantakonstantis, D., 2021. Random forests with bagging and genetic algorithms coupled with least trimmed squares regression for soil moisture deficit using SMOS satellite soil moisture. ISPRS Int. J. Geo Inf. 10 (8), 507.
Staudemeyer, R.C. and Morris, E.R., 2019. Understanding LSTM - a tutorial into Long Short-Term Memory Recurrent Neural Networks. ArXiv, abs/1909.09586.
Stefan, V.-G., Indrio, G., Escorihuela, M.-J., Quintana-Seguí, P., Villar, J.M., 2021. High-resolution SMAP-derived root-zone soil moisture using an exponential filter model calibrated per land cover type. Remote Sens. (Basel) 13 (6), 1112.
Sun, H., Cui, Y.J., 2021. Evaluating downscaling factors of microwave satellite soil moisture based on machine learning method. Remote Sens. (Basel) 13 (1).
Sun, H., Xu, Q., 2021. Evaluating machine learning and geostatistical methods for spatial gap-filling of monthly ESA CCI soil moisture in China. Remote Sens. (Basel).
Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W.H., Ma, X., Rao, Y., Bednar, J.A., Tan, A., Wang, J., Purushotham, S., Gill, T.E., Chastang, J., Howard, D., Holt, B., Gangodagamage, C., Zhao, P., Rivas, P., Chester, Z., Orduz, J., John, A., 2022. A review of earth artificial intelligence. Comput. Geosci. 159, 105034.
Sungmin, O., Orth, R., 2021. Global soil moisture data derived through machine learning trained with in-situ measurements. Sci. Data 8 (1), 170.
Sungmin, O., Orth, R., Weber, U., Park, S.K., 2022. High-resolution European daily soil moisture derived with machine learning (2003-2020). Sci. Data 9 (1).
Tandon, K., Sen, S., Kasiviswanathan, K.S., Soundharajan, B.S., Tummuru, N.R., Das, A., 2022. Integration of machine learning and particle filter approaches for forecasting soil moisture. Stoch. Env. Res. Risk A. 36 (12), 4235–4253.
Tian, J., Han, Z., Bogena, H.R., Huisman, J.A., Montzka, C., Zhang, B., He, C., 2020. Estimation of subsurface soil moisture from surface soil moisture in cold mountainous areas. Hydrol. Earth Syst. Sci. 24 (9), 4659–4674.
Tobin, K.J., Crow, W.T., Dong, J.Z., Bennett, M.E., 2019. Validation of a new root-zone soil moisture product: soil MERGE. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (9), 3351–3365.
Togneri, R., dos Santos, D.F., Camponogara, G., Nagano, H., Custodio, G., Prati, R., Fernandes, S., Kamienski, C., 2022. Soil moisture forecast for smart irrigation: the primetime for machine learning. Expert Syst. Appl. 207.
Toms, B.A., Barnes, E.A., Ebert-Uphoff, I., 2020. Physically interpretable neural networks for the geosciences: applications to earth system variability. J. Adv. Model. Earth Syst. 12 (9).
Topp, G.C., Davis, J.L., Annan, A.P., 1980. Electromagnetic determination of soil-water content - measurements in coaxial transmission-lines. Water Resour. Res. 16 (3), 574–582.
Tuia, D., Roscher, R., Wegner, J.D., Jacobs, N., Zhu, X., Camps-Valls, G., 2021. Toward a collective agenda on AI for earth science data analysis. IEEE Geosci. Remote Sens. Mag. 9 (2), 88–104.
Tuller, M., Babaeian, E., Jones, S.B., 2023. Proximal sensing of soil moisture. In: Goss, M. J., Oliver, M. (Eds.), Encyclopedia of Soils in the Environment, Second Edition. Academic Press, Oxford, pp. 591–599.
U. Wegmuller, C.M., E.G. Njoku, 1995. Canopy opacity models. In: Y.K. B. Choudhury, E. Njoku, P. Pampaloni (Editor), Passive microwave remote sensing of land–atmosphere interactions. VSP, Utrecht.
Unwin, M.J., Pierdicca, N., Cardellach, E., Rautiainen, K., Foti, G., Blunt, P., Guerriero, L., Santi, E., Tossaint, M., 2021. An introduction to the hydroGNSS GNSS reflectometry remote sensing mission. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6987–6999.
Varadharajan, C., Appling, A.P., Arora, B., Christianson, D.S., Hendrix, V.C., Kumar, V., Lima, A.R., Müller, J., Oliver, S., Ombadi, M., Perciano, T., Sadler, J.M., Weierbach, H., Willard, J.D., Xu, Z., Zwart, J., 2022. Can machine learning accelerate process understanding and decision-relevant predictions of river water quality? Hydrol. Process. 36 (4), e14565.
Vereecken, H., Huisman, J.A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogena, H., Weihermüller, L., Herbst, M., Martinez, G., Vanderborght, J., 2014. On the spatio-temporal dynamics of soil moisture at the field scale. J. Hydrol. 516, 76–96.
Vergopolan, N., Chaney, N.W., Beck, H.E., Pan, M., Sheffield, J., Chan, S., Wood, E.F., 2020. Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. Remote Sens. Environ. 242.
Verma, K., Katpatal, Y.B., 2020. Groundwater monitoring using GRACE and GLDAS data after downscaling within basaltic aquifer system. Groundwater 58 (1), 143–151.
Vogel, E., Lerat, J., Pipunic, R., Frost, A.J., Donnelly, C., Griffiths, M., Hudson, D., Loh, S., 2021. Seasonal ensemble forecasts for soil moisture, evapotranspiration and runoff across Australia. J. Hydrol. 601, 126620.
Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldana, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Zuger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Bloschl, G., Eitzinger, J., Steinnocher, K., Zeil, P., Rubel, F., 2013. The ASCAT soil moisture product: a review of its specifications, validation results, and emerging applications. Meteorol. Z. 22 (1), 5–33.
Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sens. Environ. 70 (2), 191–207.
Wang, G.J., Garcia, D., Liu, Y., de Jeu, R., Dolman, A.J., 2012. A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. Environ. Model. Software 30, 139–142.
Wang, J.R., Schmugge, T.J., 1980. An empirical-model for the complex dielectric permittivity of soils as a function of water-content. IEEE Trans. Geosci. Remote Sens. 18 (4), 288–295.
Wang, R., Wan, S., Chen, W., Qin, X., Zhang, G., Wang, L., 2024. A novel finer soil strength mapping framework based on machine learning and remote sensing images. Comput. Geosci. 182, 105479.
Wang, T., Liang, J., Liu, X., 2019. Soil moisture retrieval algorithm based on TFA and CNN. IEEE Access 7, 597–604.
Wang, Y., Zha, Y.Y., 2024. Comparison of transformer, LSTM and coupled algorithms for soil moisture prediction in shallow-groundwater-level areas with interpretability analysis. Agric Water Manag 305.
Wei, Z.S., Meng, Y.Z., Zhang, W., Peng, J., Meng, L.K., 2019. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. Remote Sens. Environ. 225, 30–44.
Weiss, T., Ramsauer, T., Löw, A., Marzahn, P., 2020. Evaluation of different radiative transfer models for microwave backscatter estimation of wheat fields. Remote Sens. (Basel) 12 (18).
Wicki, A., Lehmann, P., Hauck, C., Seneviratne, S.I., Waldner, P., Stähli, M., 2020. Assessing the potential of soil moisture measurements for regional landslide early warning. Landslides 17 (8), 1881–1896.
Wigneron, J.P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.J., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J.C., Grant, J.P., Guglielmetti, M., Hornbuckle, B., Matzler, C., Pellarin, T., Schwank, M., 2007. L-band microwave emission of the biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields. Remote Sens. Environ. 107 (4), 639–655.
Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Neural Comput. 1 (2), 270–280.
Wu, X., Ma, W., Xia, J., Bai, W., Jin, S., Calabia, A., 2021a. Spaceborne GNSS-R Soil moisture retrieval: status, development opportunities, and challenges. Remote Sens. (Basel) 13 (1), 45.
Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S., 2021b. A comprehensive survey on graph neural networks. IEEE Trans. Neural Networks Learn. Syst. 32 (1), 4–24.
Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., Shen, C., 2021. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. J. Hydrol. 603, 127043.
Xu, W., Zhang, Z.X., Long, Z.H., Qin, Q.M., 2021. Downscaling SMAP soil moisture products with convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 4051–4062.
Xu, X.D., Li, W., Ran, Q., Du, Q., Gao, L.R., Zhang, B., 2018. Multisource remote sensing data classification based on convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (2), 937–949.
Yan, Q., Huang, W., Jin, S., Jia, Y., 2020. Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. Remote Sens. Environ. 247, 111944.
Yan, R., Bai, J.J., 2020. A new approach for soil moisture downscaling in the presence of seasonal difference. Remote Sens. (Basel) 12 (17).
Yang, C., Mao, K., Guo, Z., Shi, J., Bateni, S.M., Yuan, Z., 2024. Review of GNSS-R technology for soil moisture inversion. Remote Sens. (Basel) 16 (7), 1193.
Yao, P., Shi, J., Zhao, T., Lu, H., Al-Yaari, A., 2017. Rebuilding long time series global soil moisture products using the neural network adopting the microwave vegetation index. Remote Sens. (Basel).
Yin, D.X., Xue, Z.G., Bao, D.Y., RafieeiNasab, A., Huang, Y.J., Morales, M., Warner, J.C., 2022. Understanding the role of initial soil moisture and precipitation magnitude in flood forecast using a hydrometeorological modelling system. Hydrol. Process. 36 (10).
Yueh, S., Shah, R., Xu, X., Elder, K., Starr, B., 2020. Experimental demonstration of soil moisture remote sensing using p-band satellite signals of opportunity. IEEE Geosci. Remote Sens. Lett. 17 (2), 207–211.

Zeng, L., Hu, S., Xiang, D., Zhang, X., Li, D., Li, L., Zhang, T., 2019. Multilayer soil moisture mapping at a regional scale from multisource data via a machine learning method. Remote Sens. (Basel).
Zeynoddin, M., Bonakdari, H., 2022. Structural-optimized sequential deep learning methods for surface soil moisture forecasting, case study Quebec Canada. Neural Comput. Appl. 34 (22), 19895–19921.
Zhang, D.J., Zhou, G.Q., 2016. Estimation of soil moisture from optical and thermal remote sensing: a review. Sensors 16 (8).
Zhang, J.X., Liu, K., Wang, M., 2021a. Downscaling groundwater storage data in China to a 1-km resolution using machine learning methods. Remote Sens. (Basel) 13 (3).
Zhang, L., Xue, Z.H., Liu, H., Li, H., 2023a. Enhanced generalized regression neural network with backward sequential feature selection for machine-learning-driven soil moisture estimation: a case study over the Qinghai-Tibet Plateau. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 7173–7185.
Zhang, N., Quiring, S., Ochsner, T., Ford, T., 2017. Comparison of three methods for vertical extrapolation of soil moisture in Oklahoma. Vadose Zone J. 16 (10), vzj2017.04.0085.
Zhang, Q., Yuan, Q.Q., Li, J., Wang, Y., Sun, F.J., Zhang, L.P., 2021b. Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013-2019. Earth Syst. Sci. Data 13 (3), 1385–1401.
Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D., Zhang, Y., 2022a. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge. Gondw. Res. 109, 1–17.
Zhang, Y., Huang, F.N., Li, L., Li, Q.L., Zhang, Y.K., Shangguan, W., 2023b. Real-time forecast of SMAP L3 soil moisture using spatial-temporal deep learning model with data integration. Remote Sens. (Basel) 15 (2).
Zhang, Y., Liang, S., Ma, H., He, T., Wang, Q., Li, B., Xu, J., Zhang, G., Liu, X., Xiong, C., 2023c. Generation of global 1 km daily soil moisture product from 2000 to 2020 using ensemble learning. Earth Syst. Sci. Data 15 (5), 2055–2079.
Zhang, Y.F., Liang, S.L., Zhu, Z.L., Ma, H., He, T., 2022b. Soil moisture content retrieval from Landsat 8 data using ensemble learning. ISPRS J. Photogramm. Remote Sens. 185, 32–47.
Zhao, W., Sánchez, N., Lu, H., Li, A.N., 2018. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. J. Hydrol. 563, 1009–1024.
Zhao, W., Wen, F.P., Wang, Q.M., Sanchez, N., Piles, M., 2021. Seamless downscaling of the ESA CCI soil moisture data at the daily scale with MODIS land products. J. Hydrol. 603.
Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 46 (24), 14496–14507.
Zheng, C., Jia, L., Zhao, T., 2023. A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution. Sci. Data 10 (1), 139.
Zheng, J.Y., Lu, H.S., Crow, W.T., Zhao, T.J., Merlin, O., Rodriguez-Fernandez, N., Shi, J. C., Zhu, Y.H., Su, J.B., Chuen, S.A.K., Wang, X.Y., Gou, Q.Q., 2021. Soil moisture downscaling using multiple modes of the DISPATCH algorithm in a semi-humid/humid region. Int. J. Appl. Earth Obs. Geoinf. 104.
Zhu, J.-J., Jiang, J., Yang, M., Ren, Z.J., 2023. ChatGPT and environmental research. Environ. Sci. Technol. 57 (46), 17667–17670.
Zribi, M., Muddu, S., Bousbih, S., Al Bitar, A., Tomer, S.K., Baghdadi, N., Bandyopadhyay, S., 2019. Analysis of L-Band SAR data for soil moisture estimations over agricultural areas in the tropics. Remote Sens. (Basel) 11 (9).