

Spatial Soil Moisture Prediction From In Situ Data Upscaled to Landsat Footprint: Assessing Area of Applicability of Machine Learning Models

Yi Yu¹, Member, IEEE, Brendan P. Malone², Luigi J. Renzullo³, Chad A. Burton⁴, Siyuan Tian⁵,
Ross D. Searle⁶, Thomas F. A. Bishop⁷, and Jeffrey P. Walker⁸, Fellow, IEEE

Abstract—The inherent spatial mismatch between satellite-derived and ground-observed near-surface soil moisture (SM) data necessitates cautious interpretation of point-to-pixel comparisons. Although data-driven upscaling of point-scale SM may enable statistically sound comparisons, the uncertainty across a spatial domain was less explored in previous studies. This gap underscores the need of addressing the spatial prediction uncertainties when extrapolating SM information to a broader spatial scale. Accordingly, this study presents a spatial prediction approach integrating machine learning (ML) and spatiotemporal fusion, which enables the characterization of SM variability at the Landsat satellite footprint. Spatially clustered SM from 28 in situ stations was extrapolated to a 100×100 km area at 100 m resolution over a cross-validation (CV) period (2016–2019) and an independent test period (2020–2021). The area of applicability (AOA), which represents the spatial extent within which a prediction model is considered reliable, was determined for two ML models: random forests (RFs) and extreme gradient boosting (XGB). The AOA of RF and XGB models encompassed 43.1% and 41.5% of the study area, respectively. The spatial SM predictions were further evaluated against multiple independent datasets, including field campaign data, in situ SM from different networks, and satellite retrievals. Specifically, RF-predicted SM achieved a spatial R of 0.62–0.64 against field campaign data, temporal R of 0.84–0.91 against network-recorded data, and spatiotemporal R

of 0.87 against SM active passive (SMAP) L2 data during the CV period. SM predictions within the AOA showed markedly lower uncertainties, which were further validated across an extended area (300×300 km) with diverse physiographic conditions. Overall, this study demonstrated the use of AOA in delineating the statistically reliable spatial extent for ML-based SM predictions.

Index Terms—Area of applicability (AOA), Landsat, machine learning (ML), soil moisture (SM), spatial prediction, upscaling.

I. INTRODUCTION

SOIL moisture (SM) is vital to global water, energy, and biogeochemical cycles [1], influencing the distribution of water and radiation at the land–atmosphere interface [2], [3]. Spatiotemporal SM datasets have been extensively developed and explored through remote sensing techniques [4], [5]. The remotely sensed SM data is typically acquired from microwave sensors, such as the Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) on board the Soil Moisture and Ocean Salinity (SMOS) [6], the Advanced Scatterometer (ASCAT) on board the series of Meteorological Operational (METOP) satellites [7], and the L-band radiometer on board the Soil Moisture Active Passive (SMAP) [8]. Alternatively, the simulations can be obtained from land surface models (LSMs) embedded in data assimilation systems that incorporate the remotely sensed SM [9], often from retrospective reanalyses such as European Centre for Medium-Range Weather Forecasts (ECMWF)’s fifth-generation climate reanalysis (ERA5) [10] and the National Aeronautics and Space Administration (NASA)’s Modern-Era Retrospective analysis for Research and Applications (MERRA) land data products [11]. These data have played a crucial role in studying SM dynamics within terrestrial environments.

An accurate evaluation of satellite-derived or model-estimated SM values is usually dependent on comparisons against in situ SM measurements [12], [13]. At a farming or catchment scale, there is a necessity to enhance the comprehension and application of in situ SM observations. This knowledge is often accumulated and refined over time by farm operators, while it can be supplemented through the deployment of SM sensors or networks. Some well-known SM ground networks, such as the Terrestrial Environmental Observatories (TERENO) [14], U.K. Cosmic-ray SM Observing System (COSMOS-UK) [15], the Australian Cosmic-Ray

Received 6 February 2025; revised 15 March 2025; accepted 21 April 2025. Date of publication 30 April 2025; date of current version 15 May 2025. This work was supported in part by the Grains Research and Development Corporation (GRDC) Project *SoilWaterNow: Soil Water Nowcasting for the Grains Industry* under Grant UOS2002-001RTX and in part by Australian Research Council through the Discovery Project *Towards P-Band Soil Moisture Sensing from Space* under Grant DP170102373. (Corresponding author: Yi Yu.)

Yi Yu is with the Fenner School of Environment and Society, The Australian National University, Canberra, ACT 2601, Australia, and also with CSIRO Agriculture and Food, Canberra, ACT 2601, Australia (e-mail: u6726739@anu.edu.au; yi.yu@csiro.au).

Brendan P. Malone is with CSIRO Agriculture and Food, Canberra, ACT 2601, Australia (e-mail: brendan.malone@csiro.au).

Luigi J. Renzullo and Siyuan Tian are with the Bureau of Meteorology, Parkes, ACT 2600, Australia, and also with the Fenner School of Environment and Society, The Australian National University, Canberra, ACT 2601, Australia (e-mail: luigi.renzullo@bom.gov.au; siyuan.tian@bom.gov.au).

Chad A. Burton is with the Fenner School of Environment and Society, The Australian National University, Canberra, ACT 2601, Australia (e-mail: chad.burton@anu.edu.au).

Ross D. Searle is with CSIRO Agriculture and Food, St Lucia, QLD 4067, Australia (e-mail: ross.searle@csiro.au).

Thomas F. A. Bishop is with the School of Life and Environmental Sciences, Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW 2015, Australia (e-mail: thomas.bishop@sydney.edu.au).

Jeffrey P. Walker is with the Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: jeff.walker@monash.edu).

Digital Object Identifier 10.1109/TGRS.2025.3565818

Neutron Soil Moisture Monitoring Network (CosmOz) [16], and the OzNet Hydrological Monitoring Network [17], play a pivotal role in evaluating SM from model simulations or satellite retrievals. However, the representativeness of ground measurements is constrained to a specific area and their distribution is typically sparse [18], [19]. The inherent mismatch in spatial support therefore necessitates a cautious interpretation of point-to-pixel comparisons. Researchers have consistently questioned how to effectively align the footprints between in situ data and spatial grids. As a result, extensive investigations into upscaling techniques have been undertaken [20].

Previous studies have implemented diverse approaches to upscale in situ SM data to the spatial resolution of satellite footprints, primarily leveraging regression models while incorporating remotely sensed predictor variables. The Moderate Resolution Imaging Spectroradiometer (MODIS)-derived apparent thermal inertia (ATI) approach [20] is a representative semiphysical upscaling approach that ensures consistency between upscaled SM estimates and local hydrometeorological characteristics. Concurrently, empirical methodologies, encompassing both geostatistical techniques and machine learning (ML) algorithms, have been extensively investigated, such as the spatiotemporal regression block kriging (STRBK) [21], least absolute shrinkage and selection operator (LASSO) [22], and random forests (RFs) [23]. Despite their methodological diversity, most methods impose stringent requirements on the quantity and spatial arrangement of SM sensors. Furthermore, these studies exhibited limitations in validation methodology, lacking evaluations against independent datasets (e.g., field campaign) and assessment of spatial representativeness of models. Consequently, extant upscaling approaches typically demonstrated site specificity with limited transferability across broader spatial domains [22]. A detailed comparison can be found in Table IV in the Discussion Section.

Specifically, the discussion around uncertainty magnitude within the spatial prediction remains limited. The majority of studies have been confined to localized conditions, specifically upscaling in situ SM at the precise geolocation of satellite pixels, and have primarily concentrated on cross-validation (CV) metrics, which are inadequate for comprehensively capturing the error distribution across a spatial domain. Considering that all upscaling strategies inherently carry non-negligible errors [19], and are potentially influenced by spatial autocorrelation [24], it is crucial to assess the spatial distribution of error magnitudes and extrapolate the applicability of SM prediction at a relatively broader spatial scale. Given the scarcity of independent validation data for spatial prediction results in both space and time, emerging spatial analysis tools can play a pivotal role in quantifying such spatial uncertainty [25], [26].

The area of applicability (AOA) metric [25] describes the spatial region where a prediction model is expected to maintain CV performance. This performance is learned from a spatially distributed training data that is usually collected from field sampling and/or ground stations. The AOA metric has been employed to determine the reliable spatial prediction extents for multiple geophysical variables, including soil properties [27], [28], soil organic carbon [29], and LST [30].

However, relatively little attention has been given to its use in SM prediction [31]. Results from field campaigns have consistently shown that variations in local topography, vegetation, and soil can systematically influence the relative bias of localized SM conditions when compared to broader-scale areal averages [19], [32]. The implementation of AOA can consequently help identify the reliable prediction area by accounting for the spatial variation of these predictor variables. Additionally, it is advisable to assess model performance against independent data sources, such as field campaigns or other networks [23]. These complementary assessments provide valuable context for interpreting the AOA metrics.

To enhance the extrapolation of spatial uncertainty, another area that requires improvement is the resolution of predictor variables used in spatial SM prediction. Previous studies have often relied on regression relationships between point-scale SM and predictor variables with coarse-to-moderate resolution. However, given the marked spatial variation in SM, the information provided by predictor variables at such resolutions (e.g., 1 km for MODIS) may not be directly applicable in building regression relationships with point-scale SM data. Matching the resolution of predictor variables to a level comparable to ground measurements is therefore expected to improve the quantification of spatiotemporal dynamics of SM. Moreover, as efforts have evolved to the production of fine resolution (equivalent to or finer than 1 km) SM estimates (e.g., [33], [34], [35], [36]), there is a growing need to quantify the error dynamics of SM at fine resolution (e.g., 100 m). It holds promise for capturing the intricate variations in SM across fields, leading to more accurate assessments of spatial uncertainty.

Given the above, the objectives of this study are the following: 1) to spatially predict SM at 100 m resolution with daily frequency in an agricultural region using data-driven approaches; 2) to conduct multifold spatial CV to assess the performance of two alternate ML models during 2016–2019; 3) to determine AOA of the employed models based on spatial CV performance and feature importance during 2016–2019; and 4) to evaluate the spatial SM prediction against independent data from field campaigns, additional monitoring networks, and satellite retrievals during both a CV period (2016–2019) and independent test period (2020–2021), to complement the AOA metric.

II. STUDY AREA AND DATA

Table I summarizes the data used in this study, including in situ SM measurements, gridded predictor variables, and independent SM datasets. The OzNet SM measurements were used as the response variable in the ML-based spatial prediction. The gridded predictor variables represent the multiscale physical processes that align with recent findings on the importance of precipitation, vegetation, and temperature controls on SM variability across landscapes [37].

- 1) *MODIS Products and Landsat-Resolution Products:* This captures surface energy balance and vegetation water stress through albedo, normalized difference vegetation index (NDVI), LST, and ET.

TABLE I
SUMMARY OF DATA USED IN THIS STUDY

Category	Dataset	Variable(s)	Spatial resolution	Temporal frequency	Period	Reference(s)
Ground measurements	OzNet	SM	Point	Hourly	01 Jan 2016 – 31 Dec 2021	[17]
	CosmOz	SM	Sub-hectare	Hourly	01 Jan 2016 – 31 Dec 2021	[16]
	OzFlux	SM	Point	Hourly	01 Jan 2016 – 31 Dec 2021	[38]
Gridded predictor variables	MCD43A4	Albedo and NDVI	500 m	Daily	01 Jan 2016 – 31 Dec 2021	[39]
	MOD11A1	LST	1 km	Daily	01 Jan 2016 – 31 Dec 2021	[40]
	Landsat 8	Albedo, NDVI and LST	30 m / 100 m	16-day	01 Jan 2016 – 31 Dec 2021	[41]
	CMRSET	ET _a	30 m	Monthly	01 Jan 2016 – 31 Dec 2021	[42]
	ANUClimate	T _{air} , S _{rad} , VPD, and precipitation	1 km	Daily	01 Jan 2016 – 31 Dec 2021	[43]
	DEM-S	Elevation	30 m	Static	/	[44]
Observed and retrieved SM	SLGA	AWC, clay, sand and silt	90 m	Static	/	[45], [46]
	HDAS	SM	Point	Daily	30 Sep 2019 – 18 Oct 2019; 08 Mar 2021 – 26 Mar 2021	[47], [48]
	SMAP	SM	9 km	Daily	01 Jan 2016 – 31 Dec 2021	[8]

- 2) *ANUClimate*: This quantifies atmospheric water flux components through air temperature (T_{air}), solar radiation (S_{rad}), vapor pressure deficit (VPD), and precipitation.
- 3) *DEM-S and the Soil and Landscape Grid of Australia (SLGA)*: They represent the underlying hydrological controls on water redistribution and retention.

The point-scale SM observations acquired by the hydroprobe data acquisition system (HDAS) were used to evaluate the spatial pattern of the predicted SM. The independent in situ SM observations from CosmOz and OzFlux networks and SM retrieval from the SMAP mission were jointly used to evaluate the temporal dynamics of the spatial SM prediction.

A. Study Area and In Situ Measurements

The Yanco agricultural region was designated as the study area (Fig. 1). Yanco is situated within the Murrumbidgee Catchment, with a Mediterranean climate that is representative for the majority of agricultural regions in southeastern Australia (i.e., hot dry summer and cool moist winter). Fig. 1 shows the following: (a) the land cover information in 2021 [49]; (b) a Landsat true-color composite image acquired on October 13, 2020; (c) a map of cropland extent produced by the United States Geological Survey (USGS); and (d) the smoothed digital elevation model (DEM-S) [44]. The annual total precipitation in this region averages approximately 400 mm. The in situ SM data were collected from the OzNet Hydrological Monitoring Network [17] between January 1, 2016 and December 31, 2021. OzNet site locations are indicated as black squares in Fig. 1(a) and detailed in Table II. The OzNet sites were categorized into four groups to facilitate a fourfold spatial CV (Table II). Moreover, two spatial clusters, where OzNet sites have a relatively dense distribution, were chosen, referred to as cluster A [Fig. 1(e)] and cluster B [Fig. 1(f)], respectively.

TABLE II
SUMMARY OF THE SITE INFORMATION FROM OzNET. THE “FOLD” COLUMN REPRESENTS THE NUMBERED GROUP OF THE FOURFOLD SPATIAL CV. THE “CLUSTER” COLUMN REPRESENTS THE SPATIAL CLUSTER WHERE THE SITES ARE LOCATED. COSMOZ-YANCO AND OZFLUX-YANCO ARE LOCATED WITHIN THE CLUSTER B REGION (SEE FIG. 1) BUT THEY WERE NOT INCLUDED IN THE TRAINING

Site name	Latitude (°S)	Longitude (°E)	Fold	Cluster
ya3	34.6772	146.1397	1	A
ya4a	34.7060	146.0794	2	A
ya4b	34.7031	146.1053	3	A
ya4c	34.7142	146.0943	4	A
ya4d	34.7142	146.0751	1	A
ya4e	34.7214	146.1030	4	A
ya5	34.7129	146.1277	3	A
ya7a	34.7352	146.0820	1	A
ya7b	34.7378	146.0987	4	A
ya7d	34.7544	146.0778	2	A
ya7e	34.7507	146.0949	1	A
ya9	34.7414	146.1536	2	A
yb1	34.9412	146.2765	4	B
yb3	34.9427	146.3401	2	B
yb5a	34.9653	146.3026	4	B
yb5b	34.9634	146.3184	3	B
yb5d	34.9848	146.2930	2	B
yb5e	34.9797	146.3205	3	B
yb7a	34.9885	146.2694	1	B
yb7c	34.9984	146.2785	2	B
yb7d	35.0050	146.2685	3	B
yb7e	35.0077	146.2880	4	B
yb9	35.0022	146.3398	3	B
y10	35.0053	146.3099	1	B
y5	34.7283	146.2932	3	N/A*
y7	34.8518	146.1153	4	N/A
y8	34.8470	146.4140	1	N/A
y9	34.9678	146.0163	2	N/A
CosmOz-Yanco	35.0100	146.3000	N/A	N/A
OzFlux-Yanco	34.9893	146.2907	N/A	N/A

* N/A represents they do not belong to any spatial cluster.

Two field campaigns were conducted in the coleambally irrigation area (CIA) in 2019 [47] and 2021 [48], respectively. The spatial coverage of the field campaigns spans 146.07°E–146.11°E and 34.70°S–35.73°S (Fig. 1(e);

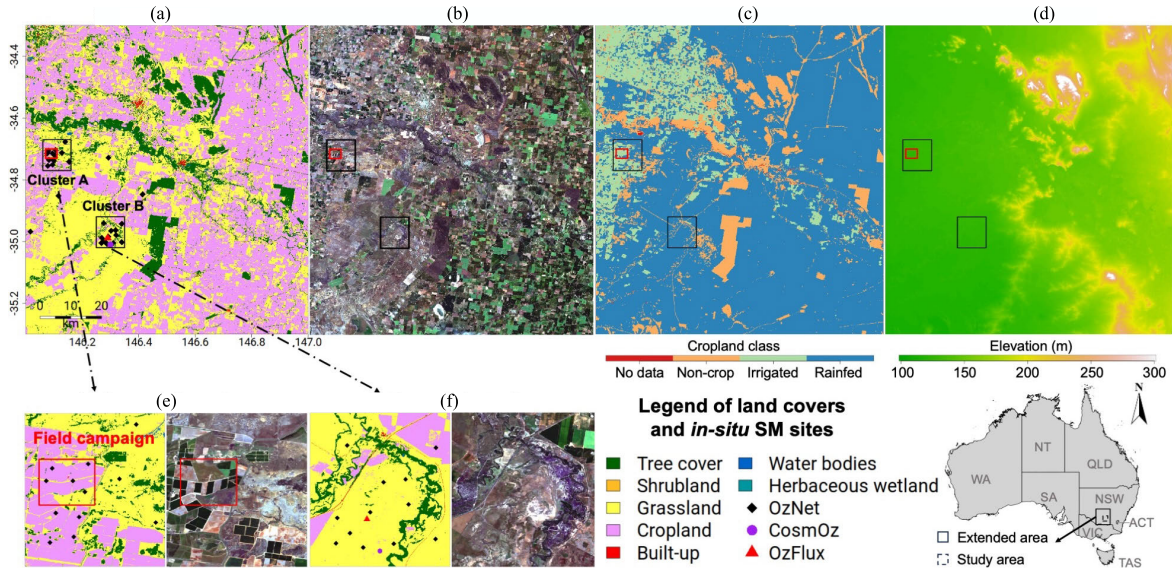


Fig. 1. (a) Land cover information of the Yanco agricultural region (highlighted in the dashed rectangle within the state of NSW) in 2021 [49]. (b) Landsat true-color composite image acquired on October 13, 2020. (c) USGS cropland extent [50]. (d) DEM-S [44]. (e) and (f) Zoomed land cover and true-color images for cluster A and B, respectively. The black dots, purple circle, and red triangle indicate in situ sites from OzNet, CosmOz, and OzFlux, respectively. The study area spans 146.00°E–147.00°E and 34.30°S–35.30°S (100 × 100 km). The cluster A spans 146.06°E–146.16°E and 34.67°S–34.77°S (10 × 10 km), with the field campaign in the coleambally irrigation area (CIA; [51]) covering 146.07°E–146.11°E and 34.70°S–34.73°S (4 × 3 km). Cluster B spans 146.25°E–146.35°E and 34.92°S–35.02°S (10 × 10 km). The AOA assessment is also extended to a broader region (144.50°E–147.50°E and 33.00°S–36.00°S; 300 × 300 km) surrounding the Yanco agricultural region. See Fig. A1 for more details about the extended area.

4 × 3 km), which was included within cluster A. Additionally, in situ SM data were collected from two other SM monitoring networks to facilitate independent evaluations, including the following: 1) the CosmOz [16] and 2) Australian and New Zealand Flux Research and Monitoring (OzFlux) [38], shown as a purple dot and a red triangle in Fig. 1(a) and (f).

The study domain is also extended to a mosaic of six Landsat tiles (path: 092–093, row: 083–085). The extended area spans 144.50°E–147.50°E and 33.00°S–36.00°S (300 × 300 km; Fig. 1). See Fig. A1 for more details.

B. Gridded Predictor Variables

1) *MODIS Data*: The MODIS products have been extensively used in moderate-resolution land surface monitoring [52]. The 500 m resolution daily nadir bidirectional reflectance distribution function adjusted reflectance (NBAR) product (MCD43A4; version 6.1) [39] and the 1 km resolution daily LST product (MOD11A1; version 6.1) [40] were collected over January 1, 2016–December 31, 2021. The 500 m daily albedo and NDVI were derived from the surface reflectance bands of MCD43A4 dataset. The MODIS albedo was calculated using the Liang method [53]. These two collections were acquired from NASA's Earthdata Search platform (<https://search.earthdata.nasa.gov/search>; accessed April 28, 2025).

2) *Landsat-Resolution Data*: The Landsat 8 mission provides data from visible to thermal infrared bands, with a spatial resolution ranging from 30 to 100 m. The 30 m resolution 16-day Landsat 8 NBAR product (version 3.0.0) [41] was collected for the period spanning January 1, 2016–December 31, 2021. With the application of the bidirectional reflectance distribution function (BRDF), it ensures consistency with the MCD43A4 product. The 30 m resolution 16-day albedo and

NDVI were derived using its surface reflectance bands. The Landsat albedo was calculated using the Liang method [53]. The 100 m resolution 16-day Landsat LST was retrieved using a split-window algorithm [54]. A 30 m resolution monthly CSIRO MODIS reflectance-based scaling evapotranspiration (CMRSET; version 2.2) Landsat actual evapotranspiration (ET_a) product [42] was also collected between January 1, 2016 and December 31, 2021. The Landsat 8 data for Australia is publicly available from the Digital Earth Australia (DEA; <https://dea.ga.gov.au/>; accessed April 28, 2025).

3) *ANUClimate*: ANUClimate (version 2.0) is a climatic dataset spanning the Australian continent, comprising gridded daily and monthly climate variables at a spatial resolution of 0.01° [43]. This dataset was generated by interpolating Australia's national point climate data using trivariate thin plate smoothing spline functions [55], which account for spatial variations in longitude, latitude, and enhanced elevation. The average air temperature (T_{air}), incoming shortwave radiation (S_{rad}), VPD, and precipitation were collected during January 1, 2016–December 31, 2021 to serve as predictor variables, all of which have a spatial resolution of 1 km and daily frequency.

4) *Static Data*: The elevation and soil data were collected as static predictor variables. Specifically, the Australian Smoothed Digital Elevation Model (DEM-S; version 1.0) [44] was collected, which eliminates the impacts of ground vegetation and has undergone a smoothing process to mitigate the influence of noise, with a spatial resolution of 1 arcsecond (approximately 30 m). The available water capacity (AWC), clay, sand, and silt data were acquired from the (SLGA; version 2.0) collection [45], [46]. Their spatial resolution is 3 arcsecond (approximately 90 m) and the depth of soil data

used herein is 0–5 cm, consistent with the near-surface layer of SM.

C. Observed and Retrieved SM

1) *Hydraprobe Data Acquisition System*: HDAS is an integrated spatial data acquisition tool functioning within a mobile GIS environment [56]. HDAS was utilized to collect SM data during the P-band radiometer inferred SM (PRISM) field campaigns in CIA in 2019 and 2021 [47], [48]. Throughout the campaigns, routine HDAS measurements were conducted across regular grids spaced at 50 m intervals to capture the spatial distribution of near-surface SM (approximately 5 cm). The acquired near-surface HDAS SM data spans September 30, 2019–October 18, 2019 and March 8, 2021–March 26, 2021. The HDAS data is available at <https://prism.monash.edu/index.html> (accessed April 28, 2025).

2) *SM Active Passive*: The SMAP mission was launched by NASA in 2015 and measures L-band microwave brightness temperature at 1.4 GHz [8]. The L2 product (version 6) was collected between January 1, 2016 and December 31, 2021, which provides radiometer SM derived from level 1 observations and ancillary information. It has an original spatial resolution of 36 km and a temporal revisit frequency of 2–3 days. The data is gridded at the 9 km Equal Area Scalable Earth-2 (EASE2) grids using the Backus–Gilbert interpolation method. The data is publicly available from the National Snow and Ice Data Center (<https://nsidc.org/data/smap>; accessed April 28, 2025).

III. METHODOLOGY

Fig. 2 presents the experimental design in four key steps. First, a nearest neighbor method was utilized to resample all predictor variables to 0.001° resolution grids under the World Geodetic System 1984 (WGS84) datum. Concurrently, a spatiotemporal fusion was implemented between MODIS and Landsat data to downscale MODIS albedo, NDVI, and LST to a 100 m resolution with daily frequency (Section III-A). Second, the ML models were trained using point-scale SM from 28 OzNet sites between 2016 and 2019 as the response variable, with the collected predictor variables serving as input features. The performance of two ML models, specifically the RFs and extreme gradient boosting (XGB), were compared through a fourfold spatial CV and a cross-cluster (CC) validation to assess the transferability of the spatial prediction approach while considering regional nuances (Section III-B). Third, the AOA metrics of RF and XGB models were determined based on their respective performance and feature importance in CV and CC (Section III-C). Fourth, the spatial SM prediction within AOA was evaluated against multiple independent datasets, including data from field campaigns, CosmOz and OzFlux networks, and SMAP SM retrievals (Section III-D).

A. Spatiotemporal Fusion

The implementation of spatiotemporal fusion is to address the resolution mismatch between key geophysical indicators

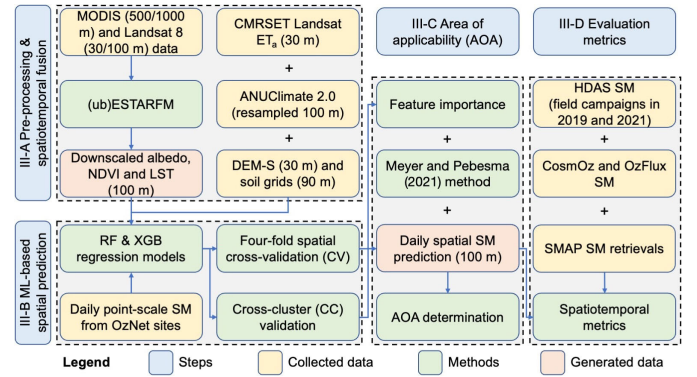


Fig. 2. Experimental design. The “steps” polygons correspond to Sections III-A–III-D.

and surface characteristics in SM prediction. While MODIS products (>500 m) provide frequent data essential for capturing daily surface dynamics, their spatial resolution is insufficient to characterize field-scale heterogeneity in SM. Through fusion with Landsat data (30–100 m), it achieves daily estimates of these predictor variables at resolutions commensurate with underlying soils (90 m) and topographic features (30 m).

The enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) [57] has been extensively applied to generate fine-resolution surface reflectance [51], [57] and demonstrated an exceptional performance in fusing surface reflectance. Furthermore, an unbiased variant (ubESTARFM) [58] has demonstrated more effectiveness in fusing LST data. Hence, ESTARFM was implemented to fuse MODIS and Landsat surface reflectance data (i.e., albedo and NDVI), while ubESTARFM was implemented to fuse MODIS and Landsat LST data, to generate daily 100 m estimates of these predictor variables. The formula of ESTARFM is given as follows:

$$F(x_{w/2}, y_{w/2}, t_p) = F(x_{w/2}, y_{w/2}, t_0) + \sum_{i=1}^N W_i \times V_i \times (C(x_i, y_i, t_p) - C(x_i, y_i, t_0)) \quad (1)$$

where $F(x_{w/2}, y_{w/2}, t_0)$ is the central pixel $(x_{w/2}, y_{w/2})$ of the fine-resolution image within a search window at time t_0 ; t_p is the prediction time; N is the number of similar pixels within the search window; W_i is the weight of the i th similar pixel; V_i is a regression coefficient between the chosen fine- and coarse-resolution pixels; and $C(x_i, y_i)$ is the pixel at the location (x_i, y_i) of a coarse-resolution image. For ubESTARFM, a local bias correction step is implemented to the central and similar fine-resolution pixels within the search window. For more details, please see [57], [58].

B. ML-Based Spatial Prediction

ML models and algorithms have been well established and applied in geoscience and remote sensing [59]. Two widely recognized ML models were employed, namely the RFs [60] and the XGB [61], to perform a comparative study. These ensemble algorithms have demonstrated robust performance in capturing nonlinear relationships in environmental data,

particularly in SM estimation applications [62]. RF's bootstrap aggregation mechanism provides protection against overfitting, while XGB's gradient boosting framework offers complementary strengths in handling complex feature interactions. Both ML approaches offer built-in feature importance assessment capabilities, enabling systematic quantification of predictor variable contributions—a critical requirement for implementing AOA. They can also be interpreted using the shapley additive explanations (SHAPs), which provide insights into the average marginal contribution of a feature value across all possible coalitions [63], [64]. The methodological selection balances algorithmic sophistication, interpretability, computational efficiency, and prediction accuracy.

1) *Fourfold Spatial CV*: A “leave-seven-sites-out” CV strategy was performed to assess the fitting performance of the ML models between 2016 and 2019. The 28 OzNet sites were randomly divided into fourfolds. In each fold, seven sites were reserved for validation purposes, while the remaining 21 sites were utilized for training the models. Specifically, each fold contains one site outside both clusters and six sites with close neighbors (Fig. 1 and Table II). The specifics of each fold are outlined in the “fold” column of Table II. This approach was chosen to ensure that the trained model captures the spatiotemporal distribution of SM data across a relatively large spatial domain.

2) *CC Validation*: The validation strategy was expanded beyond random selection of sites by conducting a CC validation. This approach was crucial due to the spatial clustering of OzNet sites (Fig. 1), necessitating an evaluation of model performance using a “leave-cluster-out” approach commonly employed in spatial ML applications [65]. Details regarding each cluster (i.e., A and B) can be found in the “cluster” column of Table II. The cluster A and B, each of which contains 12 OzNet sites, were employed reciprocally for a CC validation. Specifically, sites within cluster A were utilized for training while sites within cluster B were utilized for validation, and vice versa.

C. Area of Applicability

The AOA metric [25] serves as a tool to evaluate the spatial extent within which an ML model can be reliably applied. In spatial prediction studies, ML models are typically trained using spatially distributed data obtained from ground stations. However, the model needs to make prediction in new geographic areas, which may possess different feature combinations and relationships to those present in the model training data. AOA delineates the area where the model can apply relationships learned from training data, ensuring that the estimated CV performance remains valid. It begins with a standardization procedure that ensures equivalent scaling across all predictor variables through the implementation of z-score normalization

$$X_{i,j}^s = \frac{X_{i,j} - \bar{X}_j}{\sigma_j} \quad (2)$$

where $X_{i,j}^s$ denotes the standardized value of the j th predictor variable for the i th observation (i.e., $X_{i,j}$); and $\bar{X}_j$ and σ_j represent the arithmetic mean and standard deviation of the

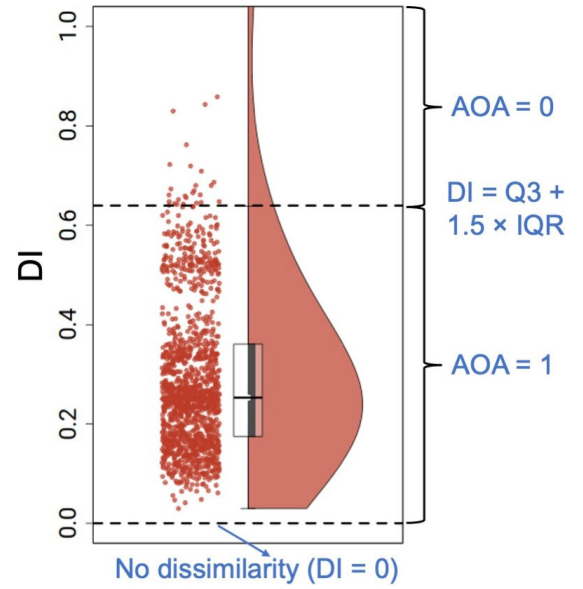


Fig. 3. Illustration of the binary determination of AOA using the RF model based on a fourfold spatial CV.

j th predictor variable, respectively. Furthermore, ML models reveal differential predictive power across variables [66]. To incorporate this heterogeneity, an importance-weighted standardization is implemented

$$X_{i,j}^{sw} = w_j X_{i,j}^s \quad (3)$$

where w_j is the weight derived from the variable importance for the j th predictor variable. The Euclidean distance between a given prediction location k and a training location i in the predictor variable space can be expressed as follows:

$$d(k, i) = \sqrt{\sum_{j=1}^p (X_{k,j}^{sw} - X_{i,j}^{sw})^2} \quad (4)$$

where d denotes the Euclidean distance and p is the number of predictor variables. The minimal distance is then given as follows:

$$d_k = \arg \min_i (d(k, i)). \quad (5)$$

The dissimilarity index (DI) for a given prediction location (i.e., DI_k) is defined as follows:

$$DI_k = d_k / \bar{d} \quad (6)$$

where $\bar{d}$ is the arithmetic mean value of all pairwise distances between training data. The DI for training instances should be computed following a CV, where distances are calculated to the nearest training point outside the instance's validation fold. The value of DI ranges from 0 to ∞ , where $DI_k = 0$ represents the prediction location k that has no dissimilarity (i.e., identical characteristics) with a training location (Fig. 3).

The outlier-removed maximum DI is used as a threshold to determine a binary AOA, which is given as follows:

$$AOA_k = \begin{cases} 1, & \text{if } DI_k \leq DI_{Q3} + 1.5 \times DI_{IQR} \\ 0, & \text{if } DI_k > DI_{Q3} + 1.5 \times DI_{IQR} \end{cases} \quad (7)$$

where $Q3$ and IQR denote the third quartile and interquartile range of a boxplot, respectively. When $AOA_k = 1$, it represents

the characteristics of predictor variables at location k that are considered similar to those of training data (Fig. 3), thus the prediction model is considered applicable at location k .

The exhibition of DI estimations and AOA determinations was advocated in conjunction with the prediction of ML models [25]. These metrics can help encapsulate spatial constraints and zones of reliable prediction for ML models, thereby complementing the traditional validation metrics and prevalent estimates of uncertainty. For more details, please see [25].

D. Evaluation Metrics

Bias, unbiased root mean squared error (ubRMSE), and Pearson correlation coefficient (R) were used to evaluate the performance of spatiotemporal fusion and ML models, which are respectively given as follows:

$$\text{Bias} = \frac{\sum_{i=1}^N (y_i - x_i)}{N} \quad (8)$$

$$\text{ubRMSE} = \sqrt{\frac{\sum_{i=1}^N ((y_i - \bar{y}) - (x_i - \bar{x}))^2}{N}} \quad (9)$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (10)$$

where y_i denotes the time series of estimated data; x_i denotes the time series of reference data; $\bar{y}$ and $\bar{x}$ are the mean values of the estimated and reference data, respectively; and N is the number of observations of the time series.

IV. RESULTS

A. Evaluation of Downscaled Predictor Variables

Fig. 4 shows the downscaled performance of (a) albedo, (b) NDVI, and (c) LST versus MODIS data during January 1, 2016–December 31, 2019 using binned density scatterplots. The validation employed a point-based comparison strategy, where both MODIS and downscaled data were extracted at precisely matched OzNet monitoring locations ($n = 14118$) through coordinate-based spatial matching. Downscaled albedo had a bias of 0, ubRMSE of 0.03, and R of 0.76, with values clustering around 0.2. NDVI showed a slight negative bias of -0.02 , ubRMSE of 0.08, and R of 0.87, with values mostly between 0.2 and 0.3. LST had a bias of 0.17 K, ubRMSE of 1.14 K, and R of 0.99, with values concentrated around 290 K and 305–315 K. Overall, the downscaled predictor variables demonstrated consistency with MODIS data, as evidenced by the established metrics and concentration patterns on the 1:1 line as observed in the scatterplots.

Fig. 5 provides an illustrative example from April 2, 2017 showcasing the spatial comparison between MODIS data and downscaled predictor variables. Fig. 5(g)–(i) presents downscaled patterns of predictor variables across the entire study area, demonstrating consistency with MODIS data but exhibiting sharpened features in comparison [Fig. 5(a)–(c)]. Upon closer examination of a zoomed area [Fig. 5(j)–(l)], the downscaled predictor variables revealed enhanced details that are more visually discernible compared to the MODIS data

[Fig. 5(d)–(f)]. Notably, features such as the vegetated areas [observed in Fig. 5(k)] are more distinct in the downscaled predictor variables. In general, the spatiotemporal fusion enabled a better capture of spatial details of predictor variables at the field scale, offering improved visualization and representation of the landscape.

B. Evaluation of Model Performances

Fig. 6 presents boxplots of (a) bias, (b) ubRMSE, and (c) R for a fourfold spatial CV of RF and XGB models, respectively, between January 1, 2016 and December 31, 2019. The median bias was approximately $0.00 \text{ m}^3/\text{m}^3$ for RF and $-0.01 \text{ m}^3/\text{m}^3$ for XGB, with bias values ranging from -0.05 to $0.05 \text{ m}^3/\text{m}^3$ for both models. The median ubRMSE was around $0.05 \text{ m}^3/\text{m}^3$ for both models, with values between 0.03 and $0.12 \text{ m}^3/\text{m}^3$. The median R was approximately 0.79 for RF and 0.78 for XGB, with RF showing slightly better performance. These metrics indicate that both models performed comparably in estimating SM contents across the OzNet sites.

Fig. 7 presents boxplots of bias, ubRMSE, and R for the CC validation of (a)–(c) RF and (e)–(f) XGB model, respectively, between January 1, 2016 and December 31, 2019. In cluster A, RF and XGB had similar performance, with median bias, ubRMSE, and R values of about $-0.02 \text{ m}^3/\text{m}^3$, $0.08 \text{ m}^3/\text{m}^3$, and 0.70, respectively (Fig. 7; cluster A columns). In cluster B, RF slightly outperformed XGB with median bias, ubRMSE, and R values of $0.05 \text{ m}^3/\text{m}^3$, $0.05 \text{ m}^3/\text{m}^3$, and 0.71, respectively, compared to XGB's $0.06 \text{ m}^3/\text{m}^3$, $0.06 \text{ m}^3/\text{m}^3$, and 0.68 (Fig. 7; cluster B columns). When trained on data from both clusters A and B, both models captured SM patterns outside these clusters (Fig. 7; remaining 4 sites columns), with RF (XGB) showing bias, ubRMSE, and R values of 0.02 ($0.01 \text{ m}^3/\text{m}^3$), 0.05 ($0.05 \text{ m}^3/\text{m}^3$), and 0.78 (0.80), respectively. In summary, these boxplots showcase the spatial transferability of the ML models, demonstrating the capability of both models to maintain consistent performance across different spatial clusters within a distance of roughly 20 km (see Fig. 1).

Fig. 8 shows the violin plots of SHAP values of six most important predictor variables of (a)–(c) RF model and (d)–(f) XGB model between January 1, 2016 and December 31, 2019 for all sites, cluster A, and cluster B, respectively. For all sites, both models identified albedo, VPD, NDVI, ET, LST, and DEM as the top predictor variables, though their order differed [Fig. 8(a)–(d)]. In cluster B for RF [Fig. 8(c)] and cluster A for XGB [Fig. 8(e)], these predictor variables remained most important. However, for RF in cluster A [Fig. 8(b)] and XGB in cluster B [Fig. 8(f)], S_{rad} emerged as a new predictor, ranking fifth and sixth, respectively. The color-coded normalized feature values (yellow to purple) reveal each predictor variable's contribution: moderate to high VPD and LST generally contribute to the reduction of predicted SM values, while high ET and NDVI would contribute to the increase. Conversely, the values of albedo, DEM, and S_{rad} exhibited more variability and lacked specific trends in predicting SM values. Overall, while predictor rankings varied by region, the most important predictor variables generally remained consistent. These variations highlight the models' reliance on

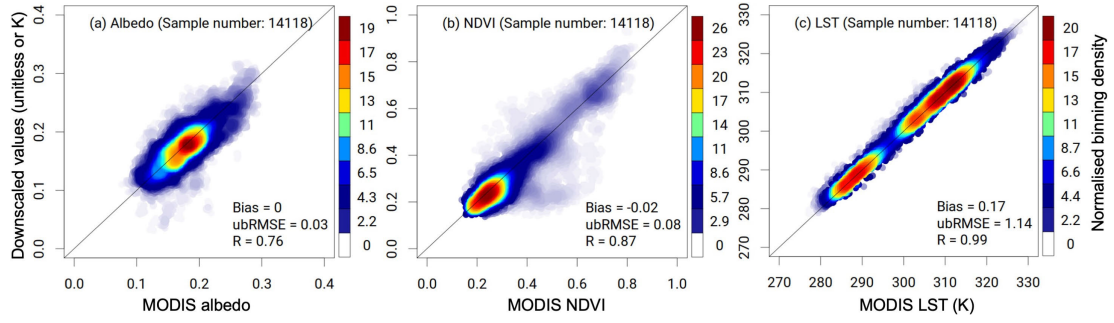


Fig. 4. Binned density scatterplots of (a) albedo, (b) NDVI, and (c) LST between January 1, 2016 and December 31, 2019 (excluding training dates of spatiotemporal fusion).

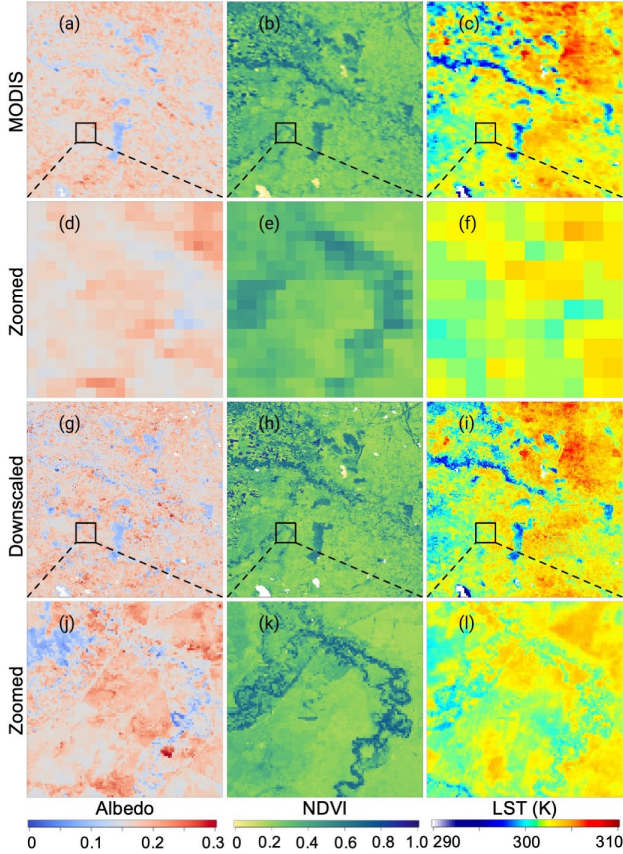


Fig. 5. Spatial comparison between MODIS and downscaled predictor variables on April 2, 2017. (a)–(c) MODIS albedo, NDVI, and LST, respectively. (d)–(f) Zoomed windows of MODIS predictor variables spanning 146.25°E–146.35°E and 34.92°S–35.02°S (i.e., cluster B). (g)–(i) Downscaled albedo, NDVI, and LST, respectively. (j)–(l) Zoomed windows of downscaled predictor variables covering the same zoomed area with (d)–(f).

regional land cover and landscape characteristics, influencing the importance and impact of specific predictor variables.

Fig. 9 presents a Spearman correlation matrix illustrating the pairwise relationships for all predictor variables used in this study. Notable observations include strong negative correlations between clay and silt/sand content, and strong positive correlations among LST, T_{avg} , VPD, and S_{rad} , reflecting their interconnected nature. NDVI demonstrated moderate negative correlations with LST, albedo, T_{avg} , VPD, and S_{rad} . Conversely, ET demonstrated moderate positive correlations with LST, NDVI, T_{avg} , VPD, and S_{rad} . This matrix identified potential collinearities among predictors, thereby complementing the feature importance as given by SHAP values.

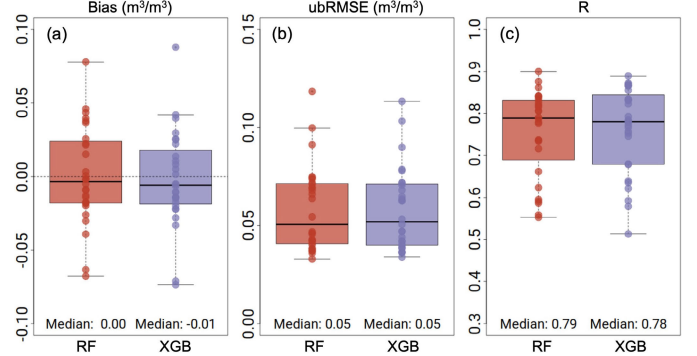


Fig. 6. Boxplots of (a) bias, (b) ubRMSE, and (c) R for a fourfold spatial CV of RF and XGB models, respectively, between January 1, 2016 and December 31, 2019. The details of sites included in each fold are given in Table II. The scatters represent the metrics of individual sites.

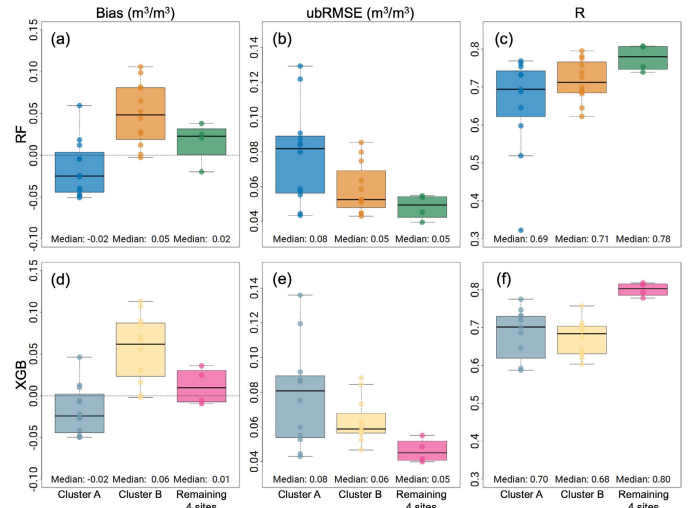


Fig. 7. Boxplots of bias, ubRMSE, and R for the CC validation of (a)–(c) RF and (d)–(f) XGB model, respectively, between January 1, 2016 and December 31, 2019. Metrics within cluster A were derived from models trained using data within cluster B, and vice versa. Metrics for the remaining four sites (see Fig. 1 and Table II) were derived from models trained using data from both clusters A and B. The details of sites included in each cluster are given in Table II. The scatters represent the metrics of individual sites.

C. AOA of Spatial SM Predictions

Fig. 10 shows the spatial distribution of the median DI (with values exceeding five masked out) and AOA based on CV and CC for (a)–(d) RF and (e)–(h) XGB, respectively, between January 1, 2016 and December 31, 2019. Both models showed low DI in the western region, regardless of CV or CC validation [Fig. 10(a)–(g)]. However, DI increased drastically in the eastern area due to no training sites and

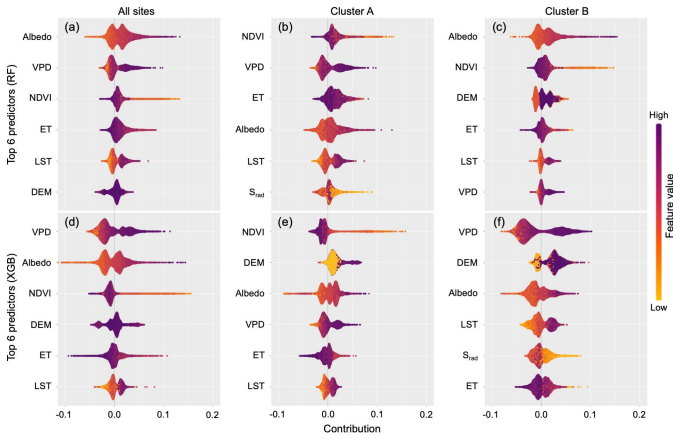


Fig. 8. Violin plots of SHAP values of six most important predictor variables of (a)–(c) RF and (d)–(f) XGB between January 1, 2016 and December 31, 2019 for all sites, cluster A, and cluster B, respectively. The predictor variables are displayed in the order of importance.

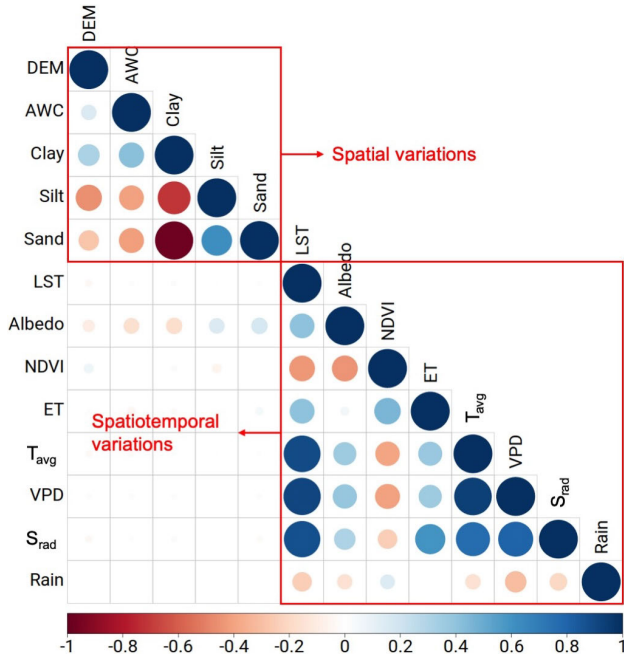


Fig. 9. Spearman correlation matrix of the pairwise relationships for the full spectrum of 13 predictor variables.

distinct landscape patterns (see Fig. 1). The DI thresholds for AOA determination were 0.64 (CV) and 0.87 (CC) for RF [Fig. 10(a)–(c)], almost double those for XGB, which were 0.34 (CV) and 0.44 (CC) [Fig. 10(e)–(g)]. AOA for RF (XGB) based on CV was 43.1% (41.5%), while for CC it was 25.4% (26.9%). Predictor variables within the determined AOA were deemed comparable to those used in training. It is expected that spatial SM prediction within AOA can demonstrate comparable performance metrics to that of the CV (see Fig. 6) and CC (see Fig. 7). The CV approach (using 21 sites each time) also demonstrated capability to effectively expand AOA compared to CC (using 12 sites each time).

Fig. 11 provides spatial examples of predicted SM using RF and XGB models on February 1, 2017 (austral summer) and August 1, 2017 (austral winter) for the entire study area, and zoomed areas based on CV or CC, respectively. AOA was displayed as the western area within the black boundaries [Fig. 11(a), (f), (k), and (p)], covering both spatial

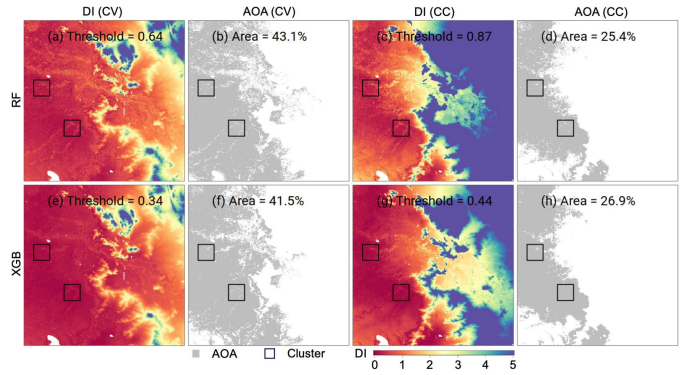


Fig. 10. Spatial distribution of median DI (with values exceeding five masked out) and AOA based on performance of CV and CC for (a)–(d) RF and (e)–(h) XGB, respectively, between January 1, 2016 and December 31, 2019. The derived DI and AOA had an identical temporal frequency with the spatial SM prediction (i.e., daily). DI was composited using its median values, and AOA was composited using its majority values (i.e., >50%). Threshold values indicate the maximum permissible DI for reliable predictions, and area percentages represent the proportion of the study region within the AOA. Black rectangles indicate the cluster locations. The mask value of 5 was empirically determined to optimize visualization clarity while maintaining representation of all relevant DI gradients that influence AOA delineation.

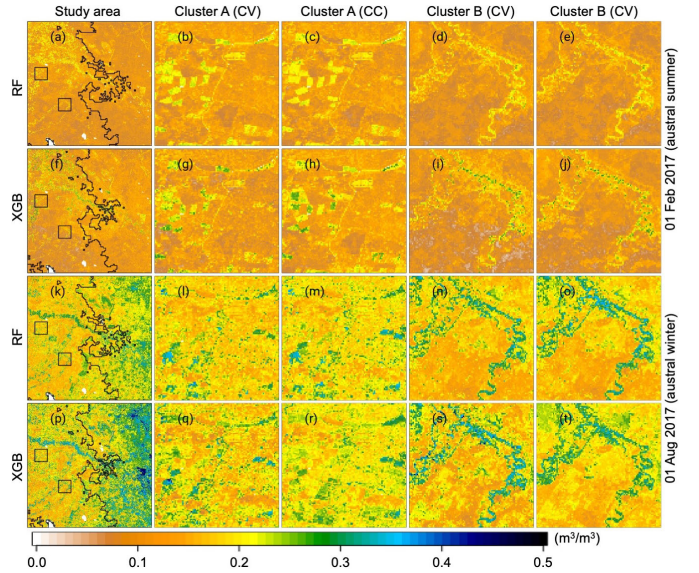


Fig. 11. Spatial examples of predicted SM using RF (a)–(e); (k)–(o) and XGB (f)–(j); (p)–(t) models on February 1, 2017 (austral summer) and August 1, 2017 (austral winter) for the study area (first column), zoomed areas based on CV (second and fourth columns), and zoomed areas based on CC (third and fifth columns), respectively. AOA was displayed as the western area of the black boundaries, encompassing both spatial clusters.

clusters. On February 1, 2017 (austral summer), predicted SM values were relatively dry ($0.1\text{--}0.2\text{ m}^3/\text{m}^3$), with both models yielding similar results [Fig. 11(a)–(f)], though XGB showed slightly higher SM values. Zoomed areas derived from CV and CC also appeared visually comparable [e.g., Fig. 11(b), (c), (g), and (h)]. On August 1, 2017 (austral winter), predicted SM values were higher ($0.2\text{--}0.5\text{ m}^3/\text{m}^3$) and more variable, especially in the eastern regions outside AOA [Fig. 11(k)–(p)]. These disparities can be attributed to both models learning only from data within the AOA, rendering conditions outside the AOA as “unknown space” and thus more uncertain. The results within zoomed areas, derived from both CV and CC, also exhibited greater variability [e.g., Fig. 11(l), (m), (q), and (r)]. In summary, this detailed visualization highlights

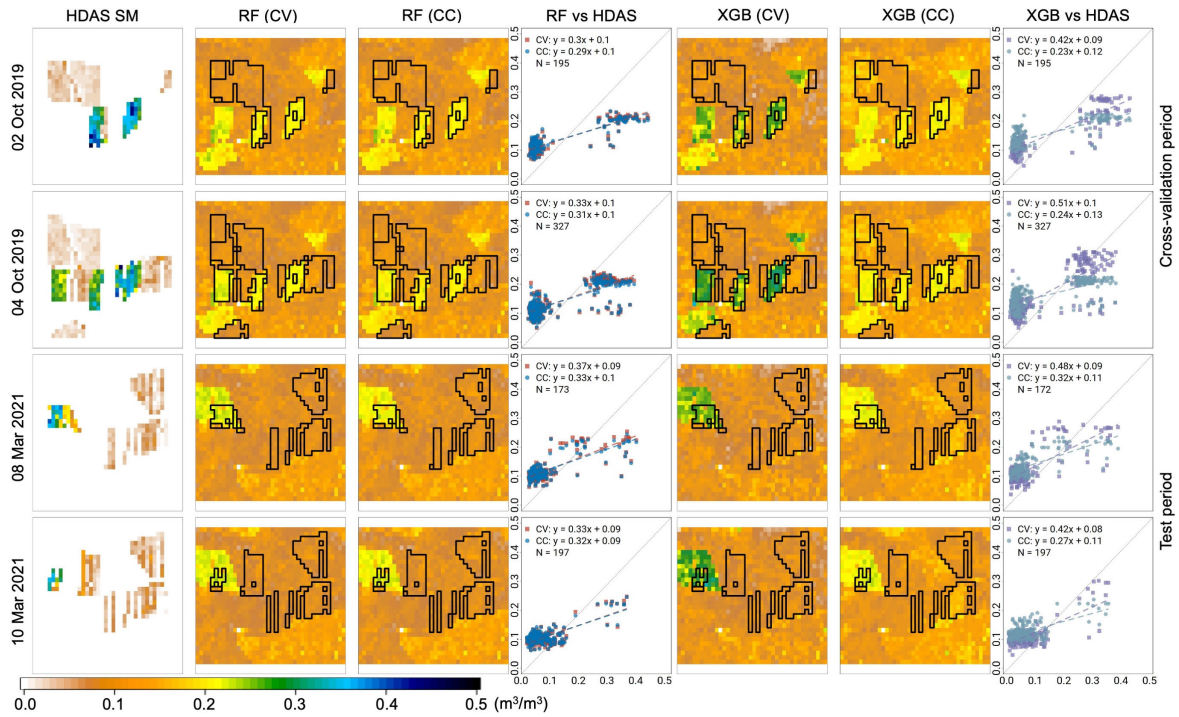


Fig. 12. Spatial examples of HDAS SM and predicted SM using RF and XGB on October 2, 2019 and October 4, 2019 (CV period) and March 8, 2021 and March 10, 2021 (test period), respectively. The first column shows the interpolated HDAS SM at 100 m resolution grids. The black polygons within second, third, fifth, and sixth columns from left are derived from the spatial extent of HDAS SM. The fourth and seventh columns from left are scatterplots of predicted SM derived from CV and CC, using RF and XGB, against HDAS SM, respectively.

the capability of AOA in assessing the reliability of predicted SM at the field scale. The observed variations in SM across landscapes underscore the importance of considering regional nuances in training strategies for accurate and context-specific predictions.

D. Evaluations Against Independent Data

The AOA metric indicated that the RF and XGB models were anticipated to reliably predict the SM dynamics within approximately 43.1% and 41.5% of the study region, respectively, as determined by CV metrics. To validate AOA, both models were assessed against independent data from field campaigns, CosmOz, OzFlux, and SMAP. Fig. 12 shows the spatial examples of HDAS SM collected during field campaigns and predicted SM using the RF and XGB models on October 2, 2019 and October 4, 2019 (within the CV period) and March 8, 2021 and March 10, 2021 (within the test period), respectively. The 4×3 km field campaign area was within the AOA (see Figs. 1–10). Both models captured the spatial patterns of wet regions. For RF, predictions from CV and CC were nearly identical (Fig. 12, second to fourth columns), but both underestimated SM in wet regions. For XGB, CV predictions had higher values than CC (Fig. 12, fifth–seventh columns) and better matched HDAS reference grids (Fig. 12, first column). In summary, both the RF and XGB models demonstrated capabilities to capture the correct spatial patterns of SM within the HDAS field campaign area, with the XGB model based on CV providing the closest prediction to HDAS SM values.

Table III details bias, ubRMSE, and R metrics for spatial SM prediction against HDAS SM on specific field campaign

dates in September to October 2019 (within the CV period) and March 2021 (within the test period). Both models maintained stable mean bias values of 0.04 – 0.05 m^3/m^3 in both periods. The ubRMSE of both models even reduced within the test period, with mean values of 0.06 m^3/m^3 compared to 0.08 – 0.09 m^3/m^3 within the CV period. The most marked difference was in R performance. The CV-based RF consistently showed better or comparable R values (0.60 – 0.62) than CV-based XGB (0.53 – 0.58). The CC-based RF also performed slightly better (0.64 – 0.65) than CC-based XGB (0.62 – 0.64). In summary, both the RF and XGB models demonstrated comparable performances in bias and ubRMSE within both the CV and test periods. However, RF showed more robust R performance with moderate values (0.60 – 0.65), while XGB exhibited more variability (0.53 – 0.64).

Figs. 13 and 14 present the time series of predicted SM (both CV and CC), CosmOz and OzFlux SM, and rainfall throughout the CV period (2016–2019) and test period (2020–2021), respectively. In the CV period, the temporal variations of RF based on CV and CC were visually identical, exhibiting high agreement with CosmOz SM [$R = 0.91$; Fig. 13(a)] and OzFlux SM [$R = 0.84$ – 0.85 ; Fig. 13(c)]. As a comparison, the temporal variations of CV-based XGB differed from those based on CC. The R performance of XGB ranged from 0.86 to 0.90 against CosmOz SM [Fig. 13(b)] and was 0.85 against OzFlux SM [Fig. 13(d)], respectively, indicating its sensitivity to training samples.

In the test period, the temporal variations of SM predicted by RF based on CV and CC remained visually similar, exhibiting moderate agreement with CosmOz SM [$R = 0.74$ – 0.75 ; Fig. 14(a)] and OzFlux SM [$R = 0.61$ – 0.63 ; Fig. 14(c)]. Comparatively, the temporal variations of SM predicted by XGB

TABLE III

METRICS OF BIAS, ubRMSE, AND R OF SPATIAL SM PREDICTION AGAINST HDAS SM ON SPECIFIC FIELD CAMPAIGN DATES IN SEPTEMBER TO OCTOBER IN 2019 (WITHIN CV PERIOD) AND MARCH IN 2021 (WITHIN TEST PERIOD). THE METRICS ARE BASED ON PIXELWISE COMPARISONS

Period	Date	Bias (m^3/m^3)				ubRMSE (m^3/m^3)				R				Sample numbers
		CV	XGB	CC	XGB	CV	XGB	CC	XGB	CV	XGB	CC	XGB	
Cross-validation	30 Sep 2019	0.02	0.04	0.03	0.03	0.10	0.11	0.11	0.11	0.81	0.81	0.86	0.85	147
	02 Oct 2019	0.02	0.03	0.01	0.02	0.09	0.11	0.10	0.10	0.89	0.81	0.92	0.90	195
	04 Oct 2019	0.05	0.05	0.03	0.03	0.07	0.10	0.08	0.09	0.83	0.69	0.81	0.80	327
	07 Oct 2019	0.06	0.05	0.04	0.04	0.09	0.09	0.09	0.09	0.51	0.41	0.51	0.51	320
	11 Oct 2019	0.07	0.06	0.05	0.05	0.09	0.09	0.08	0.08	0.37	0.30	0.42	0.42	314
	14 Oct 2019	0.04	0.03	0.03	0.03	0.08	0.07	0.07	0.07	0.34	0.30	0.46	0.45	157
	18 Oct 2019	0.06	0.05	0.06	0.06	0.05	0.06	0.05	0.05	0.63	0.54	0.63	0.65	242
	All	0.05	0.05	0.04	0.04	0.08	0.09	0.08	0.08	0.62	0.53	0.64	0.64	1702
Test period	08 Mar 2021	0.05	0.06	0.04	0.04	0.07	0.07	0.07	0.07	0.72	0.72	0.77	0.76	173
	10 Mar 2021	0.05	0.06	0.04	0.04	0.05	0.05	0.05	0.05	0.73	0.61	0.74	0.73	197
	17 Mar 2021	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.48	0.48	0.55	0.47	162
	19 Mar 2021	0.03	0.02	0.02	0.02	0.06	0.06	0.06	0.06	0.41	0.48	0.49	0.46	138
	All	0.05	0.05	0.04	0.04	0.06	0.06	0.06	0.06	0.60	0.58	0.65	0.62	670

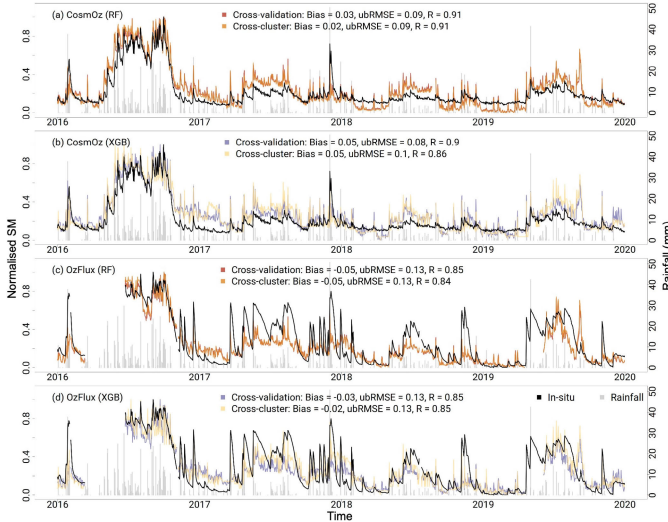


Fig. 13. Time series of predicted SM (both CV and CC), CosmOz and OzFlux SM, and rainfall during the CV period (2016–2019). The unit of SM herein is wetness (unitless), which is different with that of OzNet and HDAS SM (m^3/m^3). This is because CosmOz only provides normalized SM data ranging between 0 and 1. (a) CosmOz (RF). (b) CosmOz (XGB). (c) OzFlux (RF). (d) OzFlux (XGB).

based on CV still diverged from those based on CC. The R performance of SM predicted by XGB was 0.70–0.71 against CosmOz SM [Fig. 14(b)] and was 0.59–0.66 against OzFlux SM [Fig. 14(d)], respectively. In summary, these temporal metrics indicated that RF exhibited slightly better agreement with independent SM time series and displayed more robustness irrespective of variations in training sample numbers.

Fig. 15 presents the spatial distribution of bias, ubRMSE, and R of SM predicted by the RF and XGB models against SMAP L2 data. The evaluation covers the entire study area, cluster A, and cluster B, spanning both the CV period [2016–2019; Fig. 15(a)–(r)] and test period [2020–2021; Fig. 15(s)–(aj)], respectively. Spatial bias and ubRMSE for both models were comparable in both periods, ranging from -0.02 to $0.00 \text{ m}^3/\text{m}^3$ and 0.06 to $0.07 \text{ m}^3/\text{m}^3$, respectively. Notably, during the CV period, the spatial SM prediction showed a negative bias (approximately $-0.03 \text{ m}^3/\text{m}^3$) and

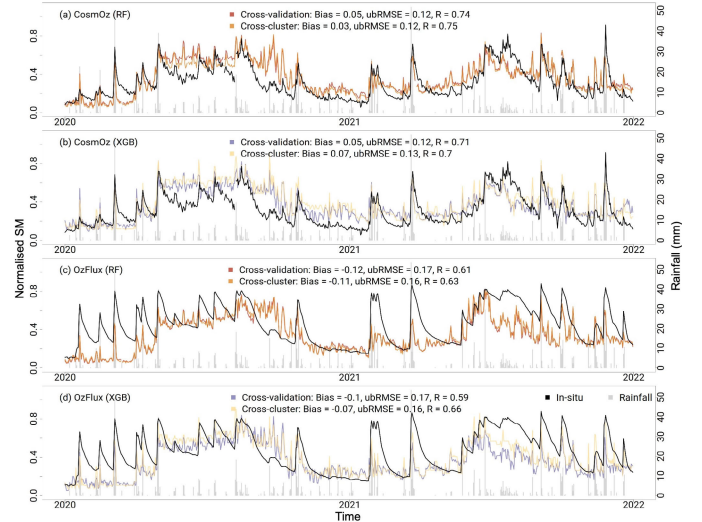


Fig. 14. Same with Fig. 13 but for the test period (2020–2021). (a) CosmOz (RF). (b) CosmOz (XGB). (c) OzFlux (RF). (d) OzFlux (XGB).

lower ubRMSE ($<0.06 \text{ m}^3/\text{m}^3$) within AOA compared to positive bias (0 – $0.03 \text{ m}^3/\text{m}^3$) and higher ubRMSE ($>0.06 \text{ m}^3/\text{m}^3$) outside AOA [Fig. 15(a), (b), (d), and (e)]. Regarding the spatial R performance, SM predicted by RF demonstrated slightly better agreement with SMAP data, with a median value of 0.87 [Fig. 15(c)], compared to XGB prediction with a median value of 0.86 [Fig. 15(f)] within the CV period. However, in the test period, both models achieved relatively moderate R agreement (0.75) with SMAP SM. Unlike bias and ubRMSE, the R performance within AOA did not show marked differences compared to areas outside the AOA. In summary, these findings highlight the capability of AOA in delineating a spatial extent conducive to reliable prediction, particularly evident in bias and ubRMSE metrics.

V. DISCUSSION

Upscaling emerges as an attractive strategy to reconcile the spatial mismatch between point-scale data and satellite-derived SM grids [21], [22], [23], [67]. However, previous studies were often restricted to localized conditions and have overlooked the

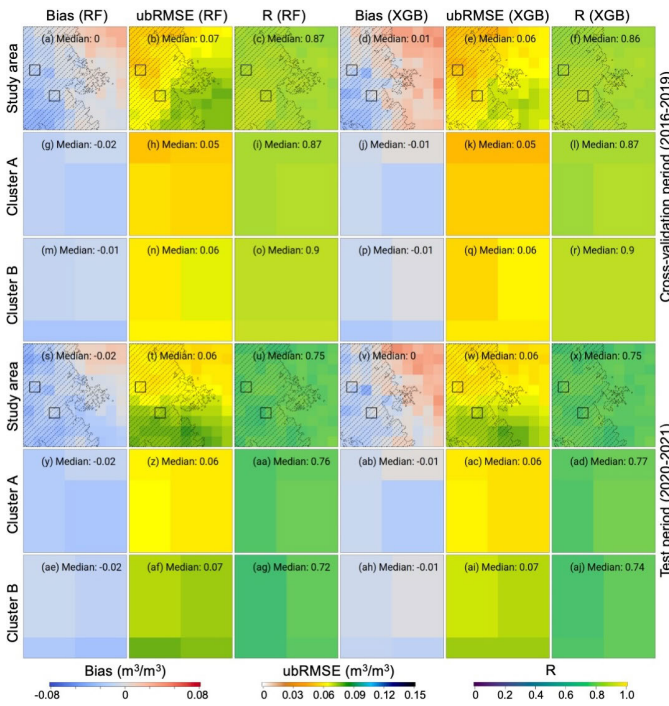


Fig. 15. Spatial distribution of bias, ubRMSE, and R of SM predicted by the RF and XGB models against SMAP L2 for the entire study area, cluster A, and cluster B within CV period [2016–2019; (a)–(r)] and test period [2020–2021; (s)–(aj)], respectively. The spatial SM prediction was aggregated to 0.1° (~ 10 km) resolution grids. AOA was displayed as the western area of the black boundaries, encompassing both spatial clusters.

quantification of spatial uncertainty in SM estimates. In this study, a data-driven approach was introduced to spatially predict SM at a 100 m resolution in an Australian agricultural region. The AOA of two ML models was determined, both of which covered roughly 40% of the study region, and evaluated against multiple independent datasets. A few discussion points are summarized as follows.

A. Further Assessment of AOA

The spatial pattern of AOA, as shown in Fig. 10, exhibited noticeable sensitivity to land cover and topographic features. To fully explore the spatial characteristics of AOA in SM prediction, a further assessment using an extended region in 2019 was conducted. The extended AOA analysis, encompassing a broader spatial extent beyond the primary study area, demonstrates alignment with natural landscape boundaries (Fig. 16), particularly evident in the correspondence between AOA delineation and cropland extent shown in Fig. A1(a) and (b). This alignment is further reinforced by the distinct gradient in DI values that traces the transition from agricultural to nonagricultural lands. The spatial validation against SMAP L2 retrievals substantiates the AOA's effectiveness (Fig. 17), particularly evident in capturing the spatial area with relatively lower bias [Fig. 17(a)], compared to positive/negative values in peripheral zones. This spatial coherence between AOA delineation and error structure demonstrates the metric's capacity to identify domains where the model's characteristics remain consistent with the training conditions, thereby providing a quantitative basis for constraining prediction reliability across heterogeneous landscapes. The DI distributions for both CV

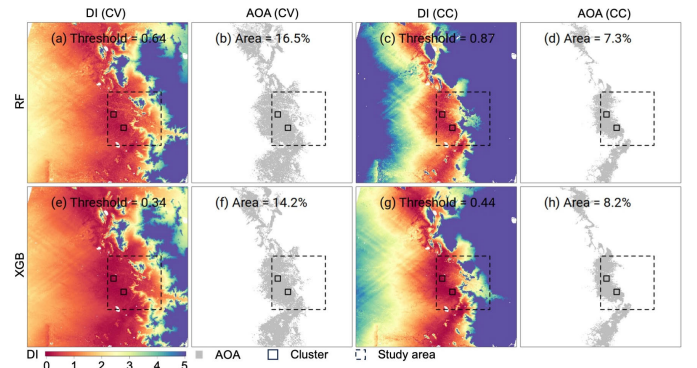


Fig. 16. Spatial distribution of median DI (with values exceeding five masked out) and AOA based on performance of CV and CC for (a)–(d) RF and (e)–(h) XGB for the extended region, respectively, between January 1, 2019 and December 31, 2019. The explanations of AOA and DI are the same as in Fig. 10. Black and dashed rectangles indicate the cluster locations and the primary study area, respectively. The detailed spatial information of the extended area can be found in Fig. A1.

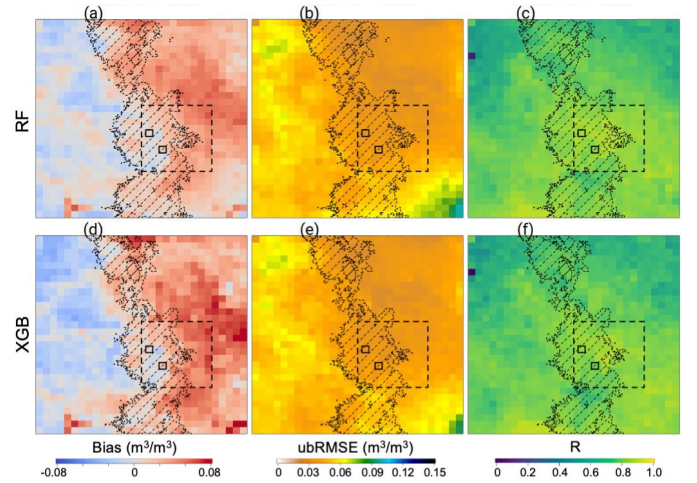


Fig. 17. Spatial distribution of bias, ubRMSE, and R metrics of SM predicted by the RF and XGB models against SMAP L2 for the extended region in 2019, respectively. The spatial SM prediction was aggregated to 0.1° (~ 10 km) resolution grids. AOA was displayed as the black polygons filled with dashed lines, encompassing the primary study area (dashed rectangle) and both spatial clusters (solid rectangles). The detailed spatial information of the extended area can be found in Fig. A1. (a) Median: 0.01. (b) Median: 0.05. (c) Median: 0.78. (d) Median: 0.02. (e) Median: 0.04. (f) Median: 0.77.

and CC validation approaches are constrained by these landscape features, with threshold values systematically higher in regions of homogeneous land cover and gentle topography (see Figs. 16–18 and A1). This systematic correspondence between AOA boundaries and model performance metrics validates the capability of AOA to identify reliable prediction zones based on landscape characteristics, while simultaneously highlighting the challenges of extrapolating SM predictions across distinct landscape units.

A quantitative analysis was also conducted to characterize surface properties within and outside the AOA boundaries, enabling the assessment of the model's applicability across distinct physiographic domains. Fig. 18 presents the distribution of normalized DEM (DEM_{norm}), normalized AWC (AWC_{norm}), clay, sand, and area equipped for irrigation (Food and Agriculture Organization) [68] for the regions (a) within AOA and (b) outside AOA. Regions with higher elevation and steeper slopes [Fig. 18(b)] may experience different precipitation patterns

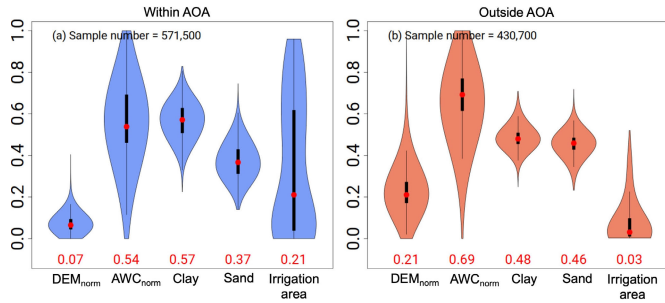


Fig. 18. Violin plots of normalized DEM (DEM_{norm}), normalized AWC (AWC_{norm}), clay, sand, and area equipped for irrigation [68] for the regions (a) within AOA and (b) outside AOA. The red dots within the violins represent the median, with the decimal numbers at the bottom indicating the numeric values. The analysis is restricted to the primary study area (Fig. 1; 100×100 km) that exhibits comparable pixel distributions within and outside AOA boundaries.

and runoff behaviors compared to low-lying, flat regions [Fig. 18(a)], thereby impacting the spatial distribution of SM. The flat region within AOA had more intensive irrigation activities [Fig. 18(a)], with the normalized irrigation area (0.21) being seven times of that outside AOA (0.03). These irrigation activities are likely to homogenize SM patterns within AOA through reduced subpixel heterogeneity, making them more consistent and predictable. While explicit incorporation of irrigation factors remains methodologically challenging due to the complex interplay between management practices and natural hydrological processes, the AOA metric effectively captures these anthropogenic influences through their manifestation in predictor variables and resulting SM patterns. Additionally, the AOA pattern provided clear insights into the regional heterogeneity within the study area (as illustrated in Fig. 1). Unlike the regions within AOA, the eastern region, marked by more rugged and varied topography, was demonstrated to have higher uncertainty (as in comparison against SMAP L2; Fig. 15). The heterogeneity in the eastern region introduces variability in microclimates and soil properties, posing challenges for the ML models to generate reliable SM prediction. The regional differences highlighted by the AOA suggest the need for localized calibration of spatial prediction models to account for spatial heterogeneity. Furthermore, the land cover heterogeneity can also be an important factor to be considered in future efforts [13], as AOA delineates regions where environmental conditions are considered similar with those represented in the training dataset. The degradation of ML model performance across heterogeneous regions can be attributed to insufficient representation of specific multivariate environmental conditions, consistent with recent findings [13] that site representativeness improves with increased land cover homogeneity. Future studies should incorporate explicit land cover similarity metrics as auxiliary variables to enhance model interpretation across diverse physiographic domains.

The ML modeling processes leveraged the SHAP method [63], [64], which offered an avenue for gaining insights into the underlying principles governing SM interactions with various predictor variables from a data-driven perspective. The implementation of SHAP integrated with the underlying model structures: RF's ensemble averaging mechanism across independently constructed trees versus XGB's

sequential gradient boosting approach optimizing residual errors. These intrinsic algorithmic architectures, through which the two models construct feature–response relationships, are reflected in the observed differential ranking patterns of predictor variables between RF and XGB models. Specifically, these differences manifested in the violin plots' distribution morphologies, with RF exhibiting more symmetrical contribution patterns [Fig. 8(a)–(c)] compared to XGB's pronounced multimodality and wider contribution ranges [Fig. 8(d)–(f)]. Despite the algorithmic variations in SHAP value distributions, both models identified consistent dominant predictor variables (i.e., albedo, DEM, ET, LST, NDVI, and VPD) in the spatial CV [Fig. 8(a)–(d)], though their ranking orders exhibited nuances within different clusters characterized by varying environmental conditions.

Regarding the absence of soil texture variables (i.e., clay, sand, and silt) among the most important predictor variables, several factors may contribute to this phenomenon. First, there is a covarying relationship between AWC and soil texture data, leading to redundant information that may reduce the importance of soil texture predictor variables in ML models. Second, since this study focused on the top 0–5 cm of soil, the homogeneity of soil mapping for AWC and texture within this shallow depth range across the study areas, particularly within the AOA envelope, could diminish the distinct contribution of soil texture variables. Third, given the spatiotemporal variabilities inherent in SM dynamics, static soil texture data may not adequately capture the dynamic nature of SM variations over time, leading to their lower importance compared to the dynamically sensed predictor variables. If the objective was to predict long-term mean SM, static soil texture data might prove more contribution, as they reflect the overall capacity of a soil type to hold moisture. However, when modeling seasonal and interannual variability, the signal from static soil texture data is likely overwhelmed by the covarying predictors that capture dynamic changes. Finally, considering that the training data clusters cover small areas, it is plausible that the soil properties are relatively homogeneous within these clusters, further reducing the distinct contribution of soil texture variables in the models.

B. Strengths and Prospects

Traditional SM upscaling approaches encompass a spectrum of spatial interpolation methodologies, from simple geometric interpolators such as arithmetic averaging [19] and Thiessen polygons [69], to distance-weighted techniques including inverse distance weighting [70] and various kriging implementations [71]. While these methods effectively utilize point-scale SM measurements, they operate independently of auxiliary remote sensing data that could potentially enhance spatial predictions. More recently, physically based approaches incorporating ATI, derived from the relationship between LST dynamics and soil thermal properties [20], have emerged as complementary techniques. However, current upscaling methodologies face several fundamental challenges. Geostatistical methods such as kriging, while highly effective near observation points, require densely distributed training data

TABLE IV
OVERVIEW OF REPRESENTATIVE SM UPSCALING STUDIES. STUDIES ARE ORDERED CHRONOLOGICALLY BY YEAR OF PUBLICATION

Ground network	Method and predictor variables	Objectives	Key findings	Potential limitations	Reference
Soil Moisture / Temperature Monitoring Network (SMTMN), Tibet Autonomous Region, China	Bayesian linear regression with MODIS-derived apparent thermal inertia (ATI).	Upscale in-situ SM to an area-averaged scale that is comparable to satellite footprints.	The upscaling approach reduced RMSE from 0.023 to 0.013 m ³ /m ³ , with the temporal pattern of upscaled SM agreeing with local hydro-meteorological knowledge.	The results lacked evaluations against independent data (e.g., from satellite missions or field campaigns).	[20]
Heihe Water Allied Telemetry Experimental Research experiment (HiWATER), Gansu Province, China	Spatiotemporal regression block kriging (STRBK) model with information of vegetation and LST.	Upscale in-situ SM observations to validate the Polarimetric L-band Multi-beam Radiometer (PLMR) SM data.	STRBK can yield upscaled SM with satisfactory accuracy for validation purposes. PLMR SM showed an RMSE of 0.03 m ³ /m ³ against upscaled SM and was considered applicable for watershed-scale SM monitoring.	The collected in-situ SM only spanned a 28-day period and the evaluations of PLMR SM were only conducted on two days.	[21]
Soil moisture Sensing Controller And oPtimal Estimator (SoilSCAPE) network, California, United States	RF model with various landscape parameters.	Upscale in-situ SM to validate SMAP products at various scales, with additional validation efforts against data from the SMAPVEX12 campaign and an additional network.	The proposed approach demonstrated better capability to predict mean SM than other upscaling methods in comparison against field measurements.	The distribution of in-situ sties was relatively clustered and the AOA of RF model was not assessed.	[23]
Soil Moisture / Temperature Monitoring Network (SMTMN), Tibet Autonomous Region, China	Bayesian data fusion method with MODIS ATI data.	Upscale in-situ SM to microwave remote sensing footprint.	The upscaled SM derived using data from randomly selected three stations was comparable with the mean values of SM from 20 stations, with the mean RMSE being 0.02 m ³ /m ³ .	The results lacked evaluations against independent data (e.g., from satellite missions or field campaigns).	[67]
A network with 20 soil monitoring stations, Twente region, Netherlands	Dutch National Hydrological model with scaling parameters derived from observed and simulated SM.	Develop upscaling methods to translate the spatial mean of point-scale SM to SMAP domain, hence to assess the performance of SMAP data retrieved using different algorithms.	The afternoon-acquired Single Channel Algorithm at Vertical Polarization (SCA-V) SMAP SM retrievals had best agreement with native spatial mean of in-situ SM, being an ubRMSE of 0.059 m ³ /m ³ , while higher bias was found when compared against upscaled SM.	The results indicated none of the SMAP products met the target accuracy using the Twente measurements, which may be attributed to frozen conditions and antecedent precipitation.	[72]
Marena, Oklahoma, In Situ Sensor Testbed (MOISST), Oklahoma, United States	Least Absolute Shrinkage and Selection Operator (LASSO) regression model with predictor variables.	Develop a broadly applicable upscaling approach that has transferrable capability between sites.	The proposed model only required point-scale SM and sand data, and achieved a reasonable accuracy, being an RMSE between 0.006 and 0.031 m ³ /m ³ among six sites.	The results were cross-validated while lacked evaluations against independent data (e.g., from satellite missions or field campaigns).	[22]
OzNet Hydrological Monitoring Network, Yanco, Australia	ML models and spatiotemporal fusion, with multi-source predictor variables.	Spatially predict SM at 100 m resolution and assess its AOA, with evaluation against independent data from field campaigns, external networks and SMAP data.	Spatial cross-validation demonstrated an ubRMSE (R) of 0.05 m ³ /m ³ (0.78). Independent evaluations demonstrated mean R values of 0.62-0.64 against field data, 0.84-0.91 against additional SM networks, and 0.87 against satellite retrievals, respectively.	The in-situ sites were spatially clustered, potentially limiting their representativeness of the general conditions of study area. The results were especially impacted by topographic features.	This study

to establish reliable spatial dependence structures and assume stationarity in their underlying covariance functions. When applied beyond the range defined by the variogram model or with limited data availability, kriging essentially becomes an extrapolator, producing uniform predictions that may not accurately reflect true spatial variability. Additionally, these methods demonstrate limited capacity in capturing nonlinear

relationships across heterogeneous landscapes where multiple environmental factors influence SM distributions. Table IV summarizes various representative SM upscaling studies from diverse ground networks.

Conversely, the proposed data-driven approach herein, distinguished by its AOA delineation, advances beyond these constraints. The AOA metric employs standardized predictor

variables, importance-weighted Euclidean distances, and outlier-removal thresholds to establish prediction reliability bounds. This approach captures complex, nonlinear patterns in predictor variables while providing quantitative metrics for prediction reliability—capabilities not readily available in traditional geostatistical frameworks. The effectiveness of AOA is evidenced by the alignment between AOA boundaries and validation metrics against SMAP retrievals (remarkably lower bias and ubRMSE within AOA), particularly in regions where traditional interpolation methods would be constrained by their inherent assumptions about spatial dependence structures.

The incorporation of diverse predictor variables enhanced its potential for application at a broader scale. Burton et al. [73] demonstrated the effectiveness of ML-based approaches in spatially predicting point-scale eddy covariance flux data across a continental scale. While acknowledging the inherent differences between SM and carbon flux data (e.g., greater spatiotemporal variabilities in SM), this study suggested a viable prediction approach of SM at larger geographic extent or higher temporal frequency. However, when dealing with large spatial extents, the representativeness and applicability of the relationship between in situ SM measurements and predictor variables for successful prediction remain uncertain [13]. Ensuring the robustness of these relationships across diverse landscapes and climatic conditions is critical for the success of spatial prediction efforts. Careful consideration of the selection and evaluation of predictor variables is crucial for improving the generalizability and reliability of SM prediction at broader scales [74]. Apart from data sourced from polar-orbiting platforms (i.e., MODIS herein), it is also worth considering data obtained from geostationary platforms, such as Himawari-8/9 [75] and Fengyun-4A [76]. These geostationary platforms present opportunities to monitor more detailed subdaily vegetation and phenology dynamics [77], LST variations [78], and gross primary productivity [79]. Integrating data from such platforms can help enhance the diurnal predictive capability of ML models. Nevertheless, the choice of predictor variables should undergo iterations of assessment to ensure the reliability and accuracy of the prediction results [74]. Accurately quantifying the contribution of essential predictor variables not only improves the operational efficiency of this approach but also aids in reducing computational resources, making it more practical for widespread use in larger-scale SM assessments.

The proposed approach integrated spatiotemporal fusion with ML models to spatially predict in situ SM measurements, achieving a spatial resolution of 100 m. This resolution, closely approximating the point scale, is particularly well-suited for agricultural applications, allowing for the effective capture of landscape heterogeneity. However, the adoption of spatiotemporal fusion techniques could introduce additional uncertainties. These uncertainties arise from the reliance on regression coefficients derived from coarse-resolution image pairs, leading to concerns about their applicability at finer resolutions. While internal bias correction steps for LST data have been employed to mitigate such errors [58], quantifying the temporal variations of surface reflectance data (albedo and NDVI), particularly in heterogeneous agricultural

regions, remains challenging. The inherent complexities of these regions can introduce variability that may not be fully captured or corrected by existing methods, emphasizing the need for careful consideration and potential refinement of fusion techniques in such environments.

Several critical aspects warrant attention and potential refinement in future studies. First, the reliability and representativeness of in situ SM measurements could influence the accuracy of the spatial prediction results [13]. The quality of in situ data is contingent on sensor calibration, monitoring techniques, and adherence to network standards. Meanwhile, the spatial prediction approach also relies on the spatial representativeness of the selected clusters for CV. Inadequate capturing of variability in land cover, soil types, or topography within these clusters may hinder the generalization of results to other regions. External factors, such as climate extremes, can also impact data continuity, as exemplified by the discontinuity caused by the 2019 wildfires in Australia affecting the OzNet network. Second, the transferability of this approach requires careful consideration. Achieving success in spatial prediction within one region or cluster does not guarantee similar performance in areas with distinct environmental conditions. The efficacy of this approach may vary when applied to regions characterized by different land cover types, climatic regimes, or soil characteristics. Future research should focus on evaluating the adaptability of these approaches in diverse environments, including vegetated or arid regions. Third and finally, the challenge of insufficient validation sites poses a potential limitation. In regions with a sparse distribution of in situ measurement sites or an uneven distribution among clusters, the robustness of CV results may be compromised. This issue is particularly pronounced in the southern hemisphere like Australia, where nonagricultural regions lack adequate representation, further complicating the application of this approach. Efforts to address these limitations and refine the methodology should be a priority in future research.

VI. CONCLUSION

Previous SM upscaling studies were often restricted to localized conditions and overlooked the quantification of spatial uncertainty in SM estimates. This study introduced a spatial prediction approach that integrates ML and spatiotemporal fusion, while implementing the AOA metric to assess the prediction capabilities of two ML models. Point-scale SM from 28 OzNet sites was extrapolated to a 100×100 km area at 100 m resolution with daily frequency, spanning both a CV period (2016–2019) and a test period (2020–2021). The AOA metrics of two ML models, RF and XGB, were determined based on their spatial CV performance and feature importance.

The AOA of RF and XGB covered 43.1% and 41.5% of the study area, respectively, with an expected ubRMSE of $0.05 \text{ m}^3/\text{m}^3$ and an R of 0.78. The spatial SM prediction delineated by AOA was then evaluated against multiple independent datasets (HDAS, CosmOz, OzFlux, and SMAP L2). The results indicated that RF-predicted SM demonstrated more robustness and better agreement with independent data. Specifically, during the CV period, RF-predicted SM achieved mean R values of 0.62–0.64 against HDAS SM,

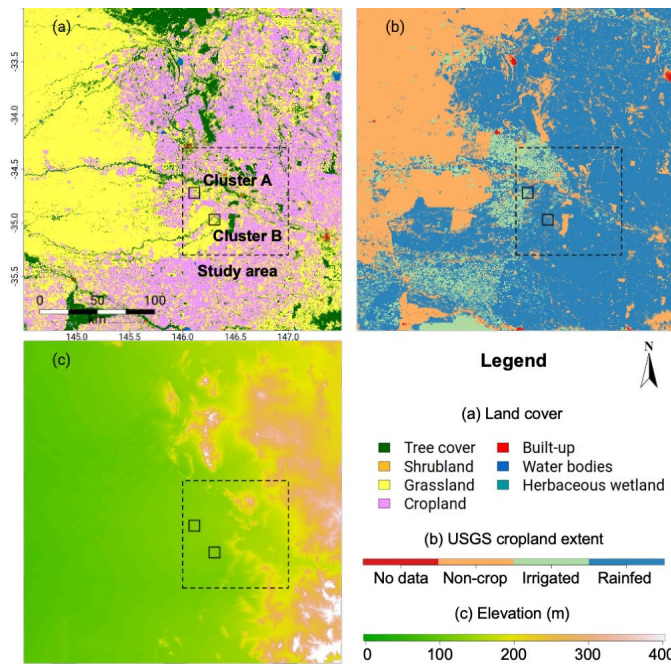


Fig. A1. (a) Land cover information of the extended area [49]. (b) USGS cropland extent [50]. (c) DEM-S [44]. The spatial extent of the extended area spans 144.50°E–147.50°E and 33.00°S–36.00°S (300 × 300 km).

0.84–0.91 against independent SM networks, and 0.87 against SMAP L2 data, respectively. Notably, spatial SM prediction within AOA showed lower uncertainties against SMAP L2 compared to areas outside AOA, which was further assessed across an extended spatial domain (Fig. A1).

In summary, this study offered a reliable approach for extrapolating point-scale SM data to match the resolution of Landsat satellite footprint, influenced by regional nuances in predictor variables. Future research directions may explore the variability of AOA by examining diverse predictor variables, and assess the transferability of the proposed approach across various environmental conditions and land cover types.

APPENDIX

ILLUSTRATION OF THE EXTENDED REGION

The extended area comprises a mosaic of six Landsat tiles (path: 092–093, row: 083–085) to evaluate westward extensibility of AOA. Fig. A1 presents (a) land cover classification information; (b) USGS cropland extent dataset differentiating between irrigated and rainfed agricultural systems; and (c) elevation gradients ranging from 0 to 400 m. The domain encompasses a physiographic transition from the agricultural core in the east (146.0°E–147.0°E) to predominantly non-agricultural regions in the west (145.0°E–146.0°E), providing a test case for assessing AOA performance across distinct landscape boundaries.

ACKNOWLEDGMENT

This research was performed as part of a Ph.D. by the first author (Yi Yu) under an academic collaboration between The Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). This research was undertaken while supported

by the ANU University Research Scholarship and an ANU-CSIRO Digital Agriculture Ph.D. Supplementary Scholarship. This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS). This research was supported by resources and expertise provided by CSIRO IMT Scientific Computing. The authors acknowledge the OzNet (<https://www.oznet.org.au>), CosmOz (<https://cosmoz.csiro.au>), and OzFlux (<https://www.ozflux.org.au>) networks for provision and processing of data. They thank the TGRS editorial team and the two anonymous reviewers for their constructively critical comments which provided the catalyst to improve an earlier version of this research.

CODE AND DATA AVAILABILITY

The analysis scripts for this research are publicly available at https://github.com/yuyi13/OzNet_AOA, with a permalink registered at Zenodo (<https://doi.org/10.5281/zenodo.15290777>). The experimental data are publicly available from the CSIRO Data Access Portal (<https://doi.org/102.100.100/705658>).

REFERENCES

- [1] S. I. Seneviratne et al., "Investigating soil moisture-climate interactions in a changing climate: A review," *Earth-Sci. Rev.*, vol. 99, nos. 3–4, pp. 125–161, May 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0012825210000139>
- [2] C. Corradini, R. Morbidelli, and F. Melone, "On the interaction between infiltration and hortonian runoff," *J. Hydrol.*, vol. 204, nos. 1–4, pp. 52–67, Jan. 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169497001005>
- [3] S. Assouline, "Infiltration into soils: Conceptual approaches and solutions," *Water Resour. Res.*, vol. 49, no. 4, pp. 1755–1772, Apr. 2013, doi: [10.1002/WRCR.20155](https://doi.org/10.1002/WRCR.20155).
- [4] G. Balsamo et al., "Satellite and in situ observations for advancing global Earth surface modelling: A review," *Remote Sens.*, vol. 10, no. 12, p. 2038, Dec. 2018. [Online]. Available: <https://www.mdpi.com/2072-4292/10/12/2038>
- [5] J. Peng et al., "A roadmap for high-resolution satellite soil moisture applications—Confronting product characteristics with user requirements," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112162. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425720305356>
- [6] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Oct. 2001.
- [7] Z. Bartalis et al., "Initial soil moisture retrievals from the METOP—A Advanced Scatterometer (ASCAT)," *Geophys. Res. Lett.*, vol. 34, no. 20, Oct. 2007, Art. no. L20401, doi: [10.1029/2007GL031088](https://doi.org/10.1029/2007GL031088).
- [8] D. Entekhabi et al., "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.
- [9] J. Dong et al., "Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation," *Remote Sens. Environ.*, vol. 242, Jun. 2020, Art. no. 111756. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425720301267>
- [10] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803>
- [11] R. H. Reichle et al., "Assessment of MERRA-2 land surface hydrology estimates," *J. Climate*, vol. 30, no. 8, pp. 2937–2960, Apr. 2017. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/30/8/jcli-d-16-0720.1.xml>
- [12] J. Zeng, Z. Li, Q. Chen, H. Bi, J. Qiu, and P. Zou, "Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations," *Remote Sens. Environ.*, vol. 163, pp. 91–110, Jun. 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425715001042>

- [13] C. Peng et al., "Spatial representativeness of soil moisture stations and its influential factors at a global scale," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4402915. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10817643?casa_token=ItAD_NGYe7IAAAAAA
- [14] S. Zacharias et al., "A network of terrestrial environmental observatories in Germany," *Vadose Zone J.*, vol. 10, no. 3, pp. 955–973, Aug. 2011, doi: [10.2136/vzj2010.0139](https://doi.org/10.2136/vzj2010.0139).
- [15] J. G. Evans et al., "Soil water content in Southern England derived from a cosmic-ray soil moisture observing system—COSMOS-U.K.," *Hydrological Processes*, vol. 30, no. 26, pp. 4987–4999, Dec. 2016. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10929>
- [16] A. Hawdon, D. McJannet, and J. Wallace, "Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia," *Water Resour. Res.*, vol. 50, no. 6, pp. 5029–5043, Jun. 2014, doi: [10.1002/2013wr015138](https://doi.org/10.1002/2013wr015138).
- [17] A. B. Smith et al., "The Murrumbidgee soil moisture monitoring network data set," *Water Resour. Res.*, vol. 48, no. 7, Jul. 2012, Art. no. W07701. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012WR011976>
- [18] L. Brocca, R. Morbidelli, F. Melone, and T. Moramarco, "Soil moisture spatial variability in experimental areas of central Italy," *J. Hydrol.*, vol. 333, nos. 2–4, pp. 356–373, Feb. 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169406004781>
- [19] W. T. Crow et al., "Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products," *Rev. Geophys.*, vol. 50, no. 2, Jun. 2012, Art. no. RG2002. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011RG000372>
- [20] J. Qin, K. Yang, N. Lu, Y. Chen, L. Zhao, and M. Han, "Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia," *Remote Sens. Environ.*, vol. 138, pp. 1–9, Nov. 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425713002150>
- [21] J. Wang, Y. Ge, G. Heuvelink, and C. Zhou, "Upscaling in situ soil moisture observations to pixel averages with spatio-temporal geostatistics," *Remote Sens.*, vol. 7, no. 9, pp. 11372–11388, Sep. 2015.
- [22] W. G. Brown, M. H. Cosh, J. Dong, and T. E. Ochsner, "Upscaling soil moisture from point scale to field scale: Toward a general model," *Vadose Zone J.*, vol. 22, no. 2, p. 20244, Mar. 2023. [Online]. Available: <https://access.onlinelibrary.wiley.com/doi/abs/10.1002/vzj2.20244>
- [23] D. Clewley et al., "A method for upscaling in situ soil moisture measurements to satellite footprint scale using random forests," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2663–2673, Jun. 2017.
- [24] Y. Xu, C. Liu, L. Wang, and L. Zou, "Exploring the spatial autocorrelation in soil moisture networks: Analysis of the bias from upscaling the Texas soil observation network (TxSON)," *Water*, vol. 15, no. 1, p. 87, Dec. 2022. [Online]. Available: <https://www.mdpi.com/2073-4441/15/1/87>
- [25] H. Meyer and E. Pebesma, "Predicting into unknown space? Estimating the area of applicability of spatial prediction models," *Methods Ecol. Evol.*, vol. 12, no. 9, pp. 1620–1633, Sep. 2021. [Online]. Available: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13650>
- [26] H. Meyer and E. Pebesma, "Machine learning-based global maps of ecological variables and the challenge of assessing them," *Nature Commun.*, vol. 13, no. 1, p. 2208, Apr. 2022, doi: [10.1038/s41467-022-29838-9](https://doi.org/10.1038/s41467-022-29838-9).
- [27] T. Hengl et al., "African soil properties and nutrients mapped at 30m spatial resolution using two-scale ensemble machine learning," *Sci. Rep.*, vol. 11, no. 1, p. 6130, Mar. 2021, doi: [10.1038/s41598-021-85639-y](https://doi.org/10.1038/s41598-021-85639-y).
- [28] D. Žižala et al., "High-resolution agriculture soil property maps from digital soil mapping methods, Czech republic," *CATENA*, vol. 212, May 2022, Art. no. 106024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0341816222000108>
- [29] T. Broeg, M. Blaschek, S. Seitz, R. Taghizadeh-Mehrjardi, S. Zepp, and T. Scholten, "Transferability of covariates to predict soil organic carbon in cropland soils," *Remote Sens.*, vol. 15, no. 4, p. 876, Feb. 2023.
- [30] L.-M. L. Valdes, M. Katurji, and H. Meyer, "A machine learning based downscaling approach to produce high spatio-temporal resolution Land Surface Temperature of the Antarctic dry valleys from MODIS data," *Remote Sens.*, vol. 13, no. 22, p. 4673, Nov. 2021.
- [31] T. Lendzioch, J. Langhammer, L. Vlček, and R. Minařík, "Mapping the groundwater level and soil moisture of a montane peat bog using UAV monitoring and machine learning," *Remote Sens.*, vol. 13, no. 5, p. 907, Feb. 2021.
- [32] J. Peng, A. Loew, O. Merlin, and N. E. C. Verhoest, "A review of spatial downscaling of satellite remotely sensed soil moisture," *Rev. Geophys.*, vol. 55, no. 2, pp. 341–366, 2017.
- [33] A. S. Abowarda et al., "Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale," *Remote Sens. Environ.*, vol. 255, Mar. 2021, Art. no. 112301.
- [34] M. Stenson, R. Searle, B. P. Malone, A. Sommer, L. J. Renzullo, and H. Di. (2021). *Australia Wide Daily Volumetric Soil Moisture Estimates. Version 1.0 [Dataset]*, doi: [10.25901/b020-nm39](https://doi.org/10.25901/b020-nm39).
- [35] Y. Yu, B. P. Malone, and L. J. Renzullo, "Empirical upscaling of point-scale soil moisture measurements for spatial evaluation of model simulations and satellite retrievals," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Athens, Greece, Jul. 2024, pp. 11496–11501. [Online]. Available: <https://ieeexplore.ieee.org/document/10642763>
- [36] E. Ghafari, J. P. Walker, L. Zhu, A. Colliander, and A. Faridhosseini, "Spatial downscaling of SMAP radiometer soil moisture using radar data: Application of machine learning to the SMAPEX and SMAPVEX campaigns," *Sci. Remote Sens.*, vol. 9, Jun. 2024, Art. no. 100122. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666017224000063>
- [37] C. Peng et al., "Global spatiotemporal trend of satellite-based soil moisture and its influencing factors in the early 21st century," *Remote Sens. Environ.*, vol. 291, Jun. 2023, Art. no. 113569. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425723001207>
- [38] J. Beringer et al., "An introduction to the Australian and New Zealand flux tower network – OzFlux," *Biogeosciences*, vol. 13, no. 21, pp. 5895–5916, Oct. 2016.
- [39] C. Schaaf and Z. Wang, "MCD43A4 MODIS/terra+aqua BRDF/Albedo Nadir BRDF adjusted ref daily L3 Global—500m V006 [Data set]," NASA EOSDIS Land Processes Distrib. Active Archive Center, 2015. Accessed: May 6, 2025. [Online]. Available: <https://doi.org/10.5067/MODIS/MCD43A4.006>
- [40] Z. Wan, "New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product," *Remote Sens. Environ.*, vol. 140, pp. 36–45, Jan. 2014.
- [41] F. Li et al., "An evaluation of the use of atmospheric and BRDF correction to standardize Landsat data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 3, pp. 257–270, Sep. 2010.
- [42] J. P. Guerschman, T. R. McVicar, J. Vleeshower, T. G. Van Niel, J. L. Peña-Arancibia, and Y. Chen, "Estimating actual evapotranspiration at field-to-continent scales by calibrating the CMRSET algorithm with MODIS, VIIRS, Landsat and Sentinel-2 data," *J. Hydrol.*, vol. 605, Feb. 2022, Art. no. 127318.
- [43] M. Hutchinson, T. Xu, J. Kesteven, I. Marang, and B. Evans. (2021). *ANUClimate 2.0*. [Online]. Available: https://pid.nci.org.au/doi/f2576_7854_4065_1457
- [44] Geosci. Aust., Canberra, ACT, Australia. (2015). *Digital Elevation Model (DEM) of Australia Derived From LiDAR 5 Metre Grid*. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/AU_GA_DEM_1SEC_v10_DEM-S
- [45] B. Malone and R. Searle, "Updating the Australian digital soil texture mapping (Part 2): Spatial modelling of merged field and lab measurements," *Soil Res.*, vol. 59, no. 5, pp. 435–451, 2021. [Online]. Available: <https://www.publish.csiro.au/paper/SR20284>
- [46] R. D. Searle, P. D. S. N. Somarathna, and B. P. Malone, "Soil and Landscape grid national soil attribute maps—Available volumetric water capacity (percent) (3 arc second resolution) version 2," CSIRO. Data Collection, 2022. [Online]. Available: <https://doi.org/10.25919/4jwj-na34>
- [47] X. Wu et al., *The P-band Radiometer Inferred Soil Moisture Experiment 2019 WORKPLAN*. Melbourne, VIC, Australia: Monash University, 2019. [Online]. Available: <https://prism.monash.edu/PRISM-19.html>
- [48] X. Wu et al., "The P-band radiometer inferred soil moisture experiment 2021 WORKPLAN," Melbourne, VIC, Australia: Monash University, 2021. [Online]. Available: <https://prism.monash.edu/PRISM-21.html>
- [49] D. Zanaga et al. (2022). *ESA WorldCover 10 M 2021 V200*. [Online]. Available: <https://zenodo.org/record/7254221>
- [50] P. Teluguntla et al., "A 30m Landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth engine cloud computing platform," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 325–340, Oct. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271618302090>
- [51] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. van Dijk, "Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sens. Environ.*, vol. 133, pp. 193–209, Jun. 2013.

- [52] C. O. Justice et al., "The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1228–1249, Jul. 1998.
- [53] S. Liang, "Narrowband to broadband conversions of land surface albedo I: Algorithms," *Remote Sens. Environ.*, vol. 76, no. 2, pp. 213–238, May 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425700002054>
- [54] J. C. Jiménez-Muñoz, J. A. Sobrino, D. Skokovic, C. Mattar, and J. Cristóbal, "Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1840–1843, Oct. 2014.
- [55] M. F. Hutchinson, "Interpolating mean rainfall using thin plate smoothing splines," *Int. J. Geographical Inf. Syst.*, vol. 9, no. 4, pp. 385–403, Jul. 1995, doi: [10.1080/02693799508902045](https://doi.org/10.1080/02693799508902045).
- [56] O. Merlin, J. P. Walker, R. Panciera, R. Young, J. D. Kalma, and E. Kim, "Soil moisture measurement in heterogeneous terrain," in *Proc. MODSIM*, Christchurch, New Zealand, Dec. 2007, pp. 2604–2610. [Online]. Available: https://mssanz.org.au/MODSIM07/papers/46_s60/SoilMoistures60_Merlin.pdf
- [57] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, Nov. 2010.
- [58] Y. Yu, L. J. Renzullo, T. R. McVicar, B. P. Malone, and S. Tian, "Generating daily 100m resolution Land Surface Temperature estimates continentally using an unbiased spatiotemporal fusion approach," *Remote Sens. Environ.*, vol. 297, Nov. 2023, Art. no. 113784. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425723003358>
- [59] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1674987115000821>
- [60] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [61] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001. [Online]. Available: <http://www.jstor.org/stable/2699986>
- [62] H. Ma et al., "Surface soil moisture from combined active and passive microwave observations: Integrating ASCAT and SMAP observations based on machine learning approaches," *Remote Sens. Environ.*, vol. 308, Jul. 2024, Art. no. 114197. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425724002153>
- [63] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 4768–4777.
- [64] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9).
- [65] H. Meyer, C. Reudenbach, S. Wöllauer, and T. Nauss, "Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction," *Ecol. Model.*, vol. 411, Nov. 2019, Art. no. 108815. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304380019303230>
- [66] M. Kuhn, "Building predictive models in R using the caret package," *J. Stat. Softw.*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: <https://www.jstatsoft.org/index.php/jss/article/view/v028i05>
- [67] S. Gao, Z. Zhu, H. Weng, and J. Zhang, "Upscaling of sparse in situ soil moisture observations by integrating auxiliary information from remote sensing," *Int. J. Remote Sens.*, vol. 38, no. 17, pp. 4782–4803, Sep. 2017, doi: [10.1080/01431161.2017.1320444](https://doi.org/10.1080/01431161.2017.1320444).
- [68] S. Siebert, V. Henrich, K. Frenken, and J. Burke, "Update of the digital global map of irrigation areas to version 5," Food Agricult. Organization United Nations, Rome, Italy, Tech. Rep., 2013.
- [69] A. H. Thiessen, "Precipitation averages for large areas," *Monthly Weather Rev.*, vol. 39, no. 7, pp. 1082–1089, Jul. 1911. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/39/7/1520-0493_1911_39_1082b_pafila_2_0_co_2.xml
- [70] P. M. Bartier and C. P. Keller, "Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW)," *Comput. Geosci.*, vol. 22, no. 7, pp. 795–799, Aug. 1996. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0098300496000210>
- [71] R. Webster and M. A. Oliver, *Geostatistics for Environmental Scientists*. Hoboken, NJ, USA: Wiley, 2007. [Online]. Available: <https://books.google.com.au/books?id=WBwSvYvNY8C>
- [72] R. van der Velde et al., "Validation of SMAP L2 passive-only soil moisture products using upscaled in situ measurements collected in Twente, The Netherlands," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 1, pp. 473–495, Jan. 2021. [Online]. Available: <https://hess.copernicus.org/articles/25/473/2021/>
- [73] C. A. Burton, L. J. Renzullo, S. W. Rifai, and A. I. J. M. Van Dijk, "Empirical upscaling of OzFlux eddy covariance for high-resolution monitoring of terrestrial carbon uptake in Australia," *Biogeosciences*, vol. 20, no. 19, pp. 4109–4134, Oct. 2023. [Online]. Available: <https://bg.copernicus.org/articles/20/4109/2023/>
- [74] Y. Yu, L. J. Renzullo, and S. Tian, "Continental scale downscaling of AWRA-L analysed soil moisture using random forest regression," in *Proc. 24th Int. Congr. Model. Simul. (MODSIM2021)*, Sydney, NSW, Australia, 2021, pp. 498–504.
- [75] K. Bessho et al., "An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites," *J. Meteorological Soc. Jpn. Ser. II*, vol. 94, no. 2, pp. 151–183, 2016.
- [76] J. Yang, Z. Zhang, C. Wei, F. Lu, and Q. Guo, "Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 8, pp. 1637–1658, Aug. 2017.
- [77] T. Miura, S. Nagai, M. Takeuchi, K. Ichii, and H. Yoshioka, "Improved characterisation of vegetation and land surface seasonal dynamics in central Japan with Himawari-8 hypertemporal data," *Sci. Rep.*, vol. 9, no. 1, p. 15692, Oct. 2019.
- [78] Y. Yu et al., "Solar zenith angle-based calibration of Himawari-8 Land Surface Temperature for correcting diurnal retrieval error characteristics," *Remote Sens. Environ.*, vol. 308, Jul. 2024, Art. no. 114176. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425724001949>
- [79] A. M. Khan et al., "The diurnal dynamics of gross primary productivity using observations from the advanced baseline imager on the geostationary operational Environmental Satellite-R series at an oak savanna ecosystem," *J. Geophys. Res., Biogeosci.*, vol. 127, no. 3, p. 2021, Mar. 2022, doi: [10.1029/2021jg006701](https://doi.org/10.1029/2021jg006701).



Yi Yu (Member, IEEE) received the B.Mgmt. degree in land resources management from Southwest University, Chongqing, China, in 2018, and the M.Sc. degree in environment (advanced) and the Ph.D. degree in hydrology and remote sensing from The Australian National University (ANU), Canberra, ACT, Australia, in 2020 and 2024, respectively. His Ph.D. research investigated the role of remotely sensed land surface temperature in improving spatiotemporal monitoring capability of agricultural drought.

He is currently a Post-Doctoral Researcher at The University of Sydney, Sydney, NSW, Australia. His research interests focus on developing hybrid approaches that integrate physical and data-driven methods to better understand land-atmosphere interactions across various spatiotemporal scales, particularly in the context of climate extremes such as drought.

Dr. Yu received the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food Director's Award—Next Generation Science in 2023. He was among the Student Paper Competition Finalists at the 2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).



Brendan P. Malone received the B.Sc. degree (Hons.) in agriculture and the Ph.D. degree in digital soil mapping from The University of Sydney, Sydney, NSW, Australia, in 2008 and 2012, respectively. His Ph.D. thesis introduced new methods for rescuing legacy soil data to create thematic, highly granular, and comprehensive soil mapping products relevant across multiple spatial scales.



Luigi J. Renzullo received the B.Sc. degree in mathematics and computing, the PostGrad.Dip. degree in applied physics, and the Ph.D. degree in remote sensing science from Curtin University of Technology, Perth, WA, Australia, in 1996, 1997, and 2004, respectively.

He was a Senior Research Scientist with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, ACT, Australia, for 15 years; and The Australian National University (ANU), Canberra, for five years. He is currently a

Researcher at Bureau of Meteorology, Parkes, ACT, Australia; and an Honorary Senior Lecturer at ANU. He has over 25 years' experience in satellite remote sensing with research spanning applications of visible, thermal, and microwave sensing systems to environmental modeling problems.

Dr. Renzullo is a member of Australian Mathematical Society and Australian Institute of Physics.



Chad A. Burton received the B.Sc. degree (Hons.) in geosciences from Monash University, Clayton, VIC, Australia, in 2016, and the M.Sc. degree in environmental change and management from the University of Oxford, Oxford, U.K., in 2018. He is currently pursuing the Ph.D. degree with the Fenner School of Environment and Society, The Australian National University (ANU), Canberra, ACT, Australia.

He is currently researching data-driven methods to enhance the understanding of Australia's terrestrial carbon cycle and ecosystem dynamics at ANU. He is also an Earth Observation Scientist at Geoscience Australia, Canberra, where he develops Earth observation-based geospatial products for monitoring environmental change.



Siyuan Tian received the master's degree in engineering (geomatics) from The University of Melbourne, Carlton, VIC, Australia, in 2013, and the Ph.D. degree in Earth science from The Australian National University, Canberra, ACT, Australia, in 2019. Since her Ph.D., research has been focused on using the Earth observations together with land surface models to understand the interaction between the vegetation and hydrological cycle.

She is currently a Research Scientist at the Bureau of Meteorology, Parkes, ACT, Australia.

Dr. Tian received the Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ) Early Career Research Excellence Award in 2023.

Ross D. Searle received the B.Sc. degree in agricultural science, majoring in soil science from The University of Queensland, Brisbane, QLD, Australia, in 1990.

He is currently at the Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT, Australia, as a Senior Experimental Scientist.

Mr. Searle is a member of the Soil Science Australia.



Thomas F. A. Bishop received the B.Sc. degree (Hons.) in agriculture and the Ph.D. degree in precision agriculture from The University of Sydney (USYD), Sydney, NSW, Australia, in 1997 and 2002, respectively.

His teaching is related to applied statistics, environmental science, and GIS. He is currently the Academic Director of Sydney Informatics Hub, Newtown, NSW, Australia, which is the Core Research Facility, USYD. He is a Professor with the School of Life and Environmental Science, USYD.

Prior to starting work at USYD in 2007, he held post-doctoral positions with the University of Florida, Gainesville, FL, USA; Rothamsted Research, Harpenden, U.K.; and the University of New South Wales, Sydney, NSW, Australia. His research interests include modeling and predicting the variation of environmental properties in space and time with an emphasis on applying this to the domains of soil, agriculture, and hydrology.

Prof. Bishop is an Associate Editor of *European Journal of Soil Science* and *Soil Research* and on the Editorial Board of *Geoderma* and *Pedosphere*.



Jeffrey P. Walker (Fellow, IEEE) received the B.E. (civil) and B.Surveying degrees (Hons.) from the University of Newcastle, Callaghan, NSW, Australia, in 1995, and the Ph.D. degree in water resources engineering from the University of Newcastle, in 1999. His Ph.D. thesis was among the early pioneering research on estimation of root-zone soil moisture from assimilation of remotely sensed surface soil moisture observations.

He then joined Goddard Space Flight Center, NASA, Greenbelt, MD, USA, to implement his soil moisture work globally. In 2001, he moved to the Department of Civil and Environmental Engineering, University of Melbourne, Parkville, VIC, Australia, as Lecturer, where he continued his soil moisture work, including development of the only Australian airborne capability for simulating new satellite missions for soil moisture. In 2010, he was appointed as a Professor at the Department of Civil Engineering and Environmental Engineering, Monash University, Clayton, VIC, Australia, where he has continued this research. He is contributing to soil moisture satellite missions at NASA, ESA, and JAXA, as a Science Team Member for the Soil Moisture Active Passive (SMAP) mission and Cal/val Team member for the Soil Moisture and Ocean Salinity (SMOS) and Global Change Observation Mission—Water (GCOM-W), respectively.

Dr. Walker is a Laureate Fellow of Australian Research Council.