Spatial downscaling of SMAP radiometer soil moisture using radar data: Application of machine learning to the SMAPEx and SMAPVEX campaigns

Elaheh Ghafari, Jeffrey P. Walker, Liujun Zhu, Andreas Colliander, Alireza Faridhosseini

PII: S2666-0172(24)00006-3

DOI: https://doi.org/10.1016/j.srs.2024.100122

Reference: SRS 100122

To appear in: *Science of Remote Sensing*

Received Date: 5 August 2023

Revised Date: 6 February 2024

Accepted Date: 6 February 2024

Spatial downscaling of SMAP radiometer soil moisture using radar data: Application of machine learning to the SMAPEx and SMAPVEX campaigns

Authors: Elaheh Ghafari[a*], Jeffrey P. Walker[b], Liujun Zhu[b,c], Andreas Colliander[d], Alireza Faridhosseini[a]

[a] *Department of Water Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*

[b] *Department of Civil Engineering, Monash University, Melbourne, Australia*

[c] *Yangtze Institute for Conservation and Development, Hohai University, Nanjing, 210024, China*

[d] *Jet Propulsion Laboratory, NASA, California Institute of Technology, Pasadena, CA 91109, USA*

Submitted to: Science of Remote Sensing

Submitted Date: 5 August 2023

*Corresponding Author: Elaheh Ghafari (Email: ghafarielaheh84@gmail.com)

Email addresses: ghafarielaheh84@gamil.com (E. Ghafari), Jeff.Walker@monash.edu (J.P. Walker), Liujun.zhu@hhu.edu.cn (L. Zhu), andreas.colliander@jpl.nasa.gov (A. Colliander), farid-h@um.ac.ir (A. Faridhosseini)

ABSTRACT:

This study developed a random forest approach for downscaling the coarse-resolution (36 km) soil moisture measured by The National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) mission to 1 km spatial resolution, utilizing airborne remotely sensed data (radar backscatter and radiometer retrieved soil moisture), vegetation characteristics (normalized difference vegetation index), soil properties, topography, and ground soil moisture measurements from before the launch of SMAP for training a random forest model. The 36 km SMAP soil moisture product was then downscaled by the trained model to 1 km resolution using the information from SMAP. The downscaled soil moisture was evaluated using airborne retrieved soil moisture observations and ground soil moisture measurements. Considering the airborne retrieved soil moisture as a reference, the results demonstrated that the proposed random forest model could downscale the SMAP radiometer product to 1 km resolution with a correlation coefficient of 0.97, unbiased Root Mean Square Error of 0.048 $m^3.m^{-3}$ and bias of 0.016 $m^3.m^{-3}$. Accordingly, the downscaled soil moisture captured the spatial and temporal heterogeneity and demonstrated the potential of the proposed machine learning model for soil moisture downscaling.

Keywords: Machine learning, Downscaling, Soil moisture, SMAP, Random forest model, SMAPEx, SMAPVEX

Highlights:

- A random forest machine learning model was developed to downscale the SMAP radiometer soil moisture.
- The random forest model was trained using active-passive microwave, landscape, and vegetation data.
- Airborne pre- / post-launch data was used to train and subsequently validate the downscaling model.
- Training to soil moisture data over 1 km grid cells decreased the training scale-mismatch.
- Assessment of variables showed the importance of horizontal polarised backscatter and terrain slope.

1. Introduction

Soil moisture is an important variable in the hydrology, climatology, and agricultural sciences, as it is an essential factor in controlling the global water, energy and carbon cycles, linking land and atmospheric parameters (Seneviratne et al. 2010). Over the last decade, the possibility of global soil moisture monitoring has been made possible by the advent of remote sensing techniques (Entekhabi et al. 2010; Kerr et al. 2012). Accordingly, L-band passive microwave at 1.41 GHz frequency has been adopted as the preferred approach due to its ability to monitor data under all weather conditions, the direct relationship between passive microwave observation and soil moisture, and the low sensitivity to atmospheric effects, surface roughness and vegetation (Gao et al. 2022; Schmugge et al. 1986). Therefore, L-band satellites such as Soil Moisture and Ocean and Salinity (SMOS) mission were launched to provide global soil moisture maps (Barre et al. 2008). However, the low spatial resolution of passive microwave sensors is a major limitation to many applications. Consequently, investigations demonstrated that combining active (radar) and passive (radiometer) microwave observations can enhance the resolution by combing their respective advantages, including the high accuracy of passive observations with the fine spatial resolution of active observations (Das et al. 2011; Entekhabi et al. 2010). This method has been termed as active passive.

On the 31st January 2015, the Soil Moisture Active Passive (SMAP) satellite was launched by the National Aeronautics and Space Administration (NASA), to provide global soil moisture maps of the top 5 cm soil surface with a temporal resolution of 2 to 3 days and spatial resolution of 9 km (Entekhabi et al. 2014). This was to be achieved by combining 1.26 GHz radar backscatter (σ) at 3 km resolution and 1.41 GHz radiometer brightness temperature (Tb) at 36 km resolution, with the aim to provide a soil moisture accuracy better than 0.04 $m^3.m^{-3}$ (Chan et al. 2016). However, the SMAP radar instrument stopped working in July 2015, leaving only the radiometer observations measured by SMAP. Consequently, investigations have focused on generating a high resolution soil moisture product by combining the SMAP radiometer with other radar observations, such as those from the Copernicus Sentinel-1 C-band radar (Das et al. 2019; Ghafari et al. 2020). Moreover, the data that was collected during the period the radar was working has provided an important experimental data set for developing and testing a variety of downscaled SMAP products using a range of data and algorithms (Colliander et al. 2017a; Sabaghy et al. 2018; Wu et al. 2016; Wu et al. 2015).

In recent years, several alternate methods have emerged for downscaling the coarse resolution SMAP and SMOS soil moisture products (Das et al. 2011; Kim and Zyl 2009; Merlin et al. 2012; Narayan et al. 2006; Piles et al. 2011). Among these approaches are machine learning methods, whereby optical and thermal observations, along with static geomorphological data at high spatial resolution are usually used as the covariates to downscale the passive microwave soil moisture product (Fang and Shen 2020; Karthikeyan and Mishra 2021; Long et al. 2019). However, investigations on utilizing radar observations as a covariate for machine learning methods has been limited (Mao et al. 2019; Zhu et al. 2021). Several investigations have shown that among all the machine learning methods used for downscaling satellite-based products, being either the derived soil moisture or the observed brightness temperature, the random forest algorithm has shown the greater performance, as it is a more flexible model due to randomization and use of an ensemble approach (Abbaszadeh et al. 2019; Hu et al. 2020; Lei et al. 2022; Mao et al. 2022; Rao et al. 2022; Zhao et al. 2018).

To ensure a robust satellite downscaling algorithm, this study used completely independent pre- and post-launch information for the training and testing phases of the machine learning model development, respectively. Moreover, a random forest model was developed, based on vegetation characteristics, topography, properties of the top 5 cm soil layer, and the soil moisture datasets available at only focus monitoring sites, for downscaling the coarse resolution SMAP passive soil moisture (36 km) to fine spatial resolution (1 km). This was achieved utilizing the third Soil Moisture Active Passive Experiment (SMAPEx-3) and Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX-12) campaigns. Previous studies commonly used the 36 km SMAP grid cell soil moisture as the 1 km soil moisture input variable to construct the downscaling model (Abbaszadeh et al. 2019; Hu et al. 2020; Rao et al. 2022). Consequently, one of the novelties of this paper is utilizing soil moisture at the downscaling target resolution of 1 km as input to the training phase of the machine learning, as provided by pre-launch campaigns, instead of the coarse passive SMAP soil moisture. Furthermore, most machine learning approaches to date have validated the output at just a few in situ points (Abowarda et al. 2021; Lei et al. 2022; Long et al. 2019). However, this study used the microwave soil moisture data retrieved from airborne passive observations across several SMAP pixels at 1 km resolution for validation, along with all available ground soil moisture measurements, to ensure the accuracy of the achieved spatial patterns in soil moisture.

2. Study area

Two field experiment sites were selected as the study areas due to their large-scale airborne and ground campaigns; the Soil Moisture Active Passive Experiments (SMAPEx) field campaigns carried out in south-eastern Australia, and the Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX-12) field campaign conducted in south central Manitoba, Canada. The extensive pre-launch data make these very suitable study areas for the purpose of this research. Combining the data from both campaigns provided a sufficiently large sample size for training the algorithm. Furthermore, these sites present complementary soil characteristics, weather status and vegetation coverage, thus providing a wide range of conditions. More detailed descriptions about the field campaigns follow.

2.1. Soil Moisture Active Passive Experiment (SMAPEx) campaigns

Five airborne field campaigns were undertaken over the period from 2010 to 2015 in south-eastern Australia, known as the Soil Moisture Active Passive Experiments (SMAPEx) (Panciera et al. 2014; Ye et al. 2020). These were conducted in the Yanco SMAP validation area in the Murrumbidgee River catchment (Fig. 1). SMAPEx-1 to SMAPEx-3 were undertaken before the SMAP launch, while SMAPEx-4 and SMAPEx-5 were conducted post-launch. These campaigns were designed with the basic target of developing the soil moisture algorithms for SMAP products at pre-launch, and for calibration and validation of SMAP observations and downscaled soil moisture at post-launch. Accordingly, during the SMAPEx campaigns, airborne passive and active observations were made similar to the SMAP observations (Wu et al. 2015), and the ground soil moisture and several kinds of ancillary data were collected coincident with SMAP overpasses. The third to fifth SMAPEx campaigns, which were utilized in this research for developing and then testing the machine learning downscaling model, were conducted in the austral spring (5th to 23rd September, 2011), autumn (30th April to 23rd May, 2015), and spring (6th to 28th September, 2015), respectively. These campaigns provided valuable datasets for developing the SMAP downscaling algorithm under Australian soil and vegetation conditions (Panciera et al. 2014). More details about the SMAPEx datasets are in the workplan reports available at https://www.smapex.monash.edu, so only a brief outline of the information is presented here.

The dataset from the SMAPEx-3 campaign included six focus areas, being a 3 km × 3 km grid cell for each, corresponding to the EASE-2 SMAP grid cells across the SMAP radiometer pixel. These were used for constructing the downscaling model during the training phase of establishing the machine learning algorithm (Fig. 1). It is notable that only data from the third SMAPEx campaign was used at this step. The datasets during the SMAPEx-4 and SMAPEx-5 experiments, covering approximately six coarse resolution SMAP grid cells over the SMAP validation flight area (Fig. 1), were utilized for validating the algorithm. The variability in soil and vegetation conditions, the availability of the soil moisture dataset measured based on the ground experiments, and the availability of the required airborne and satellite data make these selected areas appropriate for research on microwave retrieval of soil moisture from satellites. The selected study site is located in a semi-arid area with flat topography. The six selected ground-sampling sites are called YA4, YA7, YB5, YB7, YE and YF. The land use of the sites is irrigated cropping (90%) and grazing (10%) for YA4 and YA7, irrigated cropping (85%) and grazing (15%) for YF, and entirely grazing for YB5, YB7 and YE. Therefore, the two main land cover types were cropping and grazing. The soil textures are categorized as clay loam for YA4 and YA7, silty clay loam for YE, and loam for YB5, YB7 and YF. The soil texture was obtained from gravimetric samples used to extract the soil particle distribution (Monerris et al. 2011) and the CSIRO Digital Atlas of Australian Soils (1991).

The SMAPEx-3 campaign took place in the austral spring, with moderate rainfall in the first half of the period resulting in a soil moisture dry down, and winter crops in their intensive growth periods. More descriptions of SMAPEx-3 are available in Panciera et al. (2014). The SMAPEx-4 campaign took place in the austral autumn. During this experiment, crop areas with dry or burned corn stubble or rice straw residual from harvest were dominant, while some crop areas had been ploughed for seeding. Consequently, the surface roughness was high due to the deep furrows in the ploughed and harvested areas, while the grazing area was covered by short grass. The range in soil moisture conditions was around 0.1 $m^3.m^{-3}$ and the average vegetation water content was approximately 0.1 $kg.m^{-2}$. Before the campaign began, several heavy rainfall events occurred which made for heterogenous soil conditions during the dry down period in the selected area. Two medium rainfall events also occurred during the campaign, providing further heterogeneity to the soil water content distributions (Ye et al. 2020). The last campaign, SMAPEx-5, took place in the austral spring when the vegetation had high growth rates, with VWC up to approximately 2 $kg.m^{-2}$. Heavy rainfall occurred before the campaign providing

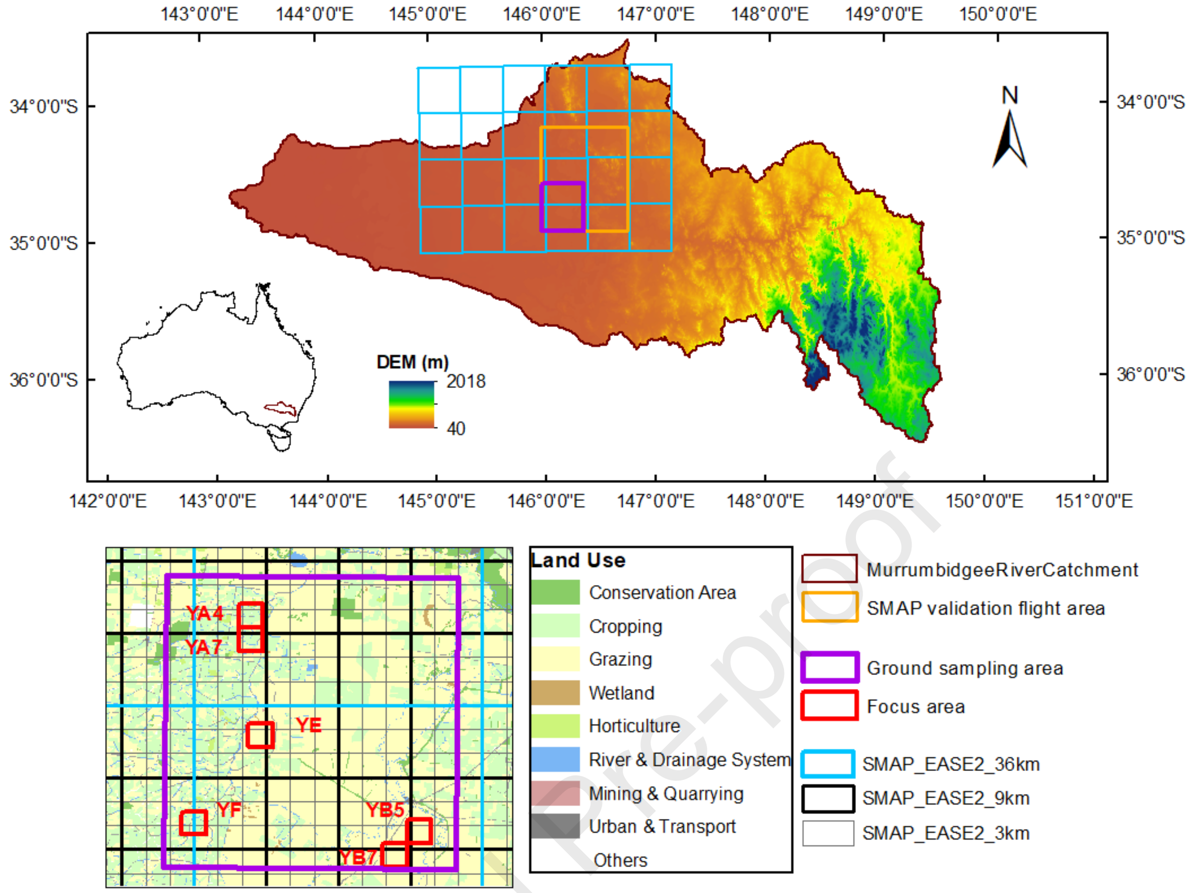

Fig. 1. The SMAPEx study site in the Murrumbidgee River catchment in south-eastern of Australia with the Digital Elevation Model (DEM), and the six focus areas used for ground sampling, together with the SMAP grid cells overlain with the land use map.

heterogeneity in soil moisture conditions along with a dry down situation. The most vegetated area during this campaign was the irrigated and dryland cropping, followed by grazing land (Ye et al. 2020).

2.2. The Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX-12) campaign

The SMAPVEX-12 field campaign was conducted at the pre-launch stage of SMAP to assist SMAP algorithm development. The campaign was conducted at the Canadian Red River Watershed in south central Manitoba, Canada (Fig. 2), mostly covered by agricultural and some forest areas (McNairn et al. 2015). The period of the SMAPVEX-12 experiment was from June 17th to July 19th, 2012, with the intent of collecting active and passive airborne observations together with ground soil moisture measurements and ancillary datasets. The size of the site was 12.8 km × 70 km, capturing forest and agricultural areas (Fig. 2). The soil texture varied from heavy clays to fine loamy sand through the east to west of the study area, leading to substantial soil moisture gradients over short distances. The site is predominately flat with a maximum slope of 2%. Ground soil moisture data were acquired by permanent soil moisture stations installed by Agriculture and Agri-food Canada, manual sampling teams, and temporary sites installed by the United States Department of Agriculture (USDA).

As shown in Fig. 2, the selected site was dominated by a mix of agricultural area, mostly including cereals and oil seeds. Overall, 67% of the site was covered by crops and approximately 15% by grassland and pasture. Seeding was undertaken in April/May and harvesting in August/September. Fifty-five agricultural fields of at least 800 m × 800 m in size were monitored throughout the SMAPVEX-12 campaign, collecting ground soil moisture measurements as shown in Fig. 2. As both cropland and grassland data were available, the SMAPVEX-12 campaign provided useful information to complement the SMAPEx campaign dataset for downscaling the SMAP soil moisture utilizing the machine learning algorithm. Further details about the campaign are available in McNairn et al. (2015), with the SMAPVEX-12 datasets accessible at https://nsidc.org/data/smap/validation-data.

3. Data

3.1. SMAP radiometer soil moisture product

The SMAP satellite provides global scale soil moisture maps of the top 5 cm, with an *ubRMSE* of less than 0.04 $m^3.m^{-3}$ (Bindlish et al. 2016). This research utilized a machine learning approach for downscaling the SMAP radiometer-based soil moisture product. The

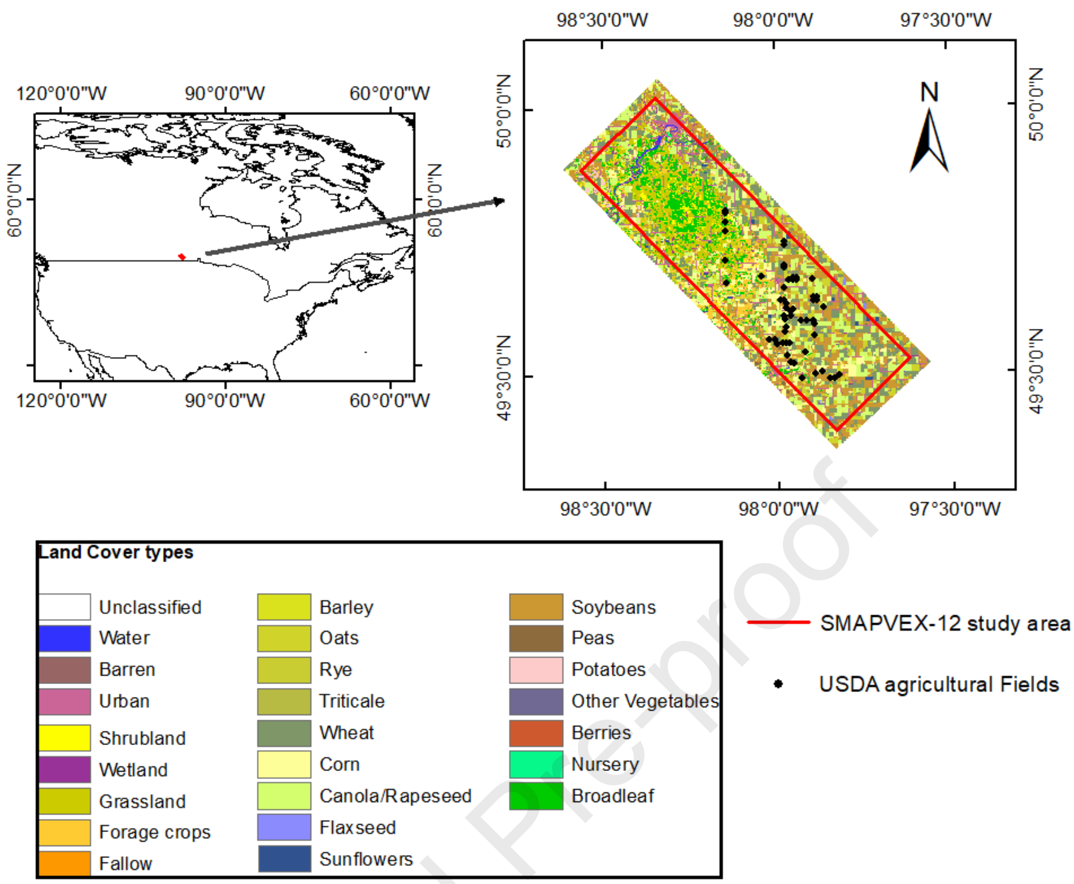

Fig. 2. Overview of the SMAPVEX-12 study site located at the Red River watershed in south-central Manitoba in Canada overlain with the land cover types and the location of USDA agricultural fields.

descending overpass of the SMAP L3 radiometer 36 km EASE-grid soil moisture product version 8 (L3_SM_P) was selected for this purpose (O'Neill et al. 2021). This product is available at https://nsidc.org/data/SPL3SMP/versions/8.

3.2. Active and passive airborne datasets

The airborne instruments used in the SMAPEx campaigns included the 1.41 GHz Polarimetric L-Band Multibeam Radiometer (PLMR) and the 1.26 GHz Polarimetric L-Band Imaging Synthetic Aperture Radar (PLIS), which provided the L-band passive (brightness temperature) and active (backscatter) microwave observations. Overall, there are nine flight dates from SMAPEx-3 (5th, 7th, 10th, 13th, 15th, 18th, 19th, 21st and 23rd September, 2011), six flight dates from SMAPEx-4 (2nd, 5th, 10th, 11th, 19th and 21st May, 2015) and eight flight dates from SMAPEx-5 (8th, 10th, 13th, 16th, 18th, 21st, 23rd and 26th September, 2015) covering several 3 dB SMAP radiometer footprints. Notably, SMAPEx-4 data was coincident with both SMAP radiometer and radar observations.

The passive airborne radiometer brightness temperature data for SMAPEx experiments was collected by the PLMR instrument with 1 km spatial resolution at horizontal and vertical (*h* and *v*) polarizations and nominal incidence angles of 17°, 21.5° and 38.5°. An accuracy of around ± 1.4 K was obtained for the calibration of PLMR brightness temperature at vertical and horizontal polarization, and an accuracy of about ± 1.5 K for thermal correction of the calibrated dataset was achieved during the SMAPEx campaigns (Ye et al. 2020). The PLMR brightness temperature observations were angle normalized from their original angles to the reference incidence angle of SMAP (~ 40º) utilizing a cumulative distribution function approach (Ye et al. 2015). An accuracy of about ± 2.4 K was achieved for angle normalization of the PLMR brightness temperature (Wu et al. 2015). As the SMAP soil moisture data did not exist for the training phase, due to being in the pre-launch period, the SMAPEx-3 airborne retrieved soil moisture at 1 km spatial resolution (Ye et al. 2020) was averaged to 36 km resolution to simulate the SMAP derived soil moisture data to train the machine learning algorithm. Additionally, the derived soil moisture observations from SMAPEx-4 and SMAPEx-5 PLMR brightness temperature at 1 km spatial resolution over the entire SMAP validation flight area (Fig. 1) were used in the testing phase of the machine learning algorithm development, for the purpose of evaluating the downscaling algorithm results. During the SMAPEx experiments, the airborne radar backscatter datasets were measured by the PLIS instrument at *hh*, *hv*, *vh* and *vv* polarizations, high temporal resolution and 10 m spatial scale (Ghafari et al. 2020; Zhu et al. 2018) with an incidence angle between 15º to 45º. The PLIS instrument provided complete coverage over the study area during SMAPEx-3, but with small gaps across the SMAPEx-4 and 5 campaigns due to the flight design. However, previous investigations on the PLIS coverage gaps demonstrated that there was a nonsignificant effect on the accuracy of the PLIS backscatter when processed to 3 km resolution for use in downscaling (Ghafari et al. 2020). Before using the PLIS observations in the machine learning technique, the data was calibrated, georeferenced, and normalized for the incidence angle, with an accuracy of 0.58 dB achieved for calibration (Zhu et al. 2018). To normalize the PLIS incidence angle to that of SMAP (40º), the method utilized for angle normalization of the PLMR observations was performed (Ye et al. 2015). An accuracy of 0.8 dB was achieved for the angle normalized backscatter data at 1 km resolution (Wu et al. 2015). Finally, the PLIS backscatter data was aggregated by linear averaging from the original grid cell (10 m) to the required resolution (1 km). In this study, the vertical and horizontal co-polarized and cross-polarized PLIS backscatter (σ_{vv}, σ_{hh} and σ_{xpol}) were used.

The airborne instrument of the SMAPVEX-12 campaign is called the Passive Active L-band Sensor (PALS), providing L-band radiometer brightness temperature with both vertical and horizontal polarization at 1.41 GHz frequency, and L-band radar backscatter with *hh*, *hv*, *vh* and *vv* polarizations at 1.26 GHz frequency. The PALS instrument was mounted to provide a single beam with a 40º incidence angle looking to the rear of the aircraft (McNairn et al. 2015). Sixteen flight dates of SMAPVEX-12 (7th, 12th, 15th, 17th, 22nd, 25th, 27th and 29th June, and 3rd, 5th, 8th, 10th, 13th, 14th, 17th and 19th July, 2012) provided active and passive airborne measurements for the machine learning algorithm training over the SMAPVEX-12 area. In this research, the calibrated co-polarized and cross-polarized PALS backscatter observations (σ_{vv}, σ_{hh} and σ_{xpol}), Version 1 (SV12PLBK) (Colliander 2014) measured over SMAPVEX-12 agricultural sampling fields (nominal size of 800 m × 800 m) were resampled through a linear averaging approach to provide the 1 km resolution radar observations, while the retrieved soil moisture data, Version 1 (SV12PLSM) achieved from PALS brightness temperature observations (Colliander 2017; Colliander et al. 2016) at 1 km spatial resolution was utilized to simulate the 36 km SMAP soil moisture. The SMAPVEX data was only used in the training step of developing the machine learning based downscaling algorithm. More description about the PALS instrument and its radar and radiometer calibration methodologies are available in McNairn et al. (2015).

3.3. MODIS Normalized Difference Vegetation Index (NDVI)

Machine learning methods are able to integrate various data sources. Utilizing vegetation index parameters in the satellite soil moisture downscaling methods has been one of the widely accepted approaches over the past decade (Fang and Lakshmi 2014; Merlin et al. 2008; Piles et al. 2011). The MODerate resolution Imaging Spectroradiometer (MODIS) is a multispectral instrument of the NASA Earth Observing System, consisting of Aqua and Terra satellites which measure the visible, near infrared, and thermal infrared signatures at 36 spectral bands every 1 to 2 days. In this study the daytime overpass of Terra, being most consistent with the SMAP overpass, was selected to extract the NDVI variable. The selected MODIS product was the version-061 daily surface spectral reflectance (MOD09GA) at 1 km spatial resolution, available at https://e4ftl01.cr.usgs.gov/MOLT/. The reflectance product is available at 500 m spatial resolution. However, for consistency with the microwave data it was resampled to 1 km resolution before calculating NDVI.

3.4. Soil texture data

Soil texture, including clay, silt and sand content, is one of the basic parameters affecting the soil moisture values, through its influence on the rate of water infiltration, soil moisture storage and soil drainage characteristics. Accordingly, several studies have shown that information on soil texture can be one of the important sources in downscaling soil moisture using machine learning (Abbaszadeh et al. 2019; Karthikeyan and Mishra 2021).

In this study, the machine learning algorithm utilized the information on soil texture (% clay, % silt, and % sand). The soil texture of the SMAPEx ground sampling site is clay loam (31% clay, 48% silt and 20% sand) for YA4 and YA7, silty clay loam (39% clay, 43% silt and 17% sand) for YE, and loam (23% clay, 47% silt and 29% sand) for YB5, YB7 and YF. The SMAPEx ground sampling soil texture data values, and also the soil texture information over the SMAP validation flight area (Fig. 1), were obtained from gravimetric experiments that extracted soil particle size distribution and the CSIRO Digital Atlas of Australian Soils (1991). The soil texture information for SMAPVEX-12 was extracted from soil texture data collected by coring devices over each agricultural field as part of the campaign. The soil texture types varied over this selected area including sand (7% clay, 4% silt and 89% sand), loamy sand (6% clay, 6% silt and 88% sand), sandy clay loam (34% clay, 14% silt and 51% sand), sandy loam (16% clay, 9% silt and 75% sand), silty clay loam (40% clay, 56% silt and 4% sand), clay (56% clay, 30% silt and 14% sand), heavy clay (67% clay, 29% silt and 4% sand), clay loam (38% clay, 19% silt and 43% sand) and silty clay (54% clay, 40% silt and 6% sand). This dataset is accessible at https://nsidc.org/data/smap/validation-data (Bullock et al. 2014).

3.5. Geographic data

Soil moisture conditions, especially in the surface layers, are affected by topographic data (Crow et al. 2012). As elevation, terrain slope and aspect have been found to be the important topographic parameters in soil moisture downscaling studies (Mascaro et al. 2011; Wilson et al. 2005), these features were selected for use in the machine learning model developed here to downscale the SMAP soil moisture. The topography of the Murrumbidgee River catchment changes from 50 m to 2000 m (Fig. 1), however, based on the 250 m topography information from the Geoscience Australia Digital Elevation Model (DEM), the elevation at 1 km spatial resolution for the SMAPEx study area only changed from 100 m to 400 m throughout the SMAP validation flight area. The terrain slope and aspect values were derived from DEM information of the SMAP validation flight area, and changed from 0º to 12º and from -1º to 360º respectively at 1 km resolution. The DEM product obtained from the ASTER Global-DEM project (https://asterweb.jpl.nasa.gov/gdem.asp) has been used for SMAPVEX, having a 30 m spatial resolution with a vertical accuracy of 7 m to 14 m. Based on the data extracted from ASTER, the mean elevations at the USDA agricultural fields varied from 237 m to 276 m when averaged to 1 km resolution, while the terrain slope and aspect values changed from 3º to 7.8º and from 127.2º to 215.1º.

3.6. Ground soil moisture observations

Each of the SMAPEx campaigns included six focus areas (3 km × 3 km) aligned with the SMAP radar grid cells, with dense soil moisture cluster monitoring stations to monitor soil moisture, along with intensive spatial ground sampling (Fig. 1). During the campaigns, intensive soil moisture values were monitored over the 0-5 cm depth concurrent with airborne overpasses at the focus areas using the Hydraprobe Data Acquisition System (HDAS) (Merlin et al. 2007). The soil moisture information was recorded on a 250 m × 250 m grid over each SMAPEx focus area. Three soil moisture values were measured at each ground sample point within a radius of one meter to consider soil moisture variations, reducing the impact of errors in measuring the data. For use in this study, these soil moisture values were aggregated through linear averaging within each 1 km grid, being the target spatial resolution.

The selected ground soil moisture of the SMAPVEX-12 experiments for this research was from the temporary soil moisture sensors installed by the United States Department of Agriculture (USDA). As mentioned earlier, there were 55 measurement sites known as agricultural fields (Fig. 2). Soil moisture values during the SMAPVEX-12 experiments varied spatially due to variations in soil texture, the topography of the area, and differences in field irrigation management. To provide valid average soil moisture measurements, sixteen sampling points with three replicates at each point were selected for every agricultural field (mostly 800 m × 800 m fields representing about 1 km spatial resolution) to measure ground soil moisture over the 0-5 cm depth. Replication was utilized to decrease the error resulting from spatial variability in soil properties. The average soil moisture data at each agricultural field was considered as the 1 km ground reference value. The soil moisture was measured using a Stevens Water Hydra Probe (McNairn et al. 2015). The information for the selected datasets is presented in Table 1.

4. Methodology

4.1. Summary of the Random Forest technique

Random forest is a machine learning method that functions as an ensemble multiple decision tree model (Breiman 1996, 2001). Importantly, the overfit situation may easily occur in the training stage with this approach, leading to poor performance during the testing phase. To overcome this problem, the random forest model makes several decision trees that work individually at the training stage, with the output data achieved by calculating the average prediction of those trees. Accordingly, the input features are divided by the random forest algorithm into several regression trees, so that each tree is produced through a bootstrap sample providing its own prediction value. Overall, the reduction in generalization error occurs due to the combination of results from several decision trees (Breiman 2001). Based on previous research, random forest is the most appropriate machine learning approach for regression and classification problems, such as downscaling of satellite products like soil moisture (He et al. 2016; Long et al. 2019; Mao et al. 2022), as it makes the decision trees using the adaptive, randomized and independent features for the relation between input and output variables (Amit and Geman 1997; Breiman 2001).

Table 1. Characteristics of the datasets utilized in the machine learning approach.

Data set	Details	Source	Spatial resolution	Temporal resolution	Time series / Dates
SMAP Level 3 soil moisture	Version 8, SMAP radiometer soil moisture product	NSIDC	36 km	2-3 days	May, 2015 (six dates) September, 2015 (eight dates)
PLMR soil moisture	Airborne soil moisture data from SMAPEx-3, SMAPEx-4 and SMAPEx-5 experiments	smapex.monash.edu	1 km	Daily	September, 2011 (nine dates) May, 2015 (six dates) September, 2015 (eight dates)
PLIS backscatter	Active airborne backscatter data from SMAPEx-3, SMAPEx-4 and SMAPEx-5 experiments	smapex.monash.edu	10 m (resampled to 1 km)	Daily	September, 2011 (nine dates) May, 2015 (six dates) September, 2015 (eight dates)
PALS soil moisture	Airborne soil moisture data from SMAPVEX-12 experiment	NSIDC	1500 m (resampled to 1 km)	Daily	June, 2012 (eight dates) July, 2012 (eight dates)
PALS backscatter	Active airborne backscatter data from SMAPVEX-12 experiment	NSIDC	500 m, and 1500 m (resampled to 1 km)	Daily	June, 2012 (eight dates) July, 2012 (eight dates)
Normalized Difference Vegetation Index (NDVI)	Extracted from MODIS MOD09GA – version 061	NASA LP DAAC	1 km	Daily	September, 2011 (nine dates) June, 2012 (eight dates) July, 2012 (eight dates) May, 2015 (six dates) September, 2015 (eight dates)
Soil Texture	Variables (% Clay, Silt, Sand)	CSIRO, and SMAPVEX-12 field surveys	1 km	Static	September, 2011 (nine dates) June, 2012 (eight dates) July, 2012 (eight dates) May, 2015 (six dates) September, 2015 (eight dates)
Terrain features	Digital Elevation Model (DEM), Terrain slope and Aspect	Geoscience Australia, and ASTER Global-DEM project	1 km	Static	September, 2011 (nine dates) June, 2012 (eight dates) July, 2012 (eight dates) May, 2015 (six dates) September, 2015 (eight dates)
Ground soil moisture	Focus areas of SMAPEx, and USDA agricultural fields of SMAPVEX-12	smapex.monash.edu, and NSIDC	Resampled to 1 km	Daily	September, 2011 (nine dates) June, 2012 (eight dates) July, 2012 (eight dates) May, 2015 (six dates) September, 2015 (eight dates)

4.2. Soil moisture downscaling method

The target of this research was to develop a random forest algorithm that leads to soil moisture at finer resolutions (i.e., 1 km), utilizing datasets sourced from before and after the

SMAP launch. The basic idea for the approach is to construct a transfer function between different input variables and the soil moisture output variable using:

$$SM_d = f(C) + \varepsilon \, , \tag{1}$$

$$C = (c_1, c_2, c_3, \dots c_N) \, , \tag{2}$$

where the SM_d is the downscaled surface soil moisture, ε is the model estimation error, and c_i demonstrates the individual input variables, including co-polarized and cross-polarized backscatter (σ_{vv}, σ_{hh} and σ_{xpol}), geographic data (elevation, terrain slope and aspect), soil texture (% clay, % silt, and % sand), airborne radiometer-based soil moisture and NDVI, and N is the dimension of input predictors (N = 11 in this study).

The training of the random forest algorithm used 11 input variables that are at or resampled to the resolution of 1 km to downscale the SMAP radiometer soil moisture product (L3_SM_P). These included retrieved soil moisture data from airborne radiometer measurements at 1 km resolution aggregated to 36 km and radar backscatter in co-polarized and cross-polarized channels (σ_{vv}, σ_{hh} and σ_{xpol}) aggregated to 1 km, NDVI as being representative of the vegetation dynamics, soil texture, and geographic data including the Digital Elevation Model **(**DEM), derived terrain slope and aspect (Table 1). These parameters have shown a strong relationship with the temporal dynamics and spatial heterogeneity of soil moisture (Abbaszadeh et al. 2019; Abowarda et al. 2021; Zhu et al. 2020). As the training phase of the random forest algorithm needs a source of soil moisture data as the output response variable, the 1 km ground soil moisture datasets were utilized for this purpose. While the 1 km resolution radiometer observations could also have been used to aid in the training, this was not done in this instance. It is notable that over the SMAPEx-3 and SMAPVEX-12 experiments, the ground soil moisture datasets were only measured at the focus areas (size of 3 km × 3 km each as shown in Fig. 1) and at the agricultural fields (size of 800 m × 800 m each as shown in Fig. 2), respectively.

Moreover, the SMAP radar backscatter data from the active passive product (SMAP_L2_SM_AP) resampled to 1 km resolution were used rather than the PLIS backscatter during the SMAPEx-4 campaign.

The dataset was split into two groups: i) the data collected before the SMAP launch to train the random forest model, and ii) the data collected after the SMAP launch, unseen by the random forest model, and thus used at the validation phase to verify the resultant downscaling model. Therefore, to investigate the main objective of this study, the data collected during SMAPEx-3 and SMAPVEX-12 (in the years 2011 and 2012, respectively) were used for the training phase of the model, and the data collected during SMAPEx-4 and SMAPEx-5 (in the year 2015) were used for the testing phase. Because the training phase was before the SMAP launch, the radiometer derived soil moisture from SMAPEx-3 and SMAPVEX-12 at 1 km resolution was aggregated to 36 km and used as the input soil moisture data in place of the SMAP 36 km soil moisture during training. In contrast, the SMAP 36 km radiometer soil moisture observations were utilized as the input at the validation stage and the 1 km resolution SMAPEx soil moisture data were used only for validation of the downscaled soil moisture. Table 2 presents numerical information regarding the available data in the training and validation phases.

The 12 columns were considered during the training of the machine learning algorithm, which include the 11 input variables (available at 1 km or resampled to 1 km) and the one output response variable (1 km ground soil moisture data). As an example, the focus area of SMAPEx-3 provides a data set with 162 rows and 12 columns, where the 162 is computed as 2×9×9, with 2 referring to the number of ground sampling focus areas with an available dataset for each day, 9 refers to the number of 1 km grid cells at each focus area (i.e., 3 km × 3 km), and the last 9 refers to the number of experiment days during the campaign. The 11 input variables were normalized from 0 to 1 before being utilized in both the training and validation

Table 2. Description of the data used for training and validation phases, including the number of 1 km grid cells, number of experiment days and the total available samples over selected campaigns.

	Campaign	**Number of 1 km grid cells**	**Number of experiment days**	**Total available samples**
Training phase	SMAPEx-3	18	9	162
	SMAPVEX-12	25–50	16	585
Validation phase	SMAPEx-4	6035	6	33234
	SMAPEx-5	6319	8	50552

phases. This step was to remove any error due to the non-equal magnitudes of the input variables (Breiman 2001; O and Orth 2021; Srivastava et al. 2013). Subsequently, the SMAP soil moisture data was downscaled via the trained algorithm, utilizing the SMAPEx-4 and SMAPEx-5 data for the input variables. Finally, the SMAP downscaled soil moisture, was evaluated utilizing the ground soil moisture datasets and the high-resolution airborne radiometer derived soil moisture.

Fig. 3 presents a schematic of the proposed random forest model for downscaling the 36 km SMAP soil moisture. The random forest algorithm requires the input variables on a 1 km grid. Therefore, the data collections which were not originally at 1 km resolution were resampled to this spatial resolution. The MATLAB built-in function TreeBagger from the MATLAB Regression Learner application was used to apply the random forest algorithm, working based on the Bagging (Bootstrap + Aggregating) approach (Breiman 1996, 2001). Using this method, the training dataset was sampled to M subgroups by the bootstrap approach, and the M individual regression decision trees fitted to train the random forest algorithm through using the input variable data. The predicted data was calculated through M replications. Finally, the average of the output values from the individual decision trees was considered as the final result value. The ensemble decision was made by averaging the M results from individual regression trees:

$$p(SM_d|C) = \frac{1}{M}\sum_{t=1}^{M} p_t(SM_d|C), \tag{3}$$

where $p_t(SM_d|C)$ is the output of each individual decision tree determining the conditional distribution of the downscaled soil moisture (SM_d) considering the multidimensional feature input vector (C).

The k-fold cross-validation technique (Hastie et al. 2009) was also included in the model to avoid overfitting. A k-value equal to 5 was selected as it showed the best performance during the training, obtained through a trial and error approach. It is also important to choose the appropriate values for minimum leaf size and number of learners applied in the random forest model during the training phase to improve the downscaling accuracy. For this purpose, different values were tested through trial and error, with a minimum leaf size equal to nine and a number of learners equal to 25 yielding the best performance of the trained random forest model in improving the downscale ing accuracy. After the training phase, the best calibrated random forest regression model was exported for implementation on the validation dataset, allowing the evaluation using unseen data. Accordingly, the SMAP radiometer soil moisture observations over the SMAP validation flight area (Fig. 1) were downscaled utilizing the calibrated random forest algorithm to 1 km resolution, and evaluated by the fine resolution ground soil moisture measurements at the focus areas averaged over 1 km grids, and also the soil moisture retrieved from the airborne brightness temperature at 1 km.

Validation of the downscaled soil moisture included quantification of statistical metrics and model errors, by comparing the estimated values with the airborne retrieved soil moisture observations and ground soil moisture measurements as the reference data. These metrics

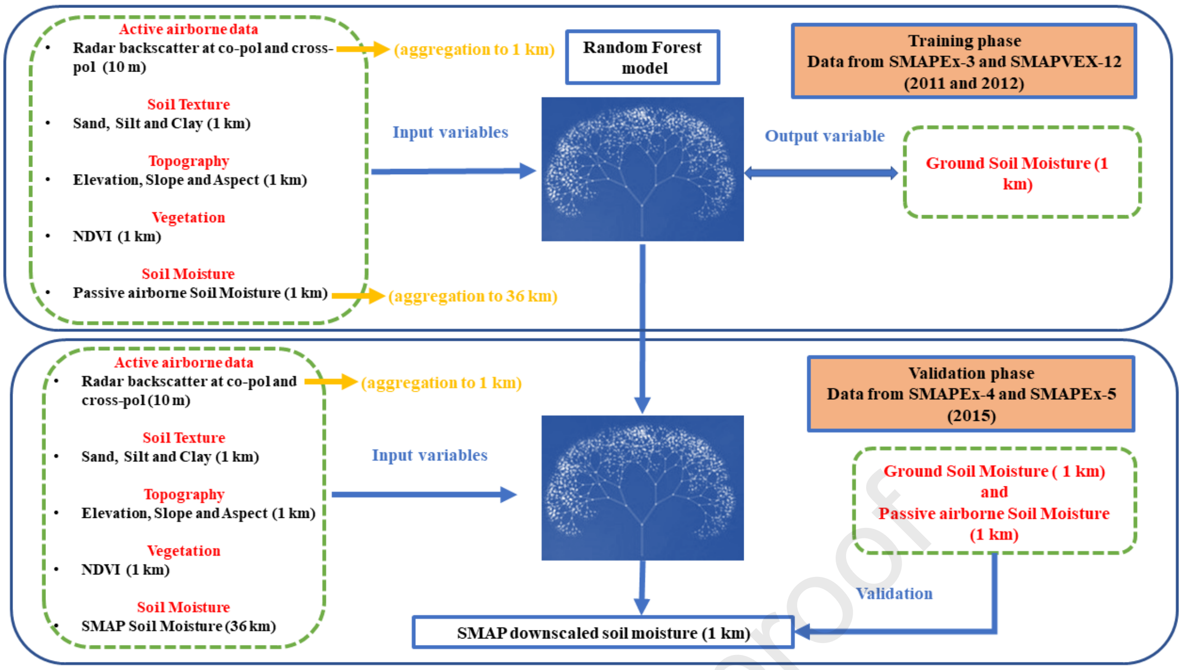

Fig.3. Flowchart of the proposed random forest downscaling model.

include the unbiased root mean square error (*ubRMSE*), Pearson correlation coefficient (*R*), and mean difference or bias. The *ubRMSE* was considered as the representative accuracy of the soil moisture in this research.

The importance of each individual variable was assessed to analyze the relative contribution of input features on the random forest downscaling accuracy. For this purpose, a leave-one-out approach was performed by removing the one input variable (i.e., radar backscatter, NDVI, DEM, terrain slope and aspect, soil texture) and implementing the random forest downscaling algorithm using the rest of the variables in order to investigate the impact of the removed variable.

5. Results and discussion

5.1. Evaluation of the soil moisture data sets

Original SMAP_L3, PLMR and PALS soil moisture observations (resampled to 1 km spatial resolution) were first evaluated against the ground soil moisture measured during the experiment periods. Fig. 4 demonstrates the evaluation of the different soil moisture observations against ground soil moisture measurements at the pixel level, including the statistical analysis values. Accordingly, the correlation coefficients between the PALS and PLMR retrieved soil moisture with the ground measurement were found to be higher than the SMAP_L3 soil moisture by 0.03 $m^3.m^{-3}$ and 0.07 $m^3.m^{-3}$ for PLMR and PALS, respectively. In contrast, the original SMAP_L3 soil moisture showed better *ubRMSE* against ground measurements than the airborne soil moisture retrieval from PLMR and PALS, achieving the lowest value equal to 0.062 $m^3.m^{-3}$. The highest *ubRMSE* was obtained between PLMR soil moisture retrieval and ground observations as 0.09 $m^3.m^{-3}$, which was partially due to standing water found in grasslands (due to heavy rainfall) at the beginning of the SMAPEx-5 campaign and crop lands (due to flood irrigation) at the end of the SMAPEx-5 campaign. Importantly, the PLMR instrument captured the soil moisture variation of these pixels. The calculation of bias statistics showed an overestimation for SMAP_L3 soil moisture of 0.013 $m^3.m^{-3}$, and an underestimation for PLMR and PALS soil moisture of -0.008 $m^3.m^{-3}$ and -0.029 $m^3.m^{-3}$ respectively when compared against ground measurements.

For comparison, the SMAP_L3 soil moisture has been resampled to 1 km resolution by applying the same soil moisture value for each 1 km pixel within each 36 km EASE grid cell. The resampled SMAP soil moisture has then been compared against the PLMR soil moisture obtained during the SMAPEx-4 and -5 experiments. This is considered as the “do-nothing” baseline performance that the downscaling algorithm must beat in order to add value. Fig. 5 shows that the comparison had a correlation coefficient of 0.66, bias of 0.016 $m^3.m^{-3}$ (SMAP_L3 higher), and *ubRMSE* of 0.121 $m^3.m^{-3}$. Thus, in order to ensure that the differences between the SMAP downscaled soil moisture and the airborne retrieved soil moisture at high spatial resolution were affected only by the machine learning downscaling algorithm, and not

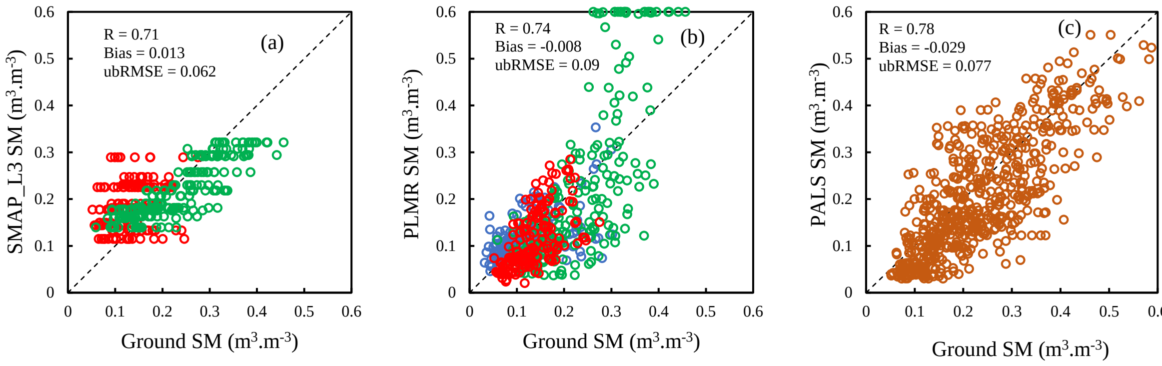

Fig. 4. Evaluation of different soil moisture observations against ground soil moisture measurements including a) 36 km SMAP_L3 SM during 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points) at the SMAPEx-4 and -5 sites, respectively; b) 1 km PLMR SM during 5–23 September, 2011 (SMAPEx-3, blue points), 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points) at the SMAPEx-3, -4 and -5 sites, respectively; and c) 1 km PALS SM during 7 June – 19 July, 2012 at the SMAPVEX-12 site (SMAPVEX-12, brown points).

because of the sensor to sensor bias, this bias value between the SMAP and PLMR soil moisture was removed before utilizing the data in the downscaling process.

5.2. Results from random forest model development

The calibration and validation of the proposed random forest algorithm was conducted using the input and output variables over selected areas. As mentioned earlier, the normalized training dataset from the SMAPEx-3 and SMAPVEX-12 experiments was partitioned into a 5-fold cross-validation. In the training phase, the 1 km ground soil moisture dataset was used in the algorithm for matching with the output response variable (see Fig. 3). The statistical results of the Ensemble TreeBagger algorithm applied at the training phase showed a good performance with R, root mean square error ($RMSE$) and mean absolute error (MAE) of 0.88, 0.05 $m^3.m^{-3}$ and 0.04 $m^3.m^{-3}$, respectively, demonstrating the capability of the calibrated

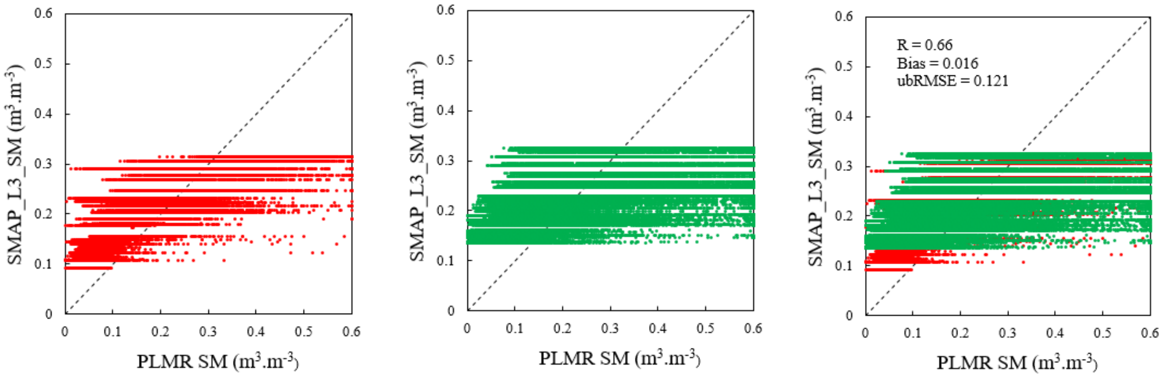

Fig. 5. Comparison of SMAP_L3 SM resampled to 1 km against 1km PLMR airborne soil moisture retrieval during 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points) over SMAPEx-4 and SMAPEx-5 flight areas, respectively.

random forest model for generalization to an unseen dataset. These results showed a better correlation coefficient and *RMSE* than Senanayake et al. (2021), which used the Gaussian process regression model over the Yanco area for downscaling of soil moisture. The statistical results of this research have been obtained by trying different numbers of decision trees and tree leaf size to achieve a suitable calibrated random forest model for the downscaling.

5.3. Assessment of the downscaling algorithm performance

5.3.1. Comparison of downscaled soil moisture with PLMR airborne retrieved data

Fig. 6 provides the scatterplots and statistical results of the SMAP downscaled soil moisture against PLMR soil moisture observations, which exhibit good agreements. The calculated *R*, bias, and *ubRMSE* were 0.97, 0.016 m^3.m^{-3} and 0.048 m^3.m^{-3}. The results show the improvement of *R* from 0.66 between the SMAP_L3 and PLMR soil moisture to 0.97 between the downscaled SMAP and PLMR soil moisture. Importantly, when utilizing the random forest algorithm trained only by the SMAPVEX-12 data there was no apparent degradation in the downscaled results (results not shown) when applied to the SMAPEx data, even though applied

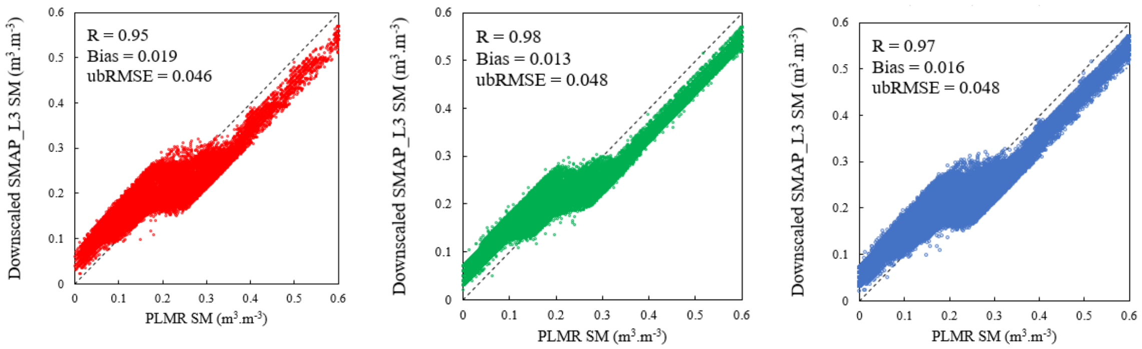

Fig. 6. Validation of downscaled SMAP soil moisture versus PLMR airborne soil moisture retrieval (1 km) during 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points) and all available data (blue points) over PLMR flight areas. All available data (blue points) include both SMAPEX-4 (red points) and SMAPEx-5 (green points) data.

to an entirely independent site, suggesting that there is some degree of transferability of the machine learning approach to locations different to those used for training. Additionally, the random forest algorithm was trained utilizing data over the entire flight areas of the SMAPEx-3 (36 km × 38 km) and SMAPVEX-12 (12.8 km × 70 km) study areas (Fig. 2) on experiment days. In this case, the calculated *R*, bias, and *ubRMSE* between the downscaled SMAP soil moisture using the random forest model and PLMR soil moisture of SMAPEx-4 and SMAPEx-5 were 0.97, 0.015 $m^3.m^{-3}$ and 0.051 $m^3.m^{-3}$, being only slightly different from the results reported in Fig. 6. However, the scatter plots indicate an overestimation at lower soil moisture values, and an underestimation between downscaled SMAP soil moisture and PLMR values at higher soil moisture values. Importantly, *ubRMSE*, the main statistical metric of the downscaling algorithm accuracy, improved from 0.121 $m^3.m^{-3}$ to 0.048 $m^3.m^{-3}$, showing good downscaling performance by the proposed random forest model.

Overall, the statistical results achieved through the comparison of the downscaled SMAP pixel with the PLMR soil moisture showed the success of the developed random forest algorithm in downscaling the SMAP soil moisture. The results of the random forest method utilized in this study are encouraging, especially when evaluated with the results of the original SMAP soil moisture reported in Fig. 5, with an improved accuracy of downscaled SMAP soil moisture against PLMR measurements, and the results of earlier studies shown in Sabaghy et al. (2020) for the same site. Consequently, the quality of PLMR observations and their full spatial coverage over the selected area have provided a good opportunity to investigate machine learning based downscaling.

In order to assess the soil moisture spatial distribution, the spatial pattern of SMAP downscaled soil moisture were investigated against the course resolution SMAP observations and the airborne retrieved soil moisture. Figs. 7 and 8 present the spatial variability in the downscaled, original SMAP soil moisture, and PLMR retrieved soil moisture over the SMAP validation flight area of SMAPEx-4 (71 km × 85 km) and SMAPEx-5 (71 km × 89 km) during each of the experiment days (D is representative of the day). The downscaled maps closely correspond to the airborne soil moisture retrieval patterns. The rainfall events on 9th and 18th May (D3 and D5) during SMAPEx-4 were clearly captured by the spatial pattern, as the soil moisture in these days showed higher values than others (Fig. 7). The dry down pattern during SMAPEx-5 from D1 to D8 corresponds to the rainfall events that preceded the campaign (Fig. 8). Overall, the downscaled soil moisture closely matched the pattern of the PLMR observations during both the SMAPEx-4 and SMAPEx-5 experiments, conducted under diverse climate and vegetation conditions.

To further analyse the capability of the downscaling model at capturing the soil moisture change, the pattern of the temporal variation of the SMAP downscaled and airborne soil moisture was investigated. There were several heavy rainfall events before both the SMAPEx-4 and SMAPEx-5 campaigns, providing heterogeneous soil moisture conditions with dry downs. Furthermore, the two additional rainfall events on the 9th and 18th of May during the SMAPEx-4 experiments are visible in Fig. 7 as increased soil moisture values. In contrast, there was no significant additional rainfall during SMAPEx-5, resulting in a prolonged dry down. Figs. 7 and 8 show that both the SMAP course resolution and the downscaled soil moisture values correspond to the temporal variability of the PLMR soil moisture in response to rainfall events. For instance, the higher amounts of soil moisture at the beginning of SMAPEx-5 are attributed to rainfall followed by a dry down with a distinct soil moisture pattern that is clearly detected. However, the consistency of the original and downscaled SMAP soil moisture with the PLMR soil moisture was affected based on the land cover and atmospheric situations. In the following, the differences are discussed according to the soil moisture dynamic ranges. For this purpose, the minimum and maximum amounts of soil moisture have been mentioned to clarify the ranges of the soil moisture.

Over the SMAPEx-4 site, the original and downscaled SMAP soil moisture varied from 0.09 $m^3.m^{-3}$ to 0.31 $m^3.m^{-3}$ and from 0.022 $m^3.m^{-3}$ to 0.57 $m^3.m^{-3}$, respectively. Over the SMAPEx-5 site, the SMAP course resolution and downscaled soil moisture varied from 0.13 $m^3.m^{-3}$ to 0.33 $m^3.m^{-3}$ and from 0.02 $m^3.m^{-3}$ to 0.57 $m^3.m^{-3}$, respectively. Overall, the range of downscaled SMAP soil moisture was more than the range of the original SMAP soil moisture over SMAPEx-4 and SMAPEx-5. In addition, the PLMR soil moisture ranged from 0 $m^3.m^{-3}$ to 0.6 $m^3.m^{-3}$ during SMAPEx-4 and SMAPEx-5. According to Figs. 7 and 8, it can be seen that the soil was generally wetter and with larger range during SMAPEx-5 than SMAPEx-4. Moreover, the vegetation water content was high with actively growing vegetation, and agricultural activities such as irrigation affecting the soil moisture ranges and the standing water, leading to increased PLMR retrieval uncertainties for some pixels. In order to minimize these errors, the bias value between the original SMAP soil moisture and the PLMR retrieved soil moisture was removed.

To investigate the spatial distribution of errors during the downscaling process, the actual difference plots between the downscaled SMAP soil moisture and PLMR observations have

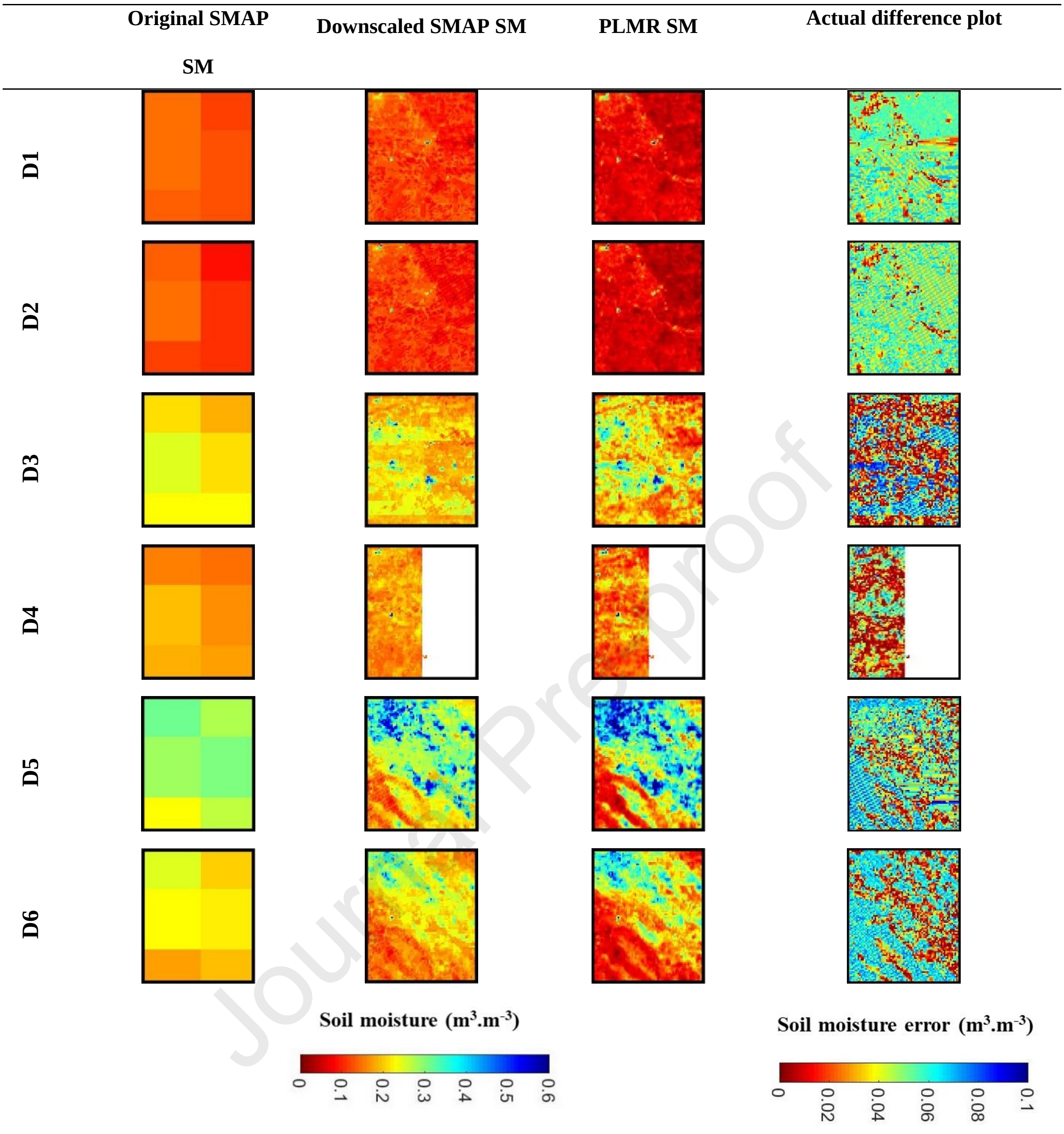

Fig. 7. Spatial distribution of original SMAP_L3 soil moisture (36 km), downscaled SMAP soil moisture (1 km), and airborne PLMR retrived soil moisture (1 km) at 5 cm depth during the period 30 April – 23 May, 2015 at SMAPEx-4 over PLMR flight area (71 km × 85 km).

also been presented in Figs. 7 and 8. Overall, the difference values gave good agreement between PLMR and downscaled soil moisture, but showed that the errors between PLMR and downscaled products at the dry and wet soil moisture conditions had more bias than under more normal soil moisture situations.

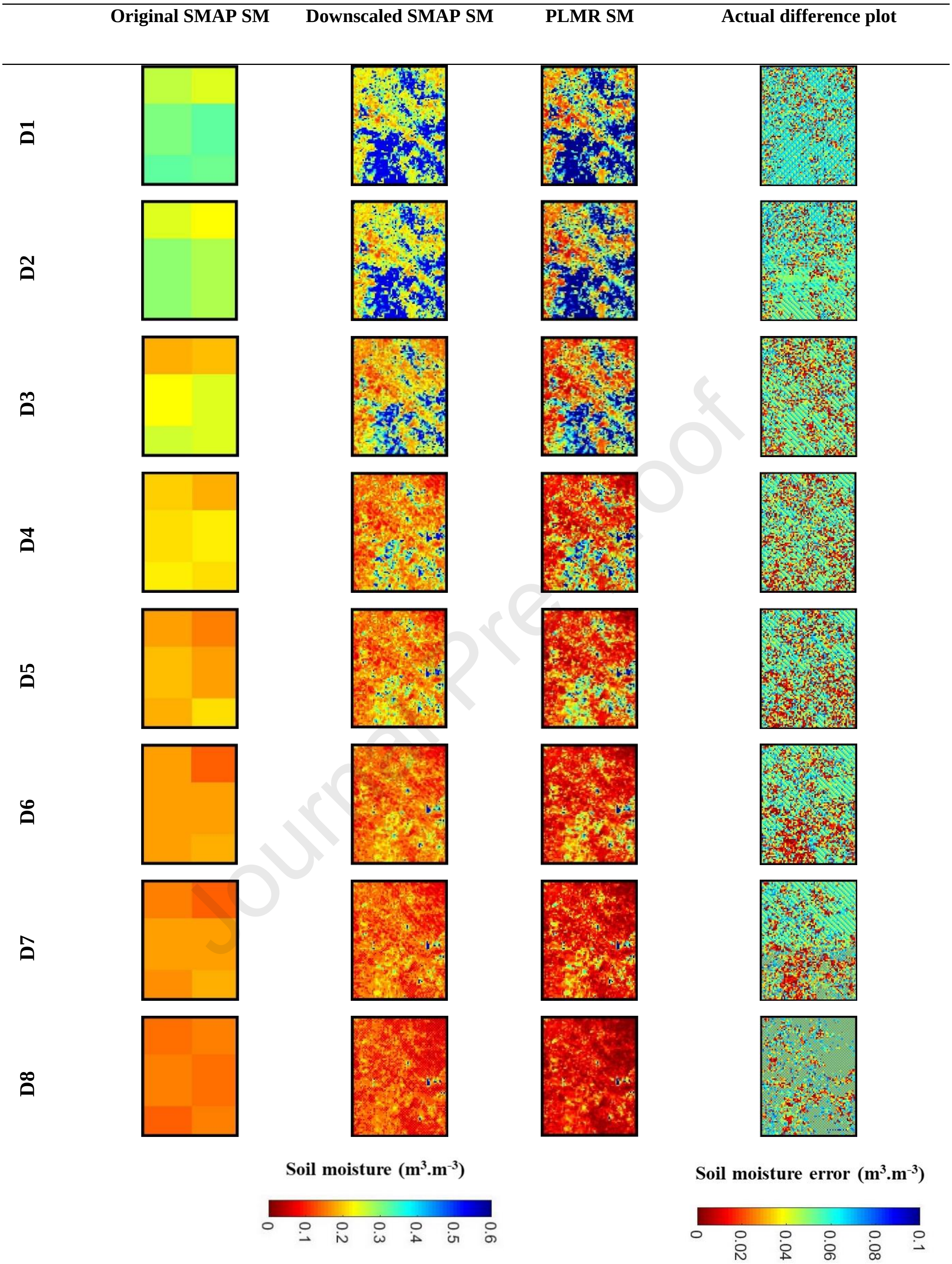

Fig. 8. Same as Fig. 7 except for SMAPEx-5 over PLMR flight area (71 km × 89 km) during the period 6–28 September, 2015.

580 **5.3.2. Comparison of downscaled soil moisture with ground measurements**

The calculated *R*, bias, and *ubRMSE* of the downscaled SMAP soil moisture against the ground data were 0.73, -0.047 $m^3.m^{-3}$ and 0.057 $m^3.m^{-3}$ for clay loam, 0.83, -0.038 $m^3.m^{-3}$ and 0.072 $m^3.m^{-3}$ for loam, and 0.8, -0.031 $m^3.m^{-3}$ and 0.06 $m^3.m^{-3}$ for silty clay loam (Fig. 9). The statistical results demonstrated that the downscaled soil moisture had a good correlation with the ground soil moisture observations over these soil texture conditions, especially for loam and silty clay loam textures, and an underestimation of downscaled soil moisture over all of the selected soil texture conditions. The *ubRMSE* showed better performance for the clay loam and silty clay loam soil textures than the loam soil texture condition.

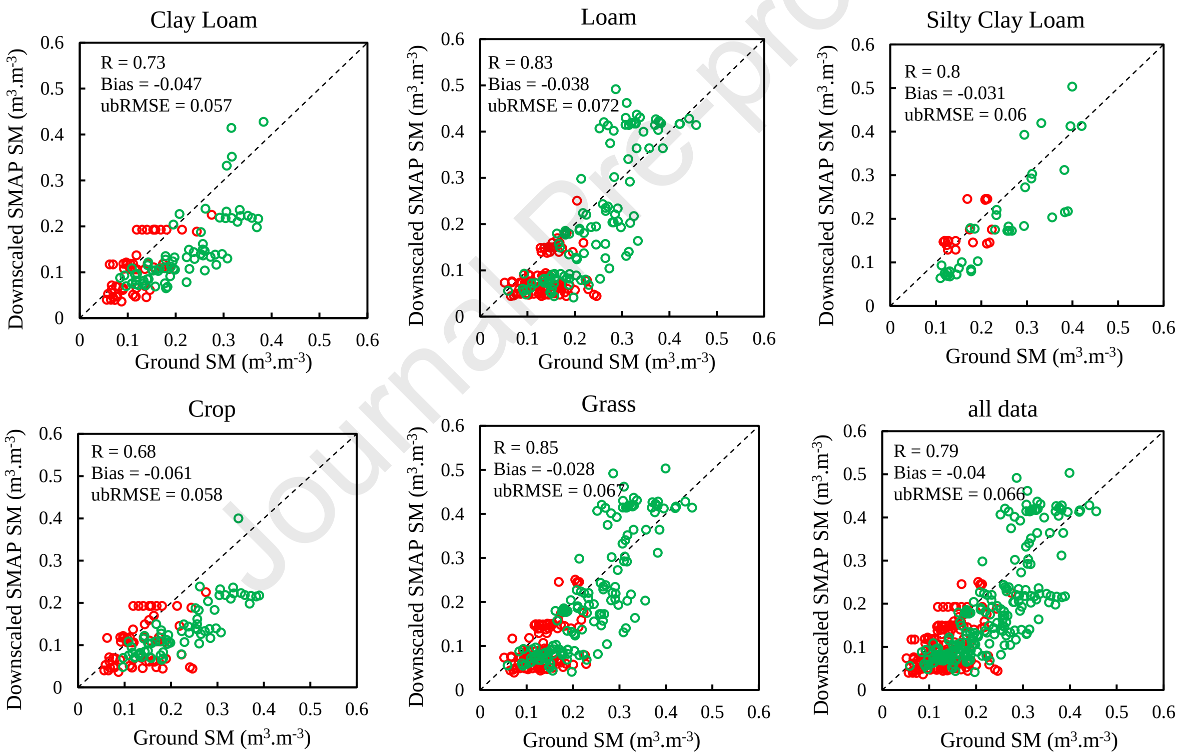

Fig. 9. Validation of downscaled SMAP soil moisture versus ground soil moisture measurements (1 km) over the SMAPEx focus area during 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points). The first row presents the results for different soil texture conditions, and the second row shows the results for different land cover types, along with the results from all available data.

The performance of the downscaled SMAP soil moisture was also assessed considering the two types of land covers. The *R*, bias, and *ubRMSE* were 0.68, -0.061 $m^3.m^{-3}$ and 0.058 $m^3.m^{-3}$ for the cropland, and 0.85, -0.028 $m^3.m^{-3}$ and 0.067 $m^3.m^{-3}$ for the grassland (Fig. 9). The bias was negative (downscaled SMAP soil moisture lower) for both cropland and grassland, and although the *R* was worse for the cropland than for the grassland, the *ubRMSE* was better for the cropland than for the grassland.

The *R*, bias, and *ubRMSE* were 0.79, -0.04 $m^3.m^{-3}$ and 0.066 $m^3.m^{-3}$ for all data over selected focus areas (Fig. 9). Although the *R* was better compared with those for SMAP_L3 (Fig. 4), the *ubRMSE* was not better in this case. Overall, the results of this study are consistent with those from Abbaszadeh et al. (2019), which utilized the random forest approach for SMAP soil moisture downscaling over the Continental United States at different soil texture conditions.

For a more detailed investigation, the performance of the downscaled SMAP soil moisture was assessed with the SMAPEx-4 and SMAPEx-5 data separately. Because these campaigns were conducted in different seasons, they provide insight into the effects of different atmospheric conditions, soil moisture variations, and variability in vegetation. The SMAPEx-4 data was collected in the austral autumn with the land surface type of bare soil in croplands and grasslands covered by short grass. In comparison, the SMAPEX-5 took place during the austral spring when the crops were in the growth stage with high vegetation water content, and grassland vegetation was at mature stages, as described earlier. Table 3 reports the statistical analysis, including *R*, bias, and *ubRMSE* between the downscaled SMAP soil moisture and ground measurements considering the soil texture and land cover scenarios for SMAPEx-4 and SMAPEx-5 experiments. The *R* showed good values for all scenarios of the SMAPEx-5 experiment. Moreover, *R* showed acceptable values for the SMAPEx-4 experiment with the exception of loam soil texture and croplands. While the *ubRMSE* values meet the SMAP soil

Table 3. The statistical metrics of soil moisture comparison between ground soil moisture and SMAP downscaled estimates according to land cover and soil texture during SMAPEx-4 and SMAPEx-5, separately.

	Campaign name	SMAPEx-4			SMAPEx-5		
		R	**Bias**	***ubRMSE***	***R***	**Bias**	***ubRMSE***
		-	$m^3.m^{-3}$	$m^3.m^{-3}$	-	$m \cdot m^{-3}$	$m^3.m^{-3}$
Soil texture	**Clay Loam**	0.76	-0.013	0.04	0.68	-0.073	0.055
	Loam	0.27	-0.055	0.051	0.82	-0.026	0.084
	Silty Clay Loam	0.56	0.005	0.038	0.86	-0.048	0.062
Land Cover	**Crop**	0.36	-0.034	0.059	0.82	-0.086	0.046
	Grass	0.53	-0.034	0.046	0.84	-0.025	0.078

moisture accuracy requirement for nearly all selected soil texture and land cover situations over SMAPEx-4, the *ubRMSE* over SMAPEx-5 showed worse values than SMAPEx-4 results, except for the cropland situation.

Overall, considering the ground soil moisture measurements as an independent reference, the proposed random forest model improved the accuracy of downscaled SMAP soil moisture over the focus areas of SMAPEx-4, when comparing with the uniform values from the original SMAP_L3 product. However, the statistics of the downscaled SMAP soil moisture did not show equal improvement for the focus areas of SMAPEx-5. It seems that at these focus areas, the downscaling performance was affected by the high vegetation water content and flood irrigation during the SMAPEx-5 experiments.

Fig. 10 and Fig. 11 present the spatial variability of the downscaled and original SMAP_L3 soil moisture, and ground soil moisture measurements over the 3 km × 3 km SMAPEx focus areas of SMAPEx-4 and SMAPEx-5 during each of the experimental days. The downscaled soil moisture maps correspond to the patterns of the ground soil moisture observations by

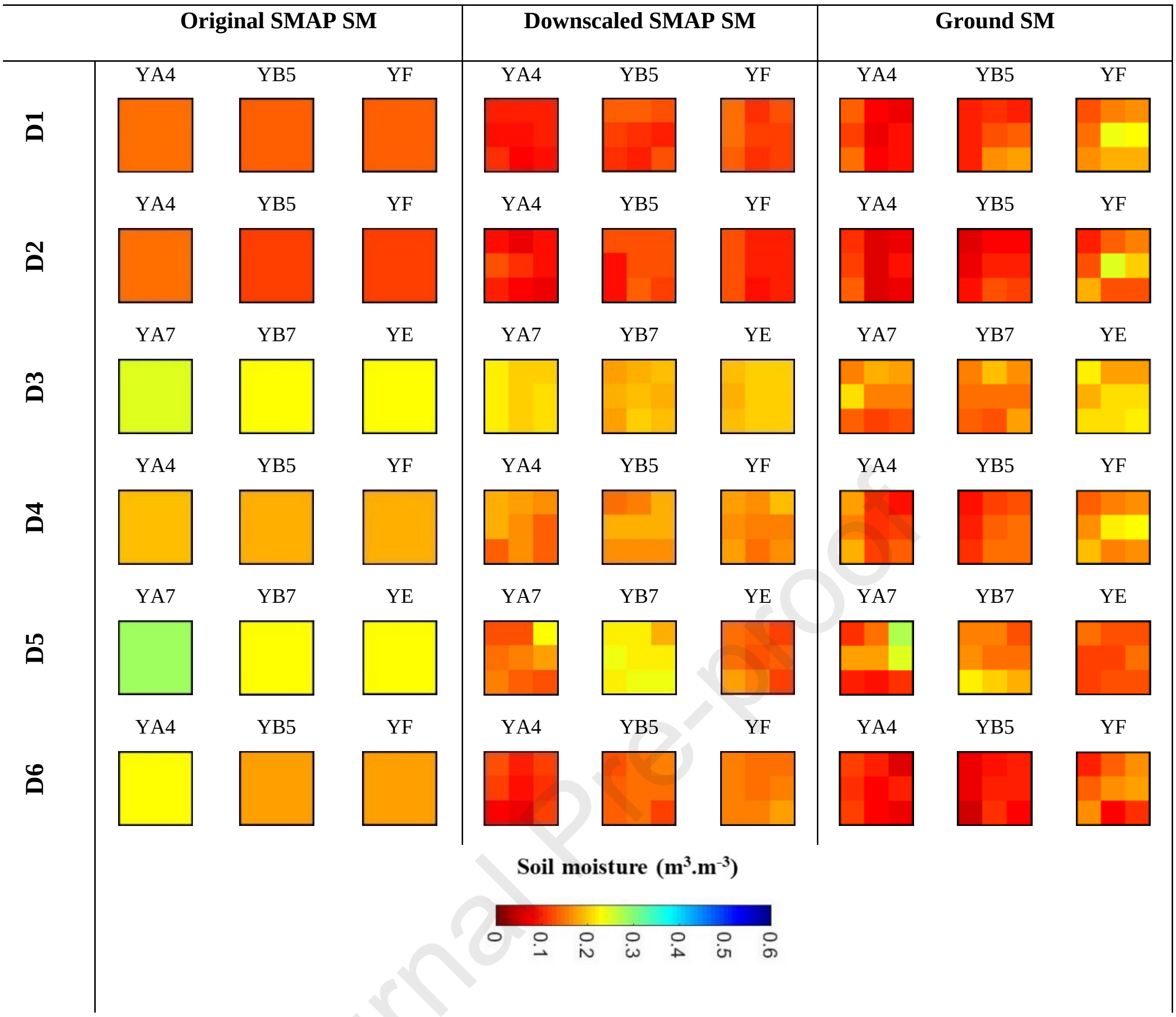

Fig. 10. Spatial distribution of original SMAP_L3 soil moisture (36 km), downscaled SMAP soil moisture (1 km) and ground soil moisture (1 km) measurments at 5 cm depth during the period 30 April – 23 May, 2015 at SMAPEx-4 focus areas (3 km × 3 km).

generally capturing the rainfall events during SMAPEx-4 (Fig. 10) and the dry down pattern of SMAPEx5 due to the rainfall events prior to the campaign (Fig. 11).

The geographic parameters such as topography, vegetation coverage, and soil texture contribute to the heterogeneity. While the topography of the SMAPEx focus areas does not change substantially, there are three distinct soil texture types. Considering the spatial distribution based on the soil textures for SMAPEx-4 (Fig. 10), the downscaled soil moisture matched the spatial pattern of the ground soil moisture qualitatively for different soil texture conditions. Considering the spatial distribution based on the land cover of SMAPEx-4, it seems that the spatial distribution of soil moisture within the focus area under both grassland and cropland showed good consistency when compared to the ground soil moisture measurements. It is notable that conditions included bare soil in the cropland and sparsely vegetated dry grassland during the SMAPEx-4 experiment. These conditions will have affected the soil drying states as well as rapid infiltration after any rainfall or irrigation.

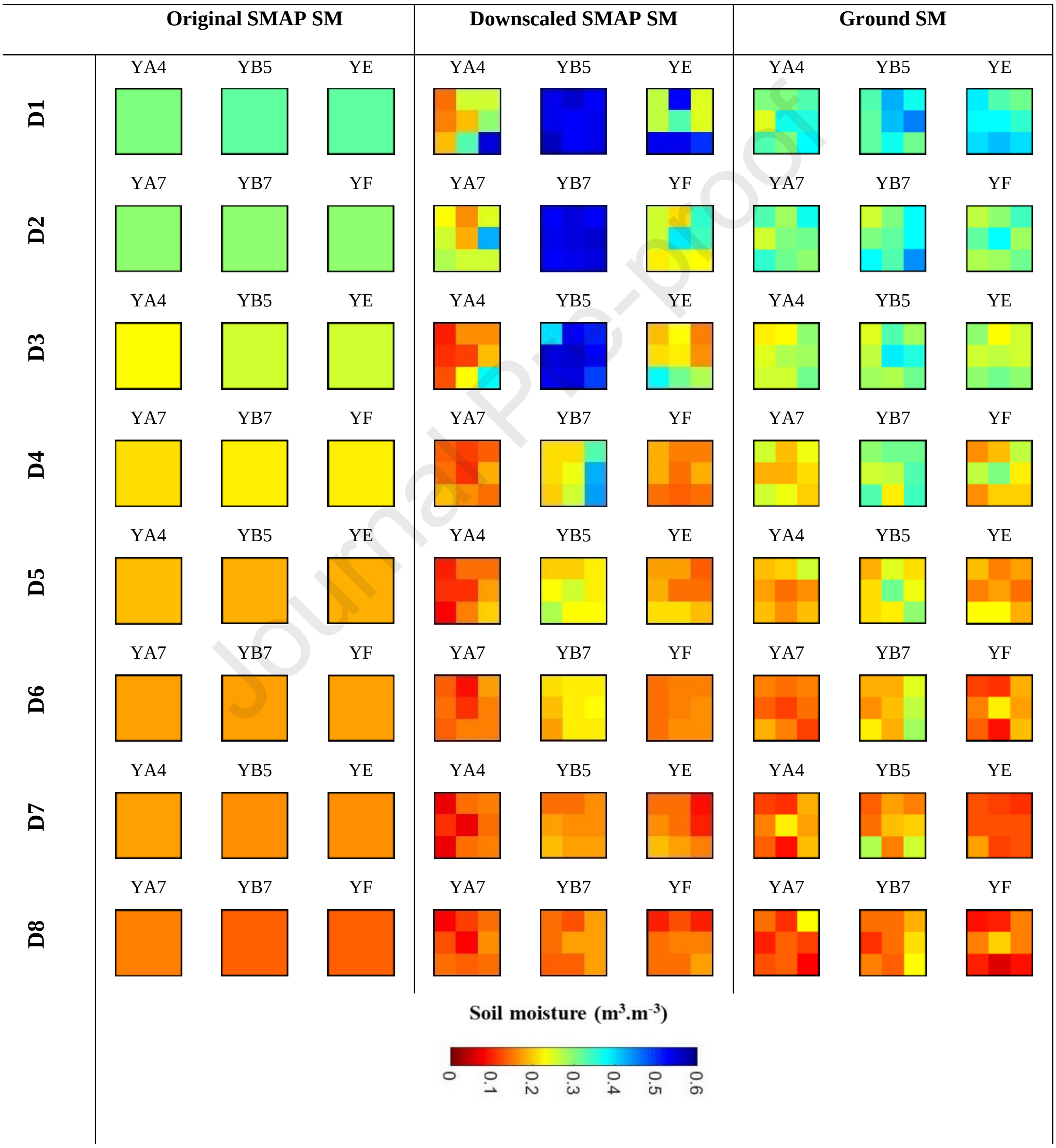

Fig. 11. Same as Fig. 10 except for SMAPEx-5 focus areas (3 km × 3 km) during the period 6–28 September, 2015.

Based on the spatial distribution maps, greater heterogeneity in the soil moisture spatial distribution was visible for vegetated and irrigated areas during SMAPEx-5. However, the greater vegetation led to an increased attenuation of the microwave signal, contributing to an underestimation of soil moisture. Considering the spatial distribution in the different soil texture variations, for SMAPEx-5 (Fig. 11) the downscaled soil moisture matched qualitatively the spatial pattern of the ground soil moisture for the clay loam texture type (YA4, YA7). Moreover, considering the spatial distribution based on the land cover of SMAPEx-5, the soil moisture spatial distribution of the focus area showed qualitatively better consistency under croplands (YA4 and YA7) than grasslands. The random forest method showed higher uncertainty under grassland (YB5 and YB7) in downscaling the SMAP soil moisture during the early days of the campaign, which were influenced by standing water. Table 3 reports the same results, with the *ubRMSE* values over SMAPEx-5 achieving 0.046 $m^3.m^{-3}$ and 0.078 $m^3.m^{-3}$ for cropland and grassland, respectively.

5.4. Results of utilizing 36 km SMAP data at 1 km in the training phase

Most of the machine learning based passive soil moisture downscaling approaches to date have focused on utilizing the coarse resolution grid cell soil moisture uniformly across all fine resolution grid cells as an input in the training phase. In order to understand the effect of such assumptions, the results from utilizing the 1 km soil moisture values at the focus areas were compared with results from utilizing 36 km grid cell average soil moisture at the same focus areas as the input in the training phase of the random forest algorithm. Fig. 12 shows the scatter plots and the statistical analysis of the downscaled SMAP soil moisture against airborne PLMR soil moisture and the ground soil moisture measurements for the two different approaches over the ground sampling focus area. When using the average soil moisture of the 36 km grid cell as the input in the training phase, the statistical metrics *R*, bias, and *ubRMSE* of the SMAP downscaled soil moisture against PLMR soil moisture were 0.53, 0.055 $m^3.m^{-3}$, and 0.083

$m^3.m^{-3}$, and against ground soil moisture were 0.54, 0.014 $m^3.m^{-3}$ and 0.074 $m^3.m^{-3}$. The accuracy of the downscaled SMAP soil moisture derived from the random forest algorithm based on the proposed approach of this paper clearly showed better performance in *ubRMSE* (by 0.04 $m^3.m^{-3}$) than the algorithm based on utilizing the averaged soil moisture. Additionally, the range of downscaled soil moisture based on utilizing the average soil moisture changed from 0.02 $m^3.m^{-3}$ to 0.15 $m^3.m^{-3}$, which was substantially lower than the range of downscaled soil moisture based on the proposed approach. Overall, utilizing 1 km grid soil moisture observations as the input in the training phase showed a better skill level in matching with observed soil moisture patterns, meaning that it can construct a well-trained downscaling algorithm.

5.5. Importance of input variables to the downscaled soil moisture

The importance of different variables must be analysed in order to realize their effectiveness on the performance of the random forest algorithm for soil moisture downscaling. In random forest models, the increased percentage of *MSE* in comparison with that achieved from utilizing all variables in the model describes the importance of different variables. When an important variable is not used in the algorithm, the *MSE* will increase, with the larger the increase in *MSE* signifying the greater the importance of that variable (Breiman 2001). Therefore, the significance of each input variable was analysed using the ablation test, in which each input variable was independently omitted from the downscaling process and the random forest algorithm applied using the remaining variables. Ten different input schemes including removal of the radar backscatter (σ_{vv}, σ_{hh}, σ_{xpol}), NDVI, DEM, slope, aspect and soil texture (% clay, silt and sand) were tried independently. Removal of soil moisture from the input variables increased the percentage of *MSE* value equal to 23.8 %, showing the highest importance in this machine learning downscaling approach. Therefore, as downscaling of the SMAP soil moisture was the main purpose of this research, the soil moisture parameter was included in the input

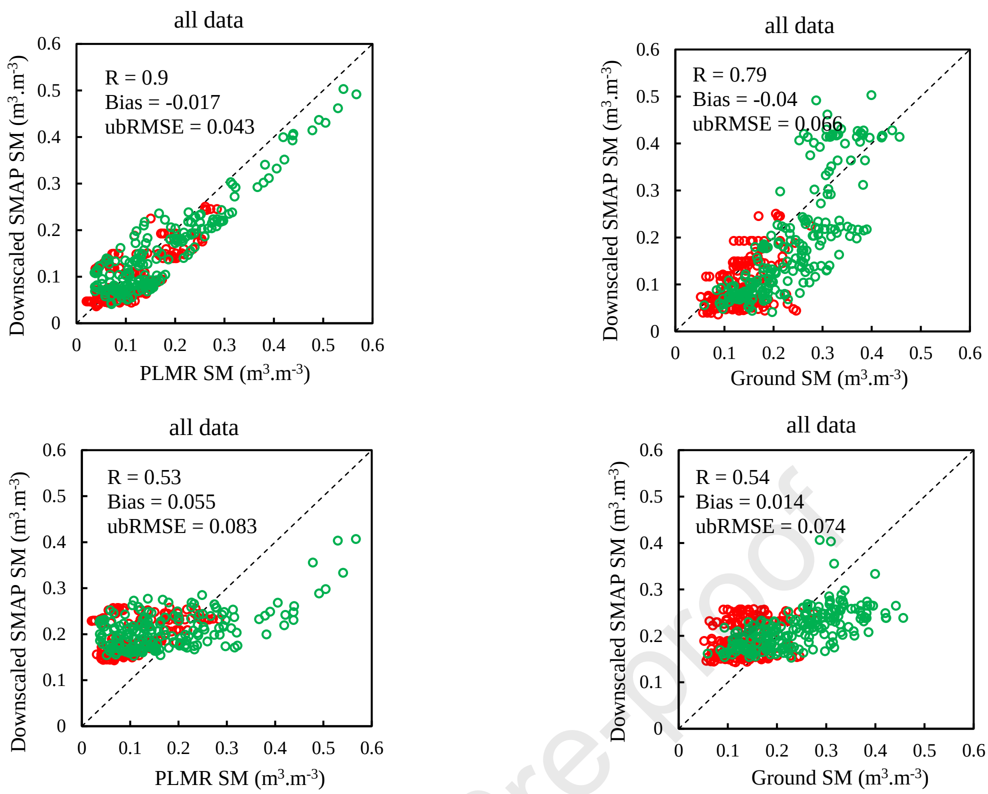

Fig. 12. Validation of downscaled SMAP soil moisture against airborne PLMR retrieved soil moisture (1 km) and ground soil moisture measurements (1 km) over the SMAPEx focus areas during 30 April – 23 May, 2015 (SMAPEx-4, red points) and 6–28 September, 2015 (SMAPEx-5, green points). The first row presents the results for utilizing 1 km soil moisture at focus areas as the input in the training phase, and the second row shows the results for utilizing the average soil moisture of the 36 km grid cell as the input in the training phase.

schemes. The *MSE* values of the downscaled SMAP soil moisture estimates relative to PLMR soil moisture were calculated separately for each input variable, with percentage of increase in *MSE* values shown in Fig. 13.

Horizontal backscatter (σ_{hh}) and slope are recognized as the most important variables (3.66 % and 3.63 %, respectively), showing more influence than other variables on the random forest accuracy. Soil texture ranked third, indicating 2.85 % and 2.76 % importance for sand and clay, respectively. The soil texture can influence water permeability, infiltration rate and water storage capacity. In this assessment, silt fraction showed least importance compared to other input variables in the proposed downscaling model. NDVI also showed high importance (2.51 %) due to the ability of presenting the vegetation status; NDVI is one of the crucial auxiliary parameters used in soil moisture retrieval, and in several soil moisture downscaling algorithms (Colliander et al. 2017a; Colliander et al. 2017b). Among the airborne co-polarized and cross-polarized backscatter products utilized in the random forest model, the importance of horizontal co-polarized backscatter (σ_{hh}) was highest (3.66 %). However, the vertical co-polarized backscatter (σ_{vv}) and cross-polarized backscatter (σ_{hv}) also showed a high influence. While the DEM had a slightly lower influence (1.73 %) compared to the high influence input variables on the results, it is one of the important input variables in the proposed random forest algorithm in this study. However, previous research has shown the high importance of a vegetation index such as NDVI in soil moisture retrieval, and the low importance of the DEM (Abowarda et al. 2021; Karthikeyan and Mishra 2021). Overall, it is suggested that utilizing all selected input variables in the downscaling model would be necessary to obtain the best downscaling accuracy. Moreover, landcover type was not included here as an option due to challenges in including categorical information in machine learning models. However, given the strong relationship in backscatter response to different landcover types and their associated land surface conditions, this could also be an important variable for use in future investigations.

5. Conclusion

This study presented a new strategy for downscaling the 36 km SMAP radiometer soil moisture product to 1 km spatial resolution. A random forest model using 1 km resolution remotely sensed backscatter, together with 1 km resolution vegetation characteristics, topography and soil properties, was used to downscale 36 km resolution passive microwave satellite soil moisture, based on training to focus areas with 1 km resolution soil moisture. The model was trained using data acquired pre-launch of SMAP, and evaluated with post-launch of SMAP airborne and field soil moisture data. Soil moisture from focus areas at 1 km spatial

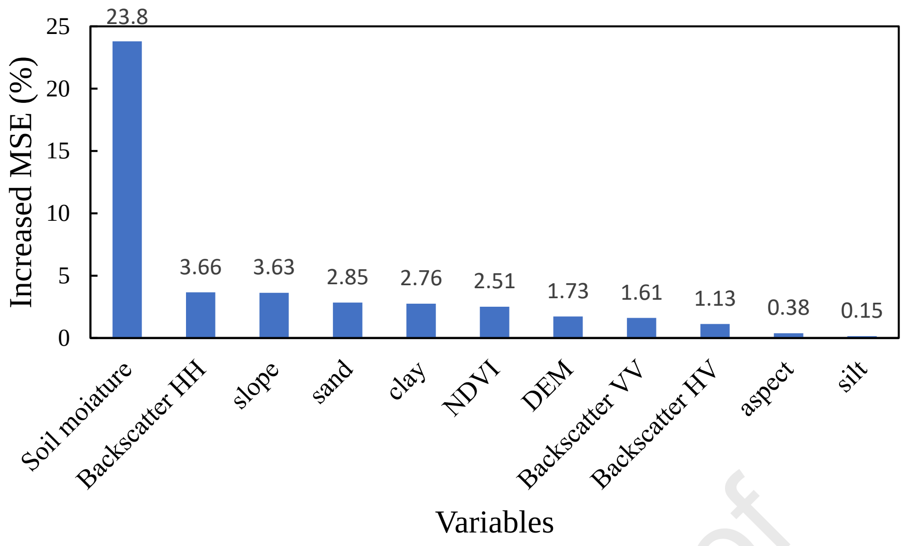

Fig. 13. Importance of input variables of the random forest model to the downscaled SMAP soil moisture calculated through increased mean square error (*MSE*) in percentage, including soil moisture, σ_{vv}, σ_{hh}, σ_{xpol}, NDVI, DEM, slope, aspect and soil texture (clay, silt and sand).

resolution were utilized to train the random forest algorithm, rather than the more traditional approach of using the SMAP 36 km soil moisture. The SMAP downscaled soil moisture product from the proposed random forest downscaling algorithm was then validated using post-launch airborne retrieved soil moisture observations and the ground soil moisture measurements from multiple points. This study was performed considering different soil characteristics and land cover conditions, including both grasslands and a variety of crops.

Based on the validation results, the downscaled SMAP soil moisture demonstrated an excellent agreement with the airborne soil moisture observations over the flight area of SMAPEx-4 and SMAPEx-5. The statistical results between the downscaled SMAP and airborne PLMR retrieved soil moisture in terms of *R*, bias and *ubRMSE* were 0.97, 0.016 $m^3.m^{-3}$ and 0.048 $m^3.m^{-3}$, respectively. Overall, compared to the original passive SMAP soil moisture product applied as a uniform field, the proposed downscaling random forest algorithm showed the ability to improve the *ubRMSE* of downscaled SMAP soil moisture from 0.121 $m^3.m^{-3}$ (Fig.

5) to 0.048 $m^3.m^{-3}$ (Fig. 6), when considering the PLMR soil moisture as a reference, being close to the SMAP soil moisture accuracy requirement. The results of this study show that the proposed random forest downscaling algorithm has the ability to be applied regionally by training to a few local pixels at 1 km in order to downscale the coarse resolution microwave soil moisture, and that training in one location (SMAPVEX-12) could be applied to another location (SMAPEx). Moreover, the statistics between the downscaled SMAP and ground soil moisture measurements over the SMAPEx focus areas achieved 0.79, -0.04 $m^3.m^{-3}$ and 0.066 $m^3.m^{-3}$ in terms of *R*, bias and *ubRMSE*, respectively. Additionally, the downscaled SMAP soil moisture observations satisfactorily captured the spatial and temporal heterogeneity relative to ground and airborne soil moisture observations.

In order to investigate the importance of using data at fine spatial resolution to train the random forest algorithm, as was conducted for this research, the results were compared with those from the strategy of utilizing the average from a 36 km grid cell. In general, the statistical metrics showed a 0.04 $m^3.m^{-3}$ improvement in terms of *ubRMSE* downscaling accuracy by using the higher spatial resolution training data, when evaluated with airborne retrieved soil moisture.

An investigation on the importance of the input variables in the random forest algorithm revealed that the best downscaling accuracy was achieved through contribution of all the input variables tested. Overall, the assessment showed that the variable importance in the random forest downscaling approach utilized in this study was in the following order: horizontal backscatter (σ_{hh}), slope, sand, clay, NDVI, DEM, vertical backscatter (σ_{vv}), cross-polarized backscatter (σ_{hv}), aspect and silt. However, the use of a landcover map should also be considered in future studies.

Acknowledgments

The SMAPEx field campaigns and related research development have been funded by the Australian Research Council Discovery (DP0984586 and DP140100572) and Australian Research Council Linkage Infrastructure Equipment and Facilities grants (LE0882509 and LE0453434). A contribution to this work was made at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

References

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2019). Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. *Water Resources Research, 55*, 324-344

Abowarda, A.S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., & Sun, Z. (2021). 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 Sensing of Environment, 255*, 112301

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Comput., 9*, 1545–1588

Barre, H.M.J.P., Duesmann, B., & Kerr, Y.H. (2008). SMOS: The Mission and the System. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 587-593

Bindlish, R., O'Neill, P., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., & Kerr, Y. (2016). Assessment of the SMAP Passive Soil Moisture Product. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 1-14

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123-140

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32

Bullock, P., Berg, A., & Wiseman, G. (2014). SMAPVEX12 Core-Based Soil Texture Data, Version 1. *NASA National Snow and Ice Data Center Distributed Active Archive Center*

Chan, S., Bindlish, R., O'Neill, P., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., & Kerr, Y. (2016). Assessment of the SMAP Passive Soil Moisture Product. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 1-14

Colliander, A. (2014). SMAPVEX12 PALS Backscatter Data, Version 1 [Data Set]. In. Boulder, Colorado USA.: NASA National Snow and Ice Data Center Distributed Active Archive Center.

Colliander, A. (2017). SMAPVEX12 PALS Soil Moisture Data, Version 1 [Data Set]. In. Boulder, Colorado USA.: NASA National Snow and Ice Data Center Distributed Active Archive Center.

Colliander, A., Fisher, J.B., Halverson, G., Merlin, O., Misra, S., Bindlish, R., Jackson, T.J., & Yueh, S. (2017a). Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI During SMAPVEX15. *IEEE Geoscience and Remote Sensing Letters, 14*, 2107-2111

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S.B., Cosh, M., Dunbar, R.S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., Aljassar, H.K., & Yueh, S. (2017b). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment, 191*, 215-231

Colliander, A., Njoku, E.G., Jackson, T.J., Chazanoff, S., McNairn, H., Powers, J., & Cosh, M.H. (2016). Retrieving soil moisture for non-forested areas using PALS radiometer measurements in SMAPVEX12 field campaign. *Remote Sensing of Environment, 184*, 86-100

Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., de Rosnay, P., Ryu, D., & Walker, J.P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics, 50*

Das, N., Entekhabi, D., & Njoku, E. (2011). An Algorithm for Merging SMAP Radiometer and Radar Data for High-Resolution Soil-Moisture Retrieval. *Geoscience and Remote Sensing, IEEE Transactions on, 49*, 1504-1512

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

Entekhabi, D., G. Njoku, E., E. O'Neill, P., Kellogg, K.H., Crow, W., Edelstein, W.N., Entin, J., D. Goodman, S., J. Jackson, T., Johnson, J., Kimball, J., Piepmeier, J., D. Koster, R., Martin, N., McDonald, K., Moghaddam, M., Moran, S., Reichle, R., Shi, J., & van Zyl, J. (2010). *The Soil Moisture Active and Passive (SMAP) mission*.

Entekhabi, D., Yueh, S., O'Neill, P.E., Kellogg, K.H., Allen, A., Bindlish, R., Brown, M., Chan, S., Colliander, A., Crow, W.T., Das, N., De Lannoy, G., Dunbar, R.S., Edelstein, W.N., Entin, J.K., Escobar, V., Goodman, S.D., Jackson, T.J., Jai, B., Johnson, J., Kim, E., Kim, S., Kimball, J., Koster, R.D., Leon, A., McDonald, K.C., Moghaddam, M., Mohammed, P., Moran, S., Njoku, E.G., Piepmeier, J.R., Reichle, R., Rogez, F., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., Weiss, B., & West, R. (2014). SMAP Handbook–Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space

Fang, B., & Lakshmi, V. (2014). AMSR-E Soil Moisture Disaggregation Using MODIS and NLDAS Data. *Remote Sensing of the Terrestrial Water Cycle*, 277-304

Fang, K., & Shen, C. (2020). Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel. *Journal of Hydrometeorology, 21*, 399-413

Gao, Y., Colliander, A., Burgin, M.S., Walker, J.P., Dinnat, E., Chae, C., Cosh, M.H., Caldwell, T.G., Berg, A., Martinez-Fernandez, J., Seyfried, M., Starks, P.J., Bosch, D.D., McNairn, H., Su, Z., & van der Velde, R. (2022). Multi-frequency radiometer-based soil moisture retrieval and algorithm parameterization using in situ sites. *Remote Sensing of Environment, 279*, 113113

Ghafari, E., Walker, J.P., Das, N.N., Davary, K., Faridhosseini, A., Wu, X., & Zhu, L. (2020). On the impact of C-band in place of L-band radar for SMAP downscaling. *Remote Sensing of Environment, 251*, 112111

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*.

He, X., Chaney, N.W., Schleiss, M., & Sheffield, J. (2016). Spatial downscaling of precipitation using adaptable random forests. *Water Resources Research, 52*, 8217-8237

Hu, F., Wei, Z., Zhang, W., Dorjee, D., & Meng, L. (2020). A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. *Journal of Hydrology, 590*, 125360

Karthikeyan, L., & Mishra, A.K. (2021). Multi-layer high-resolution soil moisture estimation using machine learning over the United States. *Remote Sensing of Environment, 266*, 112706

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A., Bitar, A.A., Cabot, F., Gruhier, C., Juglea, S.E., Leroux, D., Mialon, A., & Delwart, S. (2012). The SMOS Soil Moisture Retrieval Algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 1384-1403

Kim, Y., & Zyl, J.J.v. (2009). A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 2519-2527

Lei, F., Senyurek, V., Kurum, M., Gurbuz, A., Boyd, D., Moorhead, R., Crow, W., & Eroglu, O. (2022). Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations. *Remote Sensing of Environment, 276*, 113041

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., & Shi, C. (2019). Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. *Remote Sensing of Environment, 233*, 111364

Mao, H., Kathuria, D., Duffield, N., & Mohanty, B.P. (2019). Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework. *Water Resources Research, 55*, 6986-7009

Mao, T., Shangguan, W., Li, Q., Li, L., Zhang, Y., Huang, F., Li, J., Liu, W., & Zhang, R. (2022). A Spatial Downscaling Method for Remote Sensing Soil Moisture Based on Random Forest Considering Soil Moisture Memory and Mass Conservation. *Remote Sensing, 14*, 3858

Mascaro, G., Vivoni, E., & Deidda, R. (2011). Soil moisture downscaling across climate regions and its emergent properties. *J. Geophys. Res, 116*

McNairn, H., Jackson, T.J., Wiseman, G., Bélair, S., Berg, A., Bullock, P., Colliander, A., Cosh, M.H., Kim, S.B., Magagi, R., Moghaddam, M., Njoku, E.G., Adams, J.R., Homayouni, S., Ojo, E.R., Rowlandson, T.L., Shang, J., Goïta, K., & Hosseini, M. (2015). The Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX12): Prelaunch Calibration and Validation of the SMAP Soil Moisture Algorithms. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 2784-2801

Merlin, O., Rudiger, C., Bitar, A.A., Richaume, P., Walker, J.P., & Kerr, Y.H. (2012). Disaggregation of SMOS Soil Moisture in Southeastern Australia. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 1556-1571

Merlin, O., Walker, J., Panciera, R., Young, R., Kalma, J., & Kim, E. (2007). Soil Moisture Measurement in Heterogeneous Terrain. *MODSIM 2007 International Congress on Modelling and Simulation*

Merlin, O., Walker, J.P., Chehbouni, A., & Kerr, Y. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment, 112*, 3935-3946

Monerris, A., Walker, J., Panciera, R., Jackson, T., Gray, D., Yardley, H., & Ryu, D. (2011). *The Third Soil Moisture Active Passive Experiment*.

Narayan, U., Lakshmi, V., & Jackson, T.J. (2006). High-resolution change estimation of soil moisture using L-band radiometer and Radar observations made during the SMEX02 experiments. *IEEE Transactions on Geoscience and Remote Sensing, 44*, 1545-1554

O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., Bindlish, R., & Chaubell, J. (2021). SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 8. *Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center*

O, S., & Orth, R. (2021). Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data, 8*, 170

Panciera, R., Walker, J.P., Jackson, T.J., Gray, D.A., Tanase, M.A., Ryu, D., Monerris, A., Yardley, H., Rüdiger, C., Wu, X., Gao, Y., & Hacker, J.M. (2014). The Soil Moisture Active Passive Experiments (SMAPEx): Toward Soil Moisture Retrieval From the SMAP Mission. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 490-507

Piles, M., Camps, A., Vall-llossera, M., Corbella, I., Panciera, R., Rüdiger, C., Kerr, Y., & Walker, J. (2011). Downscaling SMOS-derived soil moisture using MODIS Visible/Infrared data. *Geoscience and Remote Sensing, IEEE Transactions on, 49*, 3156-3166

Rao, P., Wang, Y., Wang, F., Liu, Y., Wang, X., & Wang, Z. (2022). Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China. *Earth Syst. Sci. Data, 14*, 3053-3073

Sabaghy, S., Walker, J.P., Renzullo, L.J., Akbar, R., Chan, S., Chaubell, J., Das, N., Dunbar, R.S., Entekhabi, D., Gevaert, A., Jackson, T.J., Loew, A., Merlin, O., Moghaddam, M., Peng, J., Peng, J., Piepmeier, J., Rüdiger, C., Stefan, V., Wu, X., Ye, N., & Yueh, S. (2020). Comprehensive analysis of alternative downscaled soil moisture products. *Remote Sensing of Environment, 239*, 111586

Sabaghy, S., Walker, J.P., Renzullo, L.J., & Jackson, T.J. (2018). Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment, 209*, 551-580

Schmugge, T., O'Neill, P.E., & Wang, J.R. (1986). Passive microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, 12-22

Senanayake, I.P., Yeo, I.Y., Walker, J.P., & Willgoose, G.R. (2021). Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. *Science of The Total Environment, 776*, 145924

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., & Teuling, A.J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*, 125-161

Srivastava, P.K., Han, D., Ramirez, M.R., & Islam, T. (2013). Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resources Management, 27*, 3127-3144

Wilson, D.J., Western, A.W., & Grayson, R.B. (2005). A terrain and data-based method for generating the spatial distribution of soil moisture. *Advances in Water Resources, 28*, 43-54

Wu, X., Walker, J., Rüdiger, C., Panciera, R., & Gao, Y. (2016). Intercomparison of Alternate Soil Moisture Downscaling Algorithms Using Active-Passive Microwave Observations. *IEEE Geoscience and Remote Sensing Letters, PP*, 1-5

Wu, X., Walker, J.P., Rüdiger, C., Panciera, R., & Gray, D.A. (2015). Simulation of the SMAP Data Stream From SMAPEx Field Campaigns in Australia. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 1921-1934

Ye, N., Walker, J.P., & Rüdiger, C. (2015). A Cumulative Distribution Function Method for Normalizing Variable-Angle Microwave Observations. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 3906-3916

Ye, N., Walker, J.P., Wu, X., Jeu, R.d., Gao, Y., Jackson, T.J., Jonard, F., Kim, E., Merlin, O., Pauwels, V.R.N., Renzullo, L.J., Rüdiger, C., Sabaghy, S., Hebel, C.v., Yueh, S.H., & Zhu, L. (2020). The Soil Moisture Active Passive Experiments: Validation of the SMAP Products in Australia. *IEEE Transactions on Geoscience and Remote Sensing*, 1-18

Zhao, W., Sánchez, N., Lu, H., & Li, A. (2018). A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of Hydrology, 563*, 1009-1024

Zhu, L., Walker, J.P., & Shen, X. (2020). Stochastic ensemble methods for multi-SAR-mission soil moisture retrieval. *Remote Sensing of Environment, 251*, 112099

Zhu, L., Walker, J.P., Ye, N., Rüdiger, C., Hacker, J.M., Panciera, R., Tanase, M.A., Wu, X., Gray, D.A., Stacy, N., Goh, A., Yardley, H., & Mead, J. (2018). The Polarimetric L-Band Imaging Synthetic Aperture Radar (PLIS): Description, Calibration, and Cross-Validation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*, 4513-4525

Zhu, L., Webb, G.I., Yebra, M., Scortechini, G., Miller, L., & Petitjean, F. (2021). Live fuel moisture content estimation from MODIS: A deep learning approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 179*, 81-91

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: