Retrieving Heterogeneous Surface Soil Moisture at 100 m across the Globe via Synergistic Fusion of Remote Sensing and Land Surface Parameters

Jingyi Huang[1], Ankur Rashmikant Desai[2], Jun Zhu[2], Alfred E Hartemink[1], Paul Stoy[3], Steven P Loheide II[2], Yakun Zhang[2], Zhou Zhang[2], and Francisco J. Arriaga[2]

[1]UNIVERSITY OF WISCONSIN-MADISON
[2]University of Wisconsin-Madison
[3]University of Wisconsin - Madison

November 23, 2022

Abstract

Soil water is essential for maintaining global food security and for understanding hydrological, meteorological, and ecosystem processes under climate change. Successful monitoring and forecasting of soil water dynamics at high spatio-temporal resolutions globally are hampered by the heterogeneity of soil hydraulic properties in space and complex interactions between water and the environmental variables that control it. Current soil water monitoring schemes via station networks are sparsely distributed while remote sensing satellite soil moisture maps have a very coarse spatial resolution. In this study, an empirical surface soil moisture (SSM) model was established via data fusion of remote sensing (Sentinel-1 and Soil Moisture Active and Passive Mission - SMAP) and land surface parameters (e.g. soil texture, terrain) using a quantile random forest (QRF) algorithm. The model had a spatial resolution of 100 m and performed moderately well across the globe under cropland, grassland, savanna, barren, and forest soils (R = 0.53, RMSE = 0.08 m m). SSM was retrieved and mapped at 100 m every 6-12 days in selected irrigated cropland and rainfed grassland in the OZNET network, Australia. It was concluded that the high-resolution SSM maps can be used to monitor soil water content at the field scale for irrigation management. The SSM model is an additive and adaptable model, which can be further improved by including soil moisture network measurements at the field scale. Further research is required to improve the temporal resolution of the model and map soil water content within the root zone.

Hosted file

`globalwatermodel_tables_02052020.docx` available at https://authorea.com/users/529106/articles/601322-retrieving-heterogeneous-surface-soil-moisture-at-100-m-across-the-globe-via-synergistic-fusion-of-remote-sensing-and-land-surface-parameters

Retrieving Heterogeneous Surface Soil Moisture at 100 m across the Globe via Synergistic Fusion of Remote Sensing and Land Surface Parameters

Jingyi Huang[1,]*, Ankur R. Desai[2], Jun Zhu[3], Alfred E. Hartemink[1], Paul C. Stoy[4], Steven P. Loheide II[5], Yakun Zhang[1], Zhou Zhang[4], Francisco Arriaga[1]

[1]Department of Soil Science, University of Wisconsin-Madison, Madison, WI 53706, USA

[2]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

[3]Department of Statistics and Entomology, University of Wisconsin-Madison, Madison, WI 53706, USA

[4]Department of Biological Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

[5]Department of Civil & Environmental Engineering, University of Wisconsin, Madison, WI 53706, USA

E-mail: jhuang426@wisc.edu

* – Corresponding author

Keywords: soil moisture network; machine learning; food security; drought; gross primary productivity; evapotranspiration; water use efficiency;

Abstract

Soil water is essential for maintaining global food security and for understanding hydrological, meteorological, and ecosystem processes under climate change. Successful monitoring and forecasting of soil water dynamics at high spatio-temporal resolutions globally are hampered by the heterogeneity of soil hydraulic properties in space and complex interactions between water and the environmental variables that control it. Current soil water monitoring schemes via *in situ* station networks are sparsely distributed while remote sensing satellite soil moisture maps have a very coarse spatial resolution. In this study, an empirical surface soil moisture (SSM) model was established via data fusion of remote sensing (Sentinel-1 and Soil Moisture Active and Passive Mission - SMAP) and land surface parameters (e.g. soil texture, terrain) using a quantile random forest (QRF) algorithm. The model had a spatial resolution of 100 m and performed moderately well across the globe under cropland, grassland, savanna, barren, and forest soils (R^2 = 0.53, RMSE = 0.08 m^3 m^{-3}). SSM was retrieved and mapped at 100 m every 6-12 days in selected irrigated cropland and rainfed grassland in the OZNET network, Australia. It was concluded that the high-resolution SSM maps can be used to monitor soil water content at the field scale for irrigation management. The SSM model is an additive and adaptable model, which can be further improved by including soil moisture network measurements at the field scale. Further research is required to improve the temporal resolution of the model and map soil water content within the root zone.

1. Introduction

Water plays a fundamental role in terrestrial ecosystems and human society. Soil water is a critical factor for a number of terrestrial biochemical, climate, and atmospheric processes and is the source of water for most of the crops that we eat (Vereecken et al., 2014). Monitoring and forecasting soil water content and fluxes (e.g. evapotranspiration, deep drainage) are essential for maintaining global food security (Hoekstra and Mekonnen, 2012) and understanding hydrological, meteorological, and ecosystem processes under climate change (Seneviratne et al., 2010; Trugman et al., 2018; Stoy et al., 2019).

Successful monitoring and forecasting soil water content and fluxes at high spatio-temporal resolutions globally is hampered by many factors, including heterogeneity of soil hydraulic properties in space (Robinson et al., 2008), complex interactions between water, environment, and human activities (Vereecken et al., 2014), and computational challenges (Chaney et al., 2018). Current regional and continental soil water monitoring networks are too sparsely distributed (e.g. ~100 km) to be used for field-scale research and application (e.g. irrigation) while remote sensing satellite soil moisture missions often have a coarse spatial resolution (> 1 km) (Ochsner et al., 2017).

Recent technological advances provide a potential solution to mapping soil water variability at the field scale. First, high-resolution remote sensing satellite missions have been launched to monitor soil water dynamics and land surface parameters (e.g. vegetation, terrain, and soil properties) have become available (Reuter et al., 2007; Friedl et al., 2010; Hengl et al., 2017; Fisher et al., 2017), which characterize the heterogeneity of land cover, soil, and terrain features at the field scale. Second, machine learning and supercomputers have been increasingly used to model complex interactions between water content and fluxes with environmental variables (Lu et al., 2015 and 2017; Adeyemi et al., 2018; Chaney et al., 2018; Prasad et al., 2018). Therefore, it is possible to combine these remote sensing and land surface datasets for improved delineation of soil water variability at the field scale.

Though earlier attempts have made successes on mapping surface soil moisture (SSM) at finer resolutions (i.e. 500 m to 1 km) using empirical and mechanistic models with European Space Agency Sentinel-1 (ESA-Sentinel-1) and/or National Aeronautics and Space Administration - Soil Moisture Active Passive (NASA-SMAP) Mission data with different spatial and temporal resolutions (Lievens et al., 2017; Bauer-Marschallinger et al., 2018; Das et al., 2019; Guevara and Vargas, 2019; Reichle et al., 2019), one unresolved research question remains: how much further can we improve the spatial and temporal resolutions of the models to characterize the heterogeneity of SSM at scales relevant for management of food and water resources?

To answer the question, this paper will focus on two objectives: 1) to develop an empirical machine learning model that is able to retrieve and map SSM across the globe at 100-m every 6-12 days over 4 years (2016–2019) by synergistic fusion of remote sensing data from Sentinel-1 and SMAP with land surface parameters via a machine learning algorithm (quantile random forest); 2) to apply the SSM model to selected irrigated cropland and rainfed grassland in the semi-arid region of Australia to demonstrate the potential application of the high-resolution machine learning based SSM maps for irrigation management. Our working hypothesis is a combination of remote sensing and land surface parameters data will improve the model performance of SSM retrieval at the field scale (i.e. 100 m).

2. Materials and Methods

2.1. Remote Sensing and Land Surface Data

Ground SSM measurements from various soil moisture networks were used as training and validation data for the remote sensing and land surface data that were used as covariates to retrieve SSM across the globe. Both remote sensing and land surface datasets were spatially explicit with the remote sensing data time-varying and the land surface datasets time-constant (Figure 2 and Table 1).

Table 1 (near here)

2.2. NASA SMAP Mission

The SMAP mission was launched by the NASA, which provides land surface measurements across the globe with a revisit time of 2-3 days. It relies on the simultaneous measurements of L-band backscatter from an active synthetic-aperture radar (SAR) and brightness temperature from a passive L-band radiometer to retrieve SSM (Lievens et al., 2017). The sensors operate at a constant incidence angle. The use of L-band microwave signals enables detection of land surface moisture under moderate vegetation cover, through cloud cover, and during day and night. Since the failure of the radar in 2015, the SMAP mission can only retrieve SSM based on the passive radiometer. In this study, the SMAP_L3_SM_P product was used, which retrieves SSM at 0–

0.05 m with a resampled spatial resolution of 36 km × 36 km and a revisit time of 2-3 days across the globe based on a physical model using the brightness temperature and other ancillary datasets (O'Neill et al., 2015).

SMAP data were downloaded from Earth Data (https://earthdata.nasa.gov/) using the R platform (Version 3.6.0) with the package "smapr" (Version 0.2.1) from March 1st to October 1st between 2016 and 2019. The period was selected to avoid frozen soils within the various soil moisture networks because of the poor performance of SSM retrieval over frozen ground. Afterward, the data were gap-filled pixel-wise using a simple temporal moving average with a window size of 3 days using the "imputeTS" package. The small window size was selected to avoid smoothing of SSM due to its strong variability over time. This generated SSM estimates at a 36 km × 36 km resolution on a daily basis during the study period, which were used as time-varying covariates for modeling SSM.

2.3. ESA Sentinel-1 mission

Sentinel-1 mission was launched by the European Space Agency (ESA), which consists of C-band SARs situated at a two-satellite constellation operating at dual polarizations: single co-polarization with vertical transmit/vertical receive (abbreviated as VV) and dual-band cross-polarization with vertical transmit/horizontal receive (abbreviated as VH). It measures the land surface backscatter intensity at VV and VH polarizations with a varying incidence angle with a spatial resolution of 5 m × 20 m and a revisit time of 6–12 days. The use of a C-band microwave signal leads to a reduced penetration depth of Sentinel-1's sensors under moderate vegetation cover compared to SMAP. The relationship between SAR backscatter and the dielectric constant of the soil (a function of soil moisture) enables retrieval of SSM from the Sentinel-1 data. Because the empirical model of the Sentinel-1 mission only retrieves relative SSM instead of soil volumetric water content (Bauer-Marschallinger et al., 2018), and because the physical retrieval model is currently under development (Lievens et al., 2017), the backscatter and incidence angle data were selected as covariates. Here, classical physical models (e.g. Oh et al., 1992; Fung, 1994; Dubois et al., 1995) were not used to retrieve SSM from the Sentinel-1 data because researchers have reported poor performance of the physical models when SSM is large and land surface roughness is high (Merzouki et al., 2011; Lievens et al,. 2017).

The backscatter data were preprocessed using the Sentinel-1 Toolbox (https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1) within the Google Earth Engine platform (https://developers.google.com/earth-engine/sentinel1), which involves thermal noise removal, radiometric calibration, and terrain correction using Shuttle Radar Topography Mission (SRTM) 30-m digital elevation model (Rabus et al., 2003). To minimize the speckle effects of the resampled Sentinel-1 radar data (Gao et al., 2017), additional preprocessing procedures were applied using the Google Earth Engine platform (Gorelick et al., 2017). This was suggested by Bauer-Marschallinger et al. (2018) and involved dynamic masking the extreme backscatter values outside the normal ranges for VV (–5 to –25 dB) and VH (–10 to –30 dB), spatial aggregating to 100 m × 100 m, and filtering with a 3 × 3 Gaussian filter. The processed Sentinel-1 data included backscatter data and incidence angle values at a 100 m × 100 m resolution with a revisit time of 6–12 days, which were used as time-varying covariates for modeling SSM.

To facilitate the retrieving of SSM from Sentinel-1 data, a number of temporal indices were calculated from the processed Sentinel-1 backscatter images pixel-wise to account for the land surface characteristics, such as temporal minimum, mean, maximum, and standard deviation (SD) of the backscatter data. These temporal statistics of the sensor measurements over time contain characteristics of the soil and vegetation in the field (Huang et al., 2019) and were used as time-constant covariates for modeling SSM.

2.4. Terrain Parameters

In addition to remote sensing datasets that can be directed used to retrieve SSM, terrain parameters that characterize topography characteristics have been used to indirectly model or downscale SSM (Entekhabi et al., 2010; Guevara and Vargas, 2019). In this study, a 500-m aggregated version of the Digital Elevation Model (DEM) from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) was used to calculate a number of primary and secondary terrain parameters (Olaya, 2009), including slope, aspect, terrain position index (TPI), and terrain ruggedness index (TRI) using the "terrain" function from the R package "raster" (Hijmans et al., 2015). Because SMAP SSM data had a coarser resolution (36 km), a finer-resolution (e.g. 30–250 m) DEM was not used. Elevation data were not used for SSM modeling process because of the insufficient long-term SSM stations at high elevation across the world (e.g. Tibetan Plateau). In addition, topographic wetness index was not calculated because it was strongly correlated to TPI at the 500-m resolution.

2.5. Soil Properties

Soil physical and chemical properties affect soil water retention and redistribution in space and time (Mohanty and Skaggs, 2001). Although finer-resolution maps of soil properties are available in many countries where the soil moisture networks are installed (e.g. Grundy et al., 2015; Ramcharan et al., 2018; Chaney et al., 2019), a consistent global map of soil properties was preferred for SSM modeling. In this study, 250-m resolution maps of soil properties were used, which include clay and sand content, bulk density (BD), soil organic carbon content from the SoilGrids (Hengl et al., 2017), and newly mapped field capacity and permanent wilting point (Hengl and Gupta, 2019).

2.6. Land Cover

Land surface characteristics vary with different LC types and had different impacts on the spatial and temporal variations of SSM and the performance of SSM models (Entekhabi et al., 2010). To facilitate the interpretation of the SSM models, 500-m annual land cover (LC) data were downloaded during 2016 from the MODIS repository (MCD12Q1.006, available at https://lpdaac.usgs.gov/products/mcd12q1v006/). The International Geosphere-Biosphere Programme (IGBP) classification was used, among which were six merged LC types were selected, including cropland, grassland, savanna, shrubland, forest, and barren. These LC types were not used as covariates for retrieving SSM (due to the coarse resolution) but were used to evaluate the performance of the SSM models under different LC types.

2.7. Soil Moisture Monitoring Networks

Soil moisture networks have been established across the world to provide long-term climate reference measurements for meteorological monitoring, hydrological modeling, and validating of remote sensing products (Dorigo et al., 2011; Quiring et al., 2016). Here, two types of soil moisture networks were used: regional-scale and continental-scale networks (Figure 1). A summary of the number of stations used in this study is provided in Table 2. Details about the site characterization of these networks can be found in the references mentioned above.

Table 2 (near here)

Regional-scale soil moisture networks consist of the Murrumbidgee soil moisture monitoring networks of the OZNET in New South Wales, Australia (Smith et al., 2012), Soil Moisture Measurement Stations Network of the University of Salamanca, Spain (REMEDHUS) (Martínez-Fernández and Ceballos, 2005), and the Danish hydrological observatory (HOBE) in Denmark (Jensen and Illangasekare, 2011). These networks were selected because they were located at the regional scale (< 50,000 km^2) and can be used to characterize variations of surface soil moisture within catchments, and span a variety of soil moisture and climatic regimes each with significant spatial variability. It was expected that soil moisture measurements from these regional-scale networks can provide detailed information for retrieving SSM within the coarse pixels of the SMAP SSM product (36 km).

Continental-scale soil moisture networks consist of National Oceanic and Atmospheric Administration sponsored US Climate Reference Network (USCRN, Janis and Center, 2002) and the United States Department of Agriculture Natural Resources Conservation Service soil climate analysis network (SCAN, Schaefer et al., 2007). These networks are sparsely situated across the USA with several stations within each state, but they cover a variety of climate regimes, terrain parameters, land cover types, and soil texture classes. It was expected that the use of these widely spread networks can provide information on the relationship between climate regimes, terrain parameters, land cover types, and soil texture classes with SSM and improve the robustness of the SSM model.

2.8. Establishing Empirical SSM Retrieval Models

Random Forest is a nonparametric model based on similarities among observations to fit decision trees. To determine a split at a node in a tree, a random subsample of predictor variables is taken to select the predictor that minimizes the regression error. Nodes continue to be split until no further improvement in error is achieved. The prediction is achieved with an adaptive neighborhood classification and regression. Omitted observations, termed the “out-of-bag” sample, are used to compute the regression errors for trees (Breiman, 2001; Hastie et al., 2009). To estimate the quantiles of the predictions, the Quantile Random Forest (QRF) algorithm was applied using the R ‘quantregForest’ package (Version 1.3-7, Meinshausen and Meinshausen, 2017), which estimates the conditional distribution based on a weighted distribution of observed model response values (Meinshausen, 2006).

To train the QRF model and evaluate the model performance, SSM measurements from the various soil moisture networks were randomly split into training and validation datasets. To maximally represent the heterogeneous land surface conditions and variations of SSM, 75% of the measuring stations from the regional-scale (OZNET, REMEDHUS, HOBE) and continental-scale (USCRN, SCAN) networks were randomly selected as the training dataset and the remaining 25% of the stations from these four networks were used as the validation dataset (Figure 1). The coefficient of determination (R^2), mean error (bias), and root mean squared error (accuracy) were calculated for both calibration and validation datasets using the measured SSM at the stations and predicted SSM from the QRF models. The 5% and 95% quantiles of the predictions were also calculated to present the confidence of the SSM prediction.

Fig. 1–2 (near here)

2.9. Prediction of SSM on an Irrigated Farm in Australia

SSM was predicted at the field scale in Australia during the 2018 growing season. The flowchart of the algorithm is presented in Figure 2. The study area was located in the Yanco site of the OZNET, New South Wales, Australia. The annual precipitation was approximately 402 mm with annual minimum and maximum temperatures of 11.5 and 24.2 ℃. Furrow irrigation is often used over the growing season every one to two weeks. SSM was retrieved and mapped using the established QRF model across a number of irrigated fields and rainfed (totally 13,822 ha in size) on selected days during the early season from November to December 2018. This period was selected due to a reported drought in the region (BBC, 2018; BOM, 2018).

To demonstrate the usefulness of the high-resolution QRF model and evaluate the impacts of water stress on plant productivity, MODIS satellite-based 500-m 8-day cumulative gross primary productivity (GPP, g C m^{-2} per 8 days, https://lpdaac.usgs.gov/products/mod17a2hv006/) and evapotranspiration (ET, mm H_2O per 8 days, https://lpdaac.usgs.gov/products/mod16a2v006/) were downloaded across the study fields. The 8-day cumulative GPP and ET data were temporally interpolated with the centers of the 8-day periods matched with the dates of the SSM maps. Water use efficiency was calculated across the fields as the ratio of GPP to ET (g Carbon per mm H_2O).

3. Results

3.1. Model Performance of the QRF and SMAP Product

The importance of the variables is presented in Figure 3. SMAP was ranked as the most important time-varying variables, followed by Sentinel-1 backscatter data measured at VV and VH polarizations, and incidence angle of the Sentinel-1. In terms of the time-constant variables, sand content was most important, followed by the temporal mean of VH backscatter data, topographic ruggedness index, topographic position index, aspect, clay content, and other variables.

The model performance of the fitted QRF model is also presented in Figure 3. The model has an RMSE of 0.02 $m^3\ m^{-3}$ and R^2 of 0.95 for the training dataset and a reduced performance with an RMSE of 0.08 $m^3\ m^{-3}$ and R^2 of 0.53 for the validation dataset. The SSM estimates from SMAP had an overall similar performance (no significant difference) for the same validation dataset with an RMSE of 0.08 $m^3\ m^{-3}$ and R^2 of 0.50.

Table 3 and Fig. 3–4 (near here)

3.2. Model Performance of the QRF and SMAP Product within Different Land Cover Types

Pearson's correlation coefficient (r), mean error (ME), and root mean squared error (RMSE) calculated between measured SSM from the station soil moisture networks and predicted SSM from the QRF model or SMAP were used to evaluate the model performance within different land cover types. When all networks were considered (Table 3), the empirical QRF model established via combination of SMAP, Sentinel-1 and land surface parameters outperformed the SMAP SSM estimates for cropland (r = 0.73 vs. 0.64, RMSE = 0.08 vs. 0.10 $m^3\ m^{-3}$) and savanna (r = 0.73 vs. 0.54, RMSE = 0.09 vs. 0.11 $m^3\ m^{-3}$). Both QRF and SMAP models had similar performance under barren (r = 0.77 vs. 0.77, RMSE = 0.05 vs. 0.06 $m^3\ m^{-3}$) and forest (r = 0.58 vs. 0.63, RMSE = 0.09 vs. 0.08 $m^3\ m^{-3}$) soils. However, the QRF model was worse than the SMAP in grassland (r = 0.63 vs. 0.67, RMSE = 0.07 vs. 0.07 $m^3\ m^{-3}$) and shrubland soils (r = 0.22 vs. 0.63, RMSE = 0.07 vs. 0.05 $m^3\ m^{-3}$).

Similar patterns were observed for each network within different land cover types (Table 3) and for the temporal dynamics of measured and estimated SSM at several selected validation stations (Figure 4). It was evident that the QRF model was more accurate than SMAP under cropland (OZNET – Uri_Park, r = 0.81 vs. 0.79, RMSE = 0.06 vs. 0.14 $m^3 m^{-3}$) and Savanna (HOBE – 3.06, r = 0.55 vs. 0.56, RMSE = 0.06 vs. 0.12 $m^3 m^{-3}$), similar to SMAP under barren (SCAN – Lovelock NNR, r = 0.70 vs. 0.67, RMSE = 0.05 vs. 0.06 $m^3 m^{-3}$) and forest (SCAN – Reynolds Homestead, r = 0.60 vs. 0.57, RMSE = 0.08 vs. 0.15 $m^3 m^{-3}$), and worse than SMAP under grassland (REMEDHUS – Las_Arenas, r = 0.82 vs. 0.79, RMSE = 0.09 vs. 0.07 $m^3 m^{-3}$) and shrubland (USCRN – CA_Fallbrook_5_NE, r = 0.05 vs. 0.04).

It should also be noted that large variations of model performance (i.e. Pearson's r, ME, RMSE) were observed for all land cover types among different ground SSM stations, indicating strong heterogeneity of land surface parameters at the field scale. In summary, we note that the QRF model was able to successfully retrieve SSM dynamics under cropland (r = 0.73, RMSE = 0.08 $m^3 m^{-3}$), grassland (r = 0.63, RMSE = 0.07 $m^3 m^{-3}$), savanna (r = 0.73, RMSE = 0.09 $m^3 m^{-3}$), forest (r = 0.58, RMSE = 0.09 $m^3 m^{-3}$), and barren (r = 0.77, RMSE = 0.05 $m^3 m^{-3}$) soils.

3.3. Delineating SSM Variations at the Field scale via Data Fusion

Coarse-resolution SMAP SSM maps can not be used to reveal spatial SSM variations at the field scale compared to the data fusion based QRF model (Figure 5). In the selected fields within the OZNET network in Australia, SSM was retrieved and mapped using the QRD from November 9th to December 15th, 2018. During this early cropping season, SSM varied greatly over time (0.06–0.18 $m^3 m^{-3}$). Instead of showing uniform values for the whole region at different days from the 36-km SMAP model (0.19, 0.14, 0.07, and 0.18 $m^3 m^{-3}$), the mean SSM values of the QRF maps displayed strong heterogeneity in space with ranges of SSM of 0.06, 0.12, and 0.06 $m^3 m^{-3}$ under the dry (December 3rd), intermediate (transitional) (November 21st, December 15th), and wet (November 9th) conditions, respectively.

MODIS-estimated 8-day cumulative GPP and ET also displayed strong variations in space and over the study period (15–27 g C m^{-2} per 8 days and 400–1,200 mm H_2O per 8 days) (Figure 5). Note that GPP and ET values were higher in the northern parts of the region associated with the irrigated crops and lower in the southern parts of the region associated with the rainfed grassland.

Three sites were selected across the region, including two irrigated cropland fields (1 and 2) and one rainfed grassland. As shown in Figures 5 and 7, irrigated cropland 1 had a higher SSM on November 9th (0.18 $m^3 m^{-3}$) and other days than irrigated cropland 2 (0.17 $m^3 m^{-3}$) and rainfed grassland (0.15 m^3 m^{-3}). This was consistent with the higher GPP and ET values observed (e.g. November 9th) for irrigated cropland 1 (41 g C m^{-2} per 8 days and 1,760 mm H_2O per 8 days), cropland 2 (27 g C m^{-2} per 8 days and 1,000 mm H_2O per 8 days) and rainfed grassland (16 g C m^{-2} per 8 days and 670 mm H_2O per 8 days). Also note that different water use efficiency values were observed between the two irrigated cropland sites and the rainfed grassland.

Fig. 5–7 (near here)

4. Discussion

4.1. Model Performance under Different Land Cover Types

In terms of the model performance, the empirical data fusion based SSM model is superior or similar to the 36-km SMAP_L3 SSM product under many land cover types except for grassland and shrubland. Improved model accuracy under cropland and savanna is most likely due to the use of high-resolution (5–20 m) Sentinel-1 data compared to SMAP (~ 36 km), which characterize field-scale variations in SSM (Figures 5–7). However, due to the use of a C-band microwave signal, it is expected that the Sentinel-1's radar has a reduced penetration depth as compared to the L-band SMAP passive microwave radiometer under moderate to dense vegetation cover conditions (Lievens et al., 2017). This helps explain the poor performance of the SSM model under shrubland. In terms of grassland, the slightly worse performance of the SSM model can also be due to the grazing or harvesting practices that change the vegetation characteristics (e.g. leaf area).

4.2. A tradeoff between Spatial and Temporal Resolutions of Remote Sensing Soil Moisture Products

Current remote sensing soil moisture missions operating at the global scale are based on reflectance in the optical band (e.g. MODIS), passive microwave (e.g. SMOS, SMAP, SMAR2, ASCAT), active microwave (e.g. Sentinel-1, RADARSAT-2) and gravity (GRACE). As shown in Figure 8 and summarized by others (Robinson et al., 2008; Vereecken et al., 2014; Wang et al., 2009; Ochsner et al., 2013), there is a tradeoff between the spatial and temporal resolutions of these satellites. In general, optical and active microwave satellites have a fine spatial resolution less than 1 km but the temporal resolution (revisit time) is more than one week. By contrast, passive microwave satellites have a coarse spatial resolution larger than 10 km but the temporal resolution is higher (1-3 days). The gravity-based mission (GRACE) measures soil water in the deep profile, and has a large spatial resolution (> 100 km) with a temporal resolution of approximately one month.

The SSM model established here has a spatial resolution of 100 m and revisit time of 6-12 days (depending on the location of the study sites) across the globe. Many researchers have attempted to retrieve SSM at a similar (100 m) or higher (30 m) spatial resolution using Sentinel-1 data at the field scale using a larger number of ground-based SSM measurements (e.g. Alexakis et al., 2017; Gao et al., 2017; Attarzadeh et al., 2018). However, the model performance deteriorates with increasing spatial resolution. Based on the work of Bauer-Marschallinger et al. (2018) and others, the radar signal has a large noise (speckle effect) at the field scale due to the interference with heterogeneous vegetation, terrain surface, and soil properties. Upscaling Sentinel-1 data to a larger spatial resolution (e.g. 500 m) is required to reduce the sensor's noise. As such, the SSM established here may not be transformed to a finer resolution without reducing the model performance. To delineate the SSM variations at such a fine resolution (e.g. plot scale, Figure 8), soil core sampling (Li et al., 2019) or ground-based proximal soil sensors (Robinson et al., 2008; Striegl amd Loheide, 2012) should be used instead.

Fig. 8 (near here)

4.3. Irrigation Management at the Field Scale via Data Fusion of Remote Sensing and Land Surface Parameters Data

Compared to traditional in situ soil moisture sensors that are installed on the farm to monitor SSM at limited individual stations or SMAP_L3 SSM products (e.g. radiometer) that rely on space-borne sensors to monitor SSM with a very coarse resolution (36 km), the retrieved SSM maps can delineate field-scale variations in SSM (Figures 5 and 7), which can be potentially used for monitoring SSM and irrigation scheduling at the field scale. The maps of SSM identify regions with a high irrigation priority and the pixel resolution (100 m × 100 m: 1 ha) is suitable for irrigation management at the farm scale, whereby furrow irrigation is often used in this region to supply water on a field by field basis.

In terms of the temporal resolution, the rate-limiting factor of the current SSM model is the Sentinel-1 data, which are currently available 6–12 days globally (depending on the region of interest). Future research is required to gap-fill the SSM maps within the 6–12 days to obtain close to real-time SSM maps. This could be realized using space-time statistical method (Jost et al., 2005) or mechanistic models (Or and Lehmann, 2019).

In addition, future research is required to map soil water content below the surface, particularly within the root zone, as soil water content often varies greatly with depth within the soil profile. This can be potentially achieved by data assimilation of the empirical machine learning SSM model with a mechanistic water balance model (e.g. Das and Mohanty, 2006; Huang et al., 2017). Alternatively, to calculate soil water stored within the root zone, empirical and analytical models can be established based on the retrieved SSM maps over a long-term period (Arya et al., 1983; Jackson et al., 1987; Wagner et al., 1999; Gouweleeuw, 2000; Jackson, 2002; Ceballos et al., 2005; Albergel et al. 2008; Sadeghi et al., 2019a,b).

4.4. Developing an Additive and Adaptable SSM Model via Machine Learning

Only small numbers of the validation stations were available for certain land cover types (e.g. grassland in OZNET and REMEDHUS, and shrubland and forest in all networks). This could contribute to the moderate performance of QRF model under these land cover types because machine learning algorithms often require a large number of training dataset to capture the variations in the model response (i.e. SSM) and the feature space (i.e. environmental covariates). The accuracy of the SSM model also needs to be further improved in cropped areas where accurate characterization of soil water conditions is crucial for sustaining crop yield and maximizing water use efficiency.

Additional SSM measurements from vegetation-specific (e.g. cotton, olives, vegetables, fruits) ground soil moisture networks should be collected to provide dataset covering these feature spaces to improve the empirical QRF model. This is equivalent to the “spiking” techniques used to calibrate the global soil visible near-infrared spectroscopy library using local spectra data (Guerrero et al., 2010; Wetterlind and Stenberg, 2010; Viscarra Rossel et al., 2016). In this regard, the empirical data fusion-based QRF model established here is an additive and adaptable model and can be improved with addition of localized SSM measurements from in situ soil moisture networks in the future.

Fig. 9 (near here)

5. Conclusions

An empirical surface soil moisture (SSM) model was established via data fusion of remote sensing data (Sentinel-1 and SMAP) and land surface parameters (e.g. soil texture, terrain parameters) using quantile random forest (QRF) algorithm. The model had a spatial resolution of 100 m and performed moderately well (R^2 = 0.53, RMSE = 0.08 m^3 m^{-3}) across the globe under cropland, grassland, savanna, barren, and forest soils. Particularly, the empirical QRF model performed better than the 36-km SMAP SSM model under cropland and savanna soils.

SSM was retrieved and mapped at 100 m every 6-12 days during the plant growing seasons in 2018 in selected cropland and grassland fields in the OZNET network, Australia. It was concluded that the high-resolution SSM maps can be used to monitor soil water content at the field scale for irrigation management. The SSM model is an additive and adaptable model, which can be further improved by including soil moisture measurements at the field scale for specific vegetation/crop types. Further research is required to improve the temporal resolution of the SSM model and map soil water content within the root zone.

Acknowledgments

We acknowledge the USDA Hatch project for providing support for the project. We acknowledge HOBE, OZNET, REMEDHUS, SCAN, USCRN, National Soil Moisture Network, and International Soil Moisture Network for providing the surface soil moisture measurements. We acknowledge SoilGrids & Openlandmap, NASA, and US Geological Survey for providing maps of land surface parameters. We also acknowledge ESA and NASA for providing land surface and soil moisture products. The R codes for the quantile random forest model, training datasets, and ancillary data for soil moisture mapping in Australia are available at the Zenodo repository (DOI: 10.5281/zenodo.3659192).

References

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., ... & Wagner, W. (2018). Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 520-539. https://doi.org/10.1109/TGRS.2018.2858004

BBC, 2018. New South Wales drought now affects entire state. Accessed from https://www.bbc.com/news/world-australia-45107504

Bigiarini, M. Z., & Bigiarini, M. M. Z. (2013). Package "hydroGOF". R-package, available at: www. r-project.org/(last access: 7 Jan 2020).

Bureau of Meteorology, 2018. Map of root-zone soil moisture for the previous month. Accessed from http://www.bom.gov.au/climate/drought/archive/20181205.archive.shtml#tabs2=Soil-moisture

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., ... & Walker, J. P. (2019). The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sensing of Environment*, 233, 111380. https://doi.org/10.1016/j.rse.2019.111380

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., ... & Robock, A. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, 15(5), 1675-1698. https://doi.org/10.5194/hess-15-1675-2011

Dubois, P. C., Van Zyl, J., & Engman, T. (1995). Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), 915-926.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Kimball, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704-716.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., ... & Kilic, A. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, 53(4), 2618-2626. https://doi.org/10.1002/2016WR020175

Fung, A. (1994), Microwave Scattering and Emission Models and Their Applications, Artech House, Boston, Mass.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, 114(1), 168-182.

https://doi.org/10.1016/j.rse.2009.08.016

Gao, Q., Zribi, M., Escorihuela, M., & Baghdadi, N. (2017). Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. *Sensors*, 17(9), 1966. https://doi.org/10.3390/s17091966

Grundy, M. J., Rossel, R. V., Searle, R. D., Wilson, P. L., Chen, C., & Gregory, L. J. (2015). Soil and landscape grid of Australia. *Soil Research*, 53(8), 835-844. https://doi.org/10.1071/SR15191

Guerrero, C., Zornoza, R., Gómez, I., & Mataix-Beneyto, J. (2010). Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. *Geoderma*, 158(1-2), 66-77. https://doi.org/10.1016/j.geoderma.2009.12.021

Guevara, M., & Vargas, R. (2019). Downscaling satellite soil moisture using geomorphometry and machine learning. *PloS one*, 14(9), e0219639. https://doi.org/10.1371/journal.pone.0219639

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2). https://doi.org/10.1371/journal.pone.0169748

Hengl, T., & Gupta, S. (2019). Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v01) [Data set]. Zenodo. 10.5281/zenodo.2629589

Janis, M. J., & Center, S. R. C. (2002). US Climate Reference Network (USCRN) Site Identification, Survey, and Selection FY 02 Research Project for The NOAA Regional Climate Centers (RCC).

Jensen, K. H., & Illangasekare, T. H. (2011). HOBE: A hydrological observatory. *Vadose Zone Journal*, 10(1), 1-7. https://doi.org/10.2136/vzj2011.0006

Jost, G., Heuvelink, G. B. M., & Papritz, A. (2005). Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. *Geoderma*, 128(3-4), 258-273. https://doi.org/10.1016/j.geoderma.2005.04.008

Li, X., Shao, M., Zhao, C., Liu, T., Jia, X., & Ma, C. (2019). Regional spatial variability of root-zone soil moisture in arid regions and the driving factors - A case study of Xinjiang, China. *Canadian Journal of Soil Science*, 99(3), 277-291. https://doi.org/10.1139/cjss-2019-0006

Lievens, H., Reichle, R. H., Liu, Q., De Lannoy, G. J. M., Dunbar, R. S., Kim, S. B., ... &

Wagner, W. (2017). Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophysical Research Letters*, 44(12), 6145-6153. https://doi.org/10.1002/2017GL073904

Martínez-Fernández, J., & Ceballos, A. (2005). Mean soil moisture estimation using temporal stability analysis. Journal of Hydrology, 312(1-4), 28-38. https://doi.org/10.1016/j.jhydrol.2005.02.007

Meinshausen, N. (2006). Quantile regression forests. Journal of Machine Learning Research, 7(Jun), 983-999. Available on http://www.jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf

Meinshausen, N., & Meinshausen, M. N. (2017). Package 'quantregForest', version 1.3-7.

Merzouki, A., McNairn, H., & Pacheco, A. (2011). Mapping soil moisture using RADARSAT-2 data and local autocorrelation statistics. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1), 128-137. https://doi.org/10.1109/JSTARS.2011.2116769

Oh, Y., Sarabandi, K., & Ulaby, F. T. (1992). An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE transactions on Geoscience and Remote Sensing*, 30(2), 370-381. https://doi.org/10.1109/36.134086

Olaya, V. (2009). Basic land-surface parameters. Developments in Soil Science, 33, 141-169. https://doi.org/10.1016/S0166-2481(08)00006-8

O'Neill, P. E., Njoku, E. G., Jackson, T. J., Chan, S., & Bindlish, R. (2015). SMAP algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products. Jet Propulsion Lab., California Inst. Technol., Pasadena, CA, USA, JPL D-66480.

Or, D., & Lehmann, P. (2019). Surface evaporative capacitance: How soil type and rainfall characteristics affect global-scale surface evaporation. *Water Resources Research*, 55(1), 519-539. https://doi.org/10.1029/2018WR024050

Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. *The American Naturalist*, 164(5), 625-632. https://doi.org/10.1086/424970

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. A., Draper, C. S., Hagimoto, Y., ... & Larson, K. M. (2013). State of the art in large-scale soil moisture monitoring. *Soil Science Society of America Journal*, 77(6), 1888-1919. https://doi.org/10.2136/sssaj2013.03.0093

Quiring, S. M., Ford, T. W., Wang, J. K., Khong, A., Harris, E., Lindgren, T., ... & Li, Z. (2016). The North American soil moisture database: Development and applications. *Bulletin of the American Meteorological Society*, 97(8), 1441-1459. https://doi.org/10.1175/BAMS-D-13-00263.1

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Science Society of America Journal*, 82(1), 186-201. https://doi.org/10.2136/sssaj2017.04.0122

Reuter, H. I., Nelson, A., & Jarvis, A. (2007). An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science*, 21(9), 983-1008. https://doi.org/10.1080/13658810601169899

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., ... & Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7(1), 358-389. https://doi.org/10.2136/vzj2007.0143

Sadeghi, M., Tuller, M., Warrick, A. W., Babaeian, E., Parajuli, K., Gohardoust, M. R., & Jones, S. B. (2019). An analytical model for estimation of land surface net water flux from near-surface soil moisture observations. *Journal of Hydrology*, 570, 26-37. https://doi.org/10.1016/j.jhydrol.2018.12.038

Sadeghi, M., Ebtehaj, A., Crow, W. T., Gao, L., Purdy, A. J., Fisher, J. B., ... & Tuller, M. (2019). Global Estimates of Land Surface Water Fluxes from SMOS and SMAP Satellite Soil Moisture Data. *Journal of Hydrometeorology*, (2019). https://doi.org/10.1175/JHM-D-19-0150.1

Schaefer, G. L., Cosh, M. H., & Jackson, T. J. (2007). The USDA natural resources conservation service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic Technology*, 24(12), 2073-2077. https://doi.org/10.1175/2007JTECHA930.1

Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., ... & Richter, H. (2012). The Murrumbidgee soil moisture monitoring network data set. *Water Resources Research*, 48(7). https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012WR011976

Stoy, P. C., El-Madany, T., Fisher, J. B., Gentine, P., Gerken, T., Good, S. P., ... & Wohlfahrt12,

G. (2019). Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. *Biogeosciences Discussions*, 10. https://doi.org/10.5194/bg-16-3747-2019

Striegl, A. M., & Loheide II, S. P. (2012). Heated distributed temperature sensing for field scale soil moisture monitoring. *Groundwater*, 50(3), 340-347. https://doi.org/10.1111/j.1745-6584.2012.00928.x

Trugman, A. T., Medvigy, D., Mankin, J. S., & Anderegg, W. R. L. (2018). Soil moisture stress as a major driver of carbon cycle uncertainty. *Geophysical Research Letters*, 45(13), 6495-6503. https://doi.org/10.1029/2018GL078131

Vereecken, H., Huisman, J. A., Pachepsky, Y., Montzka, C., Van Der Kruk, J., Bogena, H., ... & Vanderborght, J. (2014). On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology*, 516, 76-96. https://doi.org/10.1016/j.jhydrol.2013.11.061

Viscarra Rossel, R., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., ... & Aïchi, H. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198-230. https://doi.org/10.1016/j.earscirev.2016.01.012

Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China*, 3(2), 237-247. https://doi.org/10.1007/s11707-009-0023-7

Wei, J., & Dirmeyer, P. A. (2012). Dissecting soil moisture-precipitation coupling. *Geophysical Research Letters*, 39(19). https://doi.org/10.1029/2012GL053038

Wetterlind, J., & Stenberg, B. (2010). Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. *European Journal of Soil Science*, 61(6), 823-843. https://doi.org/10.1111/j.1365-2389.2010.01283.x

Figure Captions

Figure 1. Locations of regional-scale soil moisture monitoring networks HOBE (Denmark), OZNET (Australia), REMEDHUS (Spain), and continental-scale soil moisture networks SCAN and USCRN (USA). Note: Training and validation stations were highlighted in different colors.

Figure 2. Flowchart of the global surface soil water model established using data fusion and machine learning.

Figure 3. Variable importance of the quantile random forest (QRF) model and a comparison of model performance on training and validation datasets generated based on data fusion based QRF model and SMAP-L3 surface soil moisture (SSM) product.

Figure 4. Plots of measured surface soil moisture (SSM) (black lines) under various land cover types at several soil moisture stations and estimated SSM from SMAP (blue lines) and the data fusion based quantile random forest (QRF) model (red lines, with 5 and 95 percentiles marked in dashed lines). Note: no SSM estimates were made during October to February.

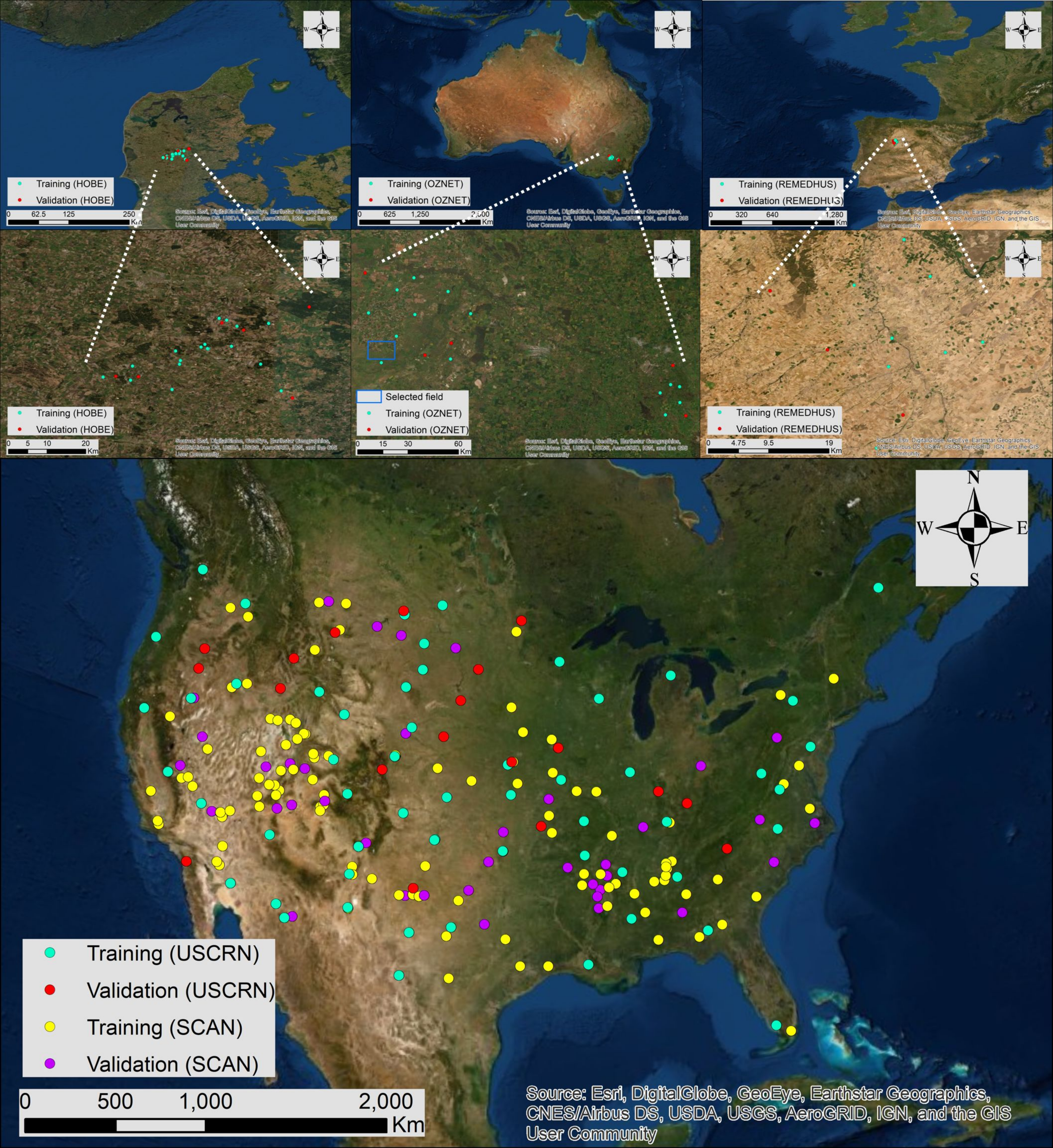
Figure 5. Predicted surface soil moisture (SSM, m^3 m^{-3}) from the quantile random forest (QRF) model during the 2018 cropping season across selected fields within OZNET network in New South Wales, Australia.

Figure 6. Maps of MODIS estimated cumulative gross primary productivity (GPP, g C m^{-2} per 8-day), evapotranspiration (ET, mm H_2O per 8-day), and water use efficiency (WUE, g C per mm H_2O) during the 2018 growing season across selected fields within OZNET network in New South Wales, Australia.

Figure 7. Plots of measured and predicted surface soil moisture (SSM, m^3 m^{-3}) from the quantile random forest (QRF) model and NASA-SMAP, MODIS cumulative gross primary productivity (GPP, g C m^{-2} per 8-day), evapotranspiration (ET, mm H_2O per 8-day), and water use efficiency (WUE, g C per mm H_2O) at three selected sites (irrigated cropland 1 and 2, rainfed grassland) during the 2018 growing season within OZNET network in New South Wales, Australia.

Figure 8. Spatial and temporal resolutions of current remote sensing soil moisture monitoring satellites. Note: satellites used in this study are highlighted in black and other satellites designed to monitor soil moisture are marked in grey.

Figure1.



Training (HOBE)
Validation (HOBE)
0 62.5 125 250 Km
Training (OZNET)
Validation (OZNET)
0 625 1,250 2,500 Km
Training (REMEDHUS)
Validation (REMEDHUS)
0 320 640 1,280 Km
Training (HOBE)
Validation (HOBE)
0 5 10 20 Km
Selected field
Training (OZNET)
Validation (OZNET)
0 15 30 60 Km
Training (REMEDHUS)
Validation (REMEDHUS)
0 4.75 9.5 19 Km
N
W
E
S
Training (USCRN)
Validation (USCRN)
Training (SCAN)
Validation (SCAN)
0 500 1,000 2,000 Km
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

Figure2.


Ground SSM measurements
Regional-scale SSM networks: HOBE, OZNET, REMEDHUS
Continental-scale SSM networks: SCAN, USCRN
Time-varying covariates
NASA SMAP (L3_SM_P):
SSM (36 km, 2-3 days)
ESA Sentinel-1 (VV & VH):
backscatter & incident angle (100 m, 6-12 days)
Time-constant covariates
ESA Sentinel-1 (VV & VH): temporal statistics (100 m)
GMTED2010 DEM: terrain parameters (250 m)
SoilGrids: clay, sand, BD, SOC, FC, PWP (250 m)
Training (75%):
Cropland, Grassland, Savanna, Shrubland, Forest, Barren
Validation (25%):
Cropland, Grassland, Savanna, Shrubland, Forest, Barren
Data fusion & machine learning: quantile random forest
Predicted mean and SD of SSM (100 m, 6-12 days)

Figure3.

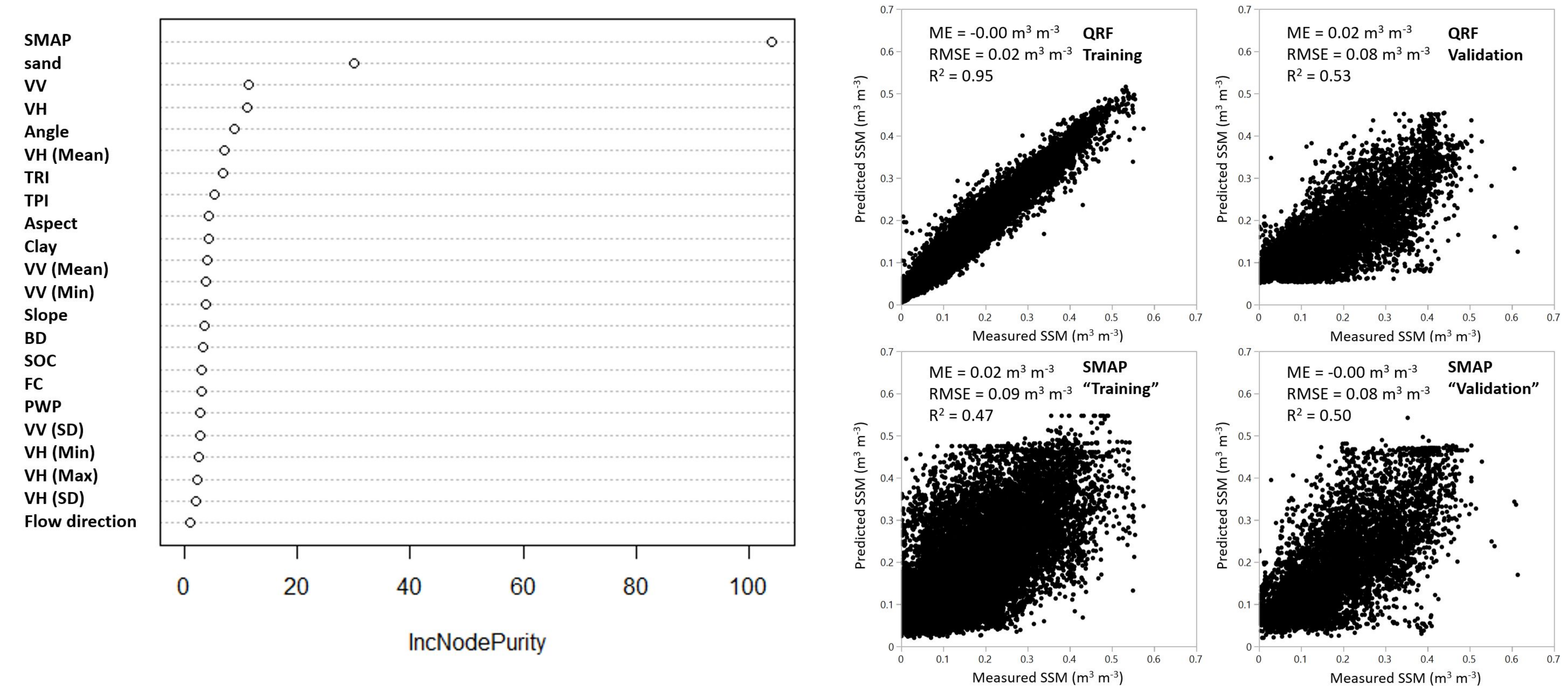
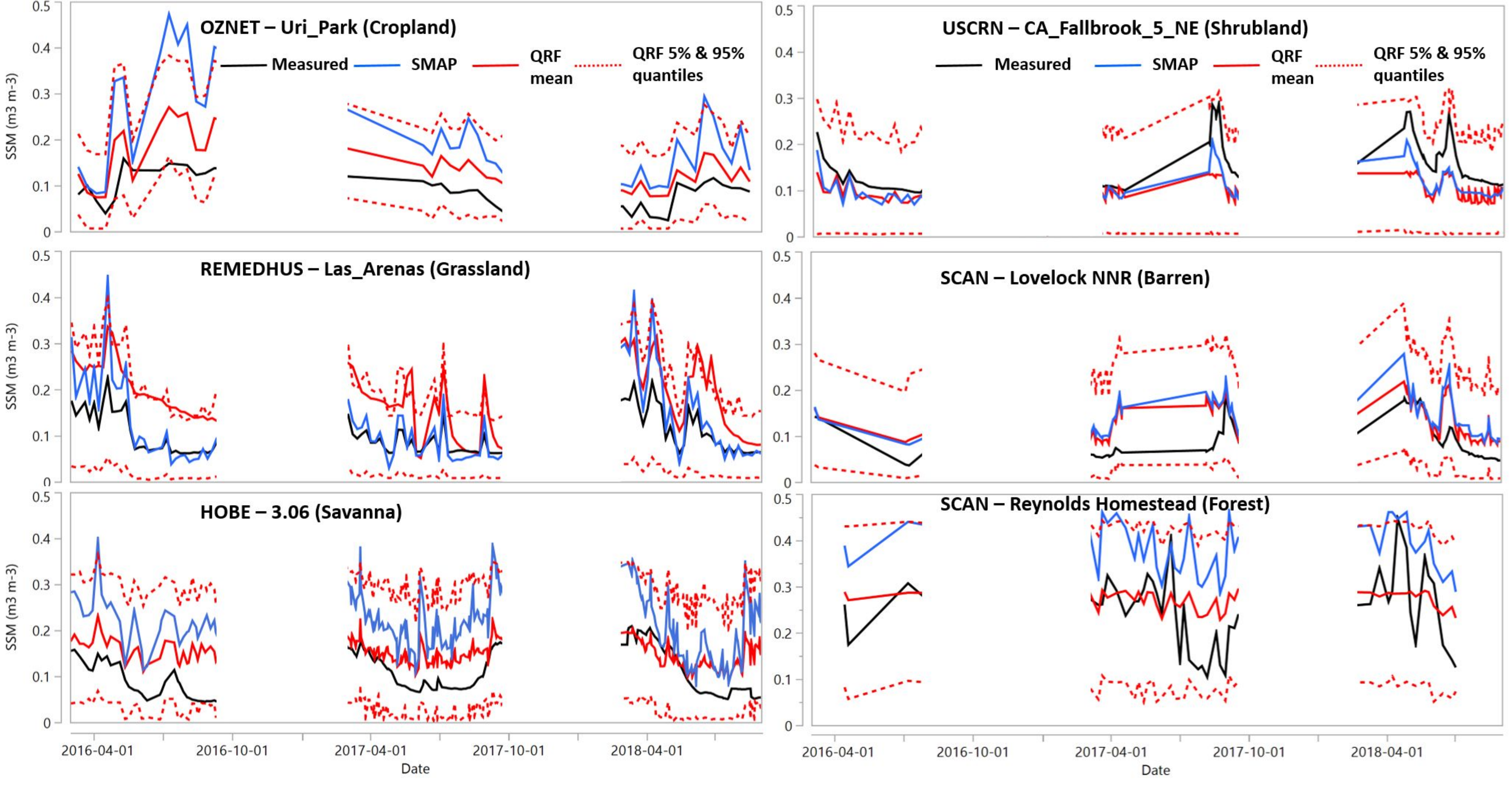
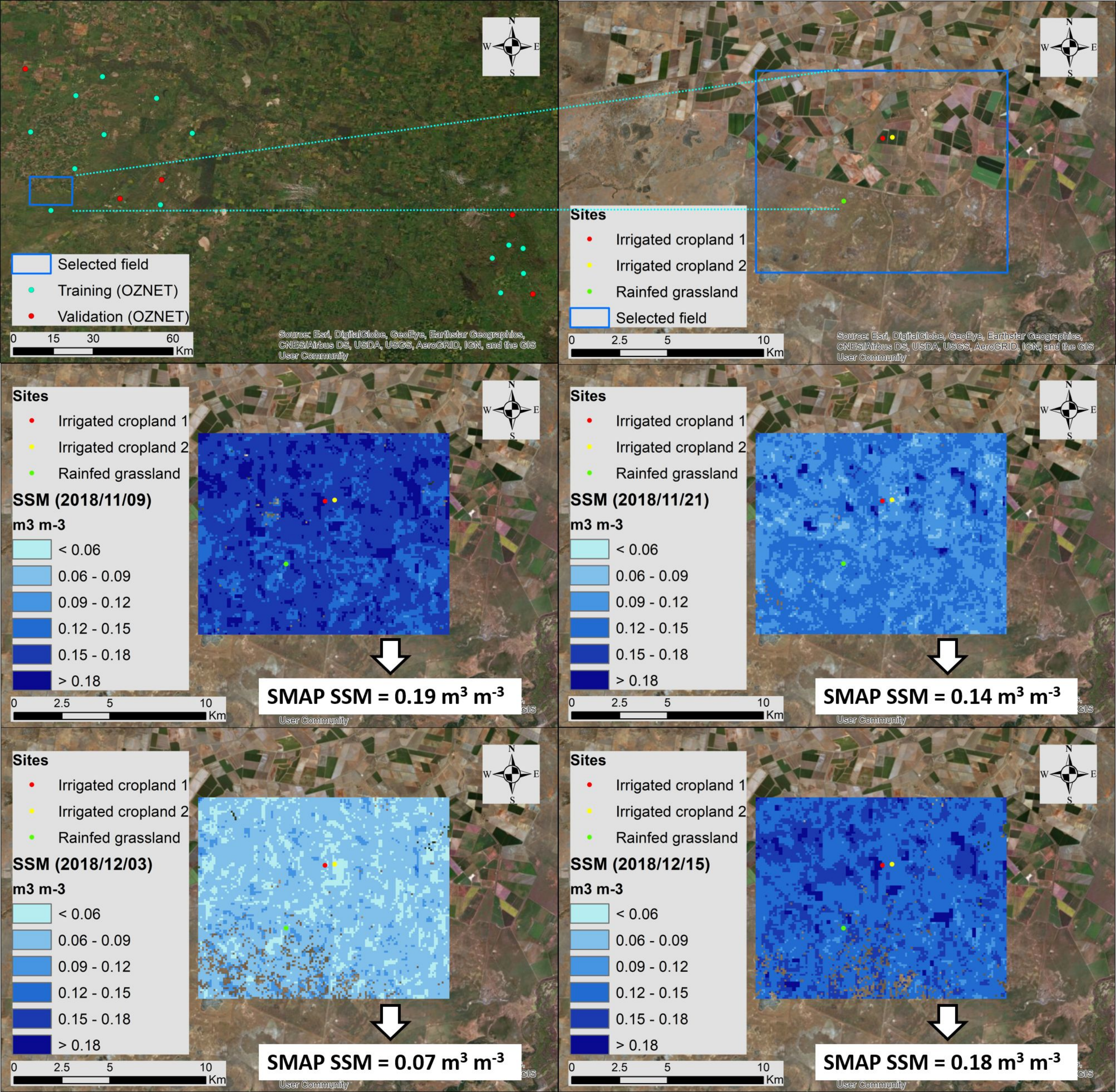

SMAP
sand
VV
VH
Angle
VH (Mean)
TRI
TPI
Aspect
Clay
VV (Mean)
VV (Min)
Slope
BD
SOC
FC
PWP
VV (SD)
VH (Min)
VH (Max)
VH (SD)
Flow direction
0
20
40
60
80
100
IncNodePurity
ME = -0.00 m3 m-3
RMSE = 0.02 m3 m-3
R2 = 0.95
QRF
Training
ME = 0.02 m3 m-3
RMSE = 0.08 m3 m-3
R2 = 0.53
QRF
Validation
ME = 0.02 m3 m-3
RMSE = 0.09 m3 m-3
R2 = 0.47
SMAP
"Training"
ME = -0.00 m3 m-3
RMSE = 0.08 m3 m-3
R2 = 0.50
SMAP
"Validation"
Predicted SSM (m3 m-3)
Measured SSM (m3 m-3)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Figure4.



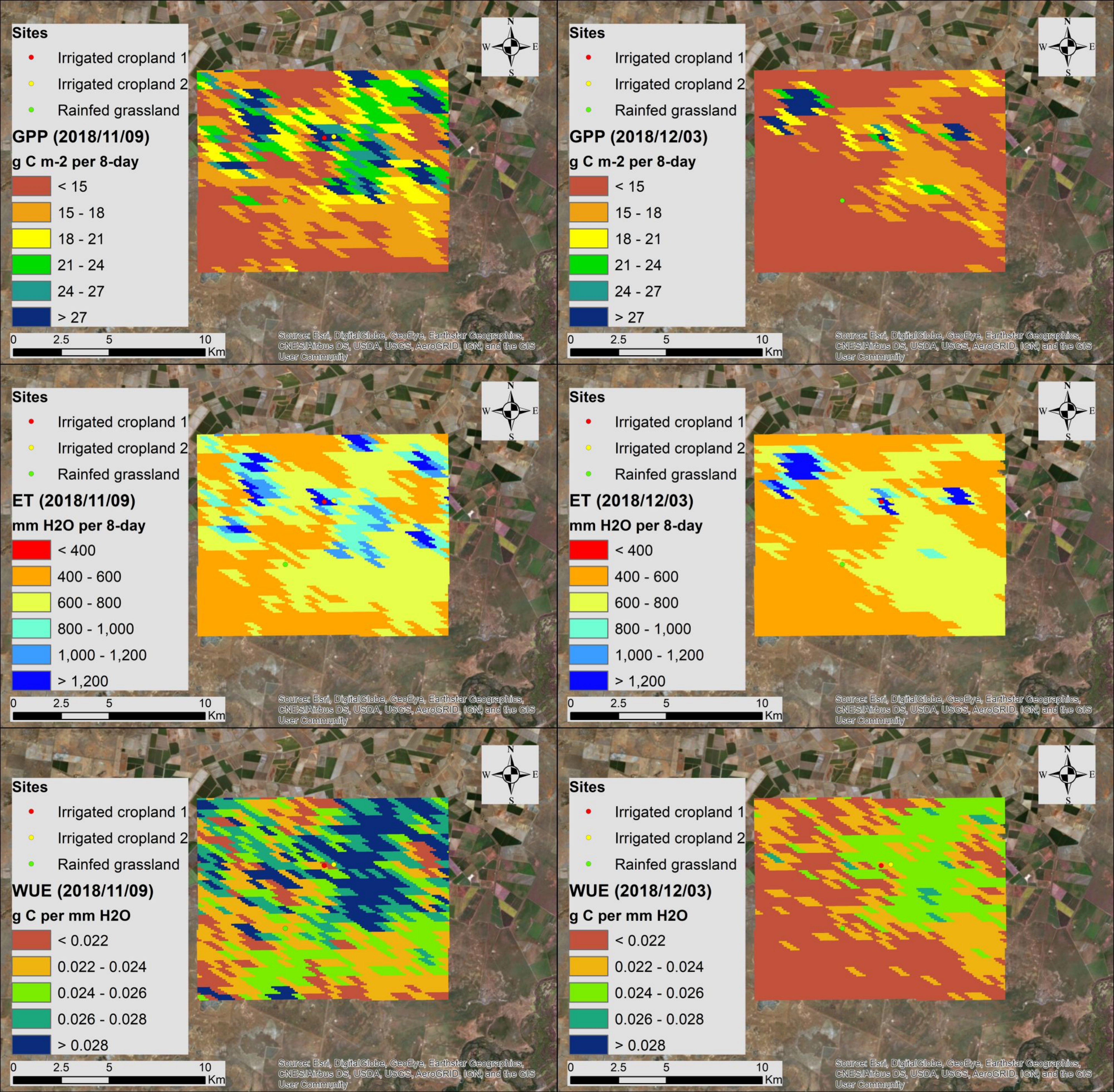
OZNET – Uri_Park (Cropland)
Measured
SMAP
QRF mean
QRF 5% & 95% quantiles
USCRN – CA_Fallbrook_5_NE (Shrubland)
REMEDHUS – Las_Arenas (Grassland)
SCAN – Lovelock NNR (Barren)
HOBE – 3.06 (Savanna)
SCAN – Reynolds Homestead (Forest)
SSM (m3 m-3)
0
0.1
0.2
0.3
0.4
0.5
2016-04-01
2016-10-01
2017-04-01
2017-10-01
2018-04-01
Date

Figure5.



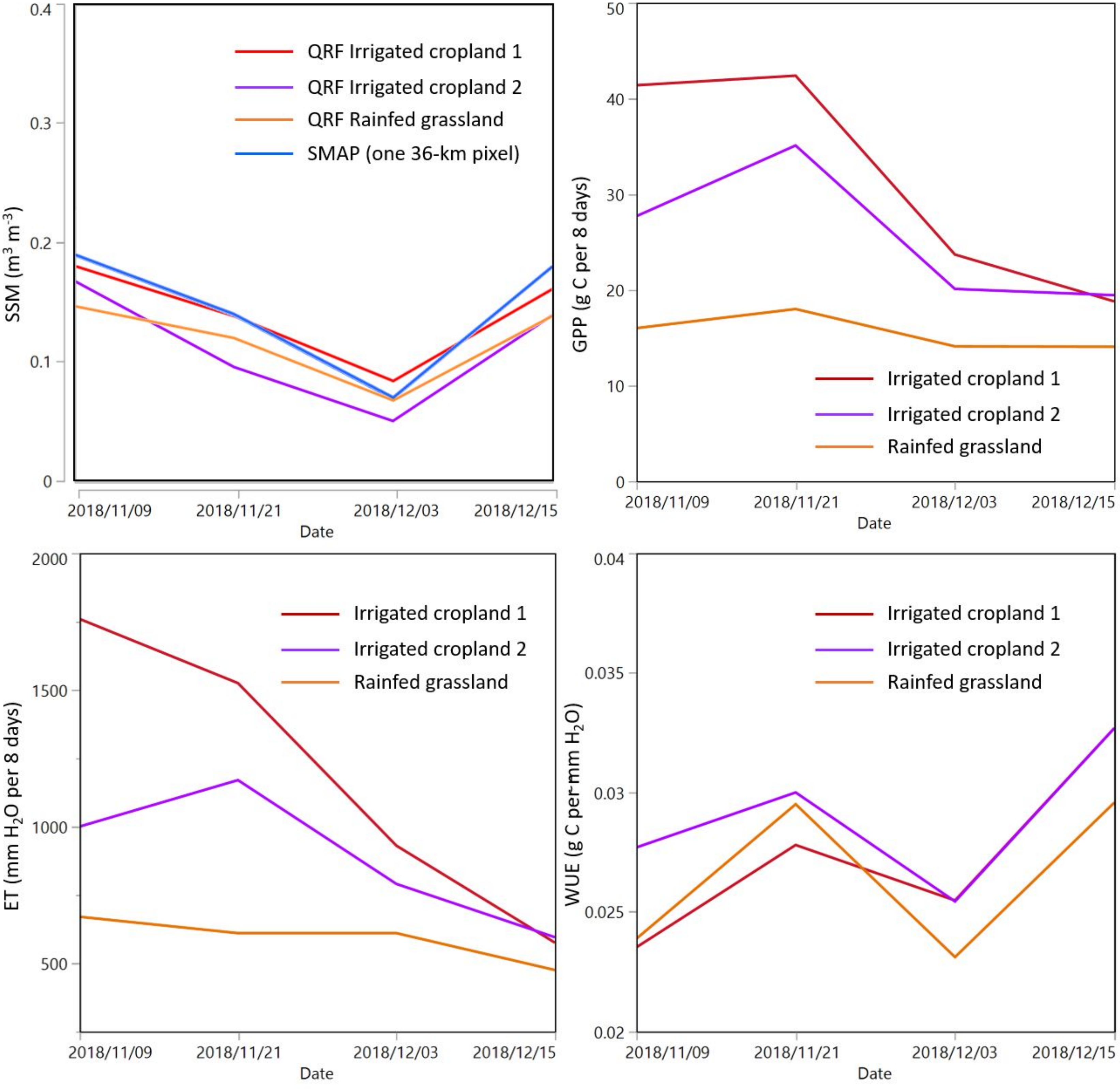
Selected field
Training (OZNET)
Validation (OZNET)
0 15 30 60 Km
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community
Sites
Irrigated cropland 1
Irrigated cropland 2
Rainfed grassland
Selected field
0 2.5 5 10 Km
SSM (2018/11/09)
m3 m-3
< 0.06
0.06 - 0.09
0.09 - 0.12
0.12 - 0.15
0.15 - 0.18
> 0.18
SMAP SSM = 0.19 m³ m⁻³
SSM (2018/11/21)
SMAP SSM = 0.14 m³ m⁻³
SSM (2018/12/03)
SMAP SSM = 0.07 m³ m⁻³
SSM (2018/12/15)
SMAP SSM = 0.18 m³ m⁻³
N
S
W
E

Figure6.



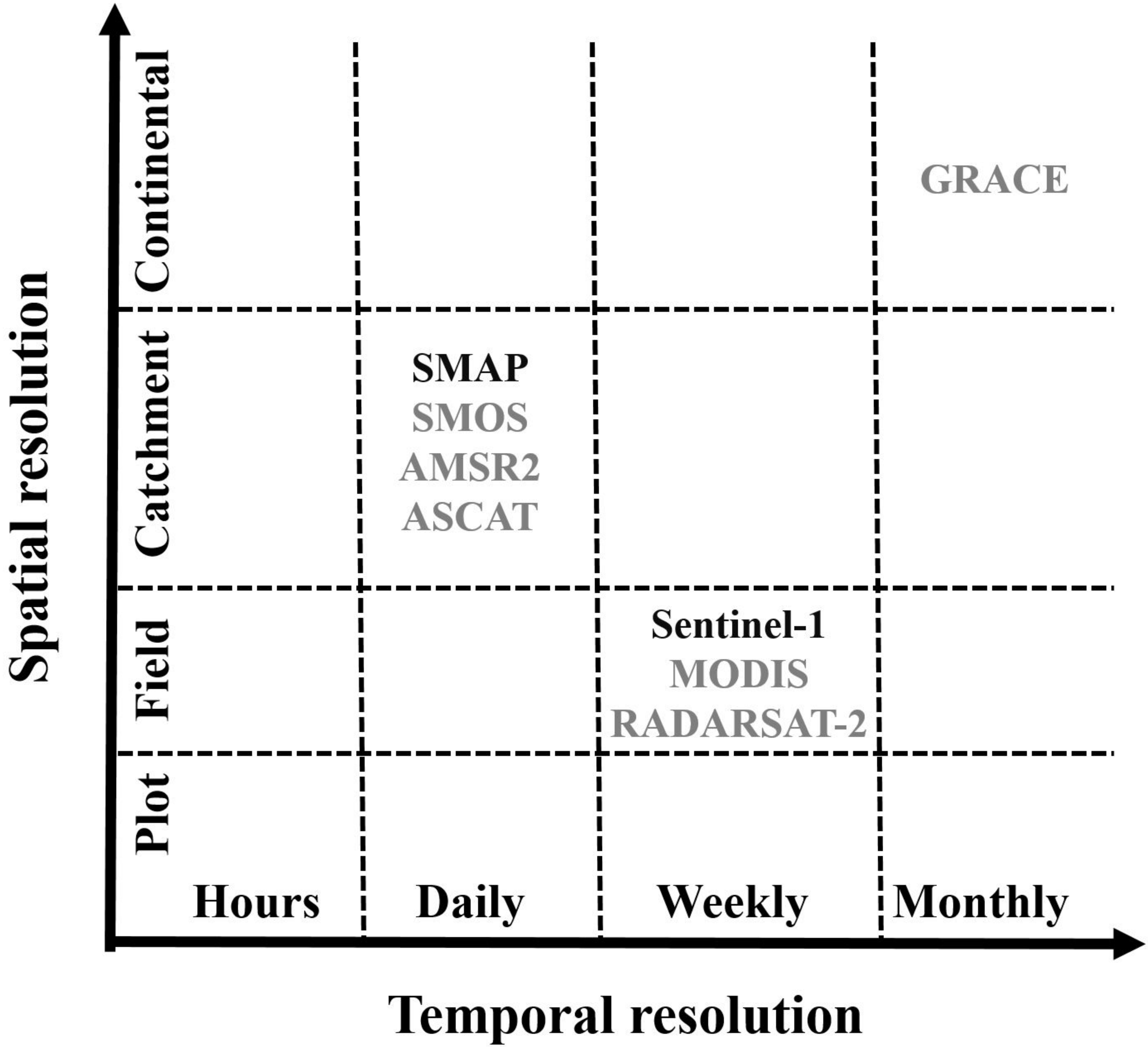
Sites
Irrigated cropland 1
Irrigated cropland 2
Rainfed grassland
GPP (2018/11/09)
g C m-2 per 8-day
< 15
15 - 18
18 - 21
21 - 24
24 - 27
> 27
GPP (2018/12/03)
ET (2018/11/09)
mm H2O per 8-day
< 400
400 - 600
600 - 800
800 - 1,000
1,000 - 1,200
> 1,200
ET (2018/12/03)
WUE (2018/11/09)
g C per mm H2O
< 0.022
0.022 - 0.024
0.024 - 0.026
0.026 - 0.028
> 0.028
WUE (2018/12/03)
0
2.5
5
10
Km
N
S
E
W
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

Figure7.



QRF Irrigated cropland 1
QRF Irrigated cropland 2
QRF Rainfed grassland
SMAP (one 36-km pixel)
SSM (m^3 m^{-3})
0
0.1
0.2
0.3
0.4
2018/11/09
2018/11/21
2018/12/03
2018/12/15
Date
GPP (g C per 8 days)
0
10
20
30
40
50
Irrigated cropland 1
Irrigated cropland 2
Rainfed grassland
ET (mm H_2O per 8 days)
500
1000
1500
2000
WUE (g C per mm H_2O)
0.02
0.025
0.03
0.035
0.04

Figure8.



Spatial resolution
Continental
Catchment
Field
Plot
GRACE
SMAP
SMOS
AMSR2
ASCAT
Sentinel-1
MODIS
RADARSAT-2
Hours
Daily
Weekly
Monthly
Temporal resolution

Table 1 Remote sensing and land surface datasets used for modeling surface soil moisture (SSM) and interpretation.

Dataset	Measuring/Estimated variable	Spatial resolution	Measuring interval
NASA-SMAP (L3_SM_P)	Brightness temperature retrieved to surface soil moisture (0–0.05 m)	36 km	2-3 days
ESA-Sentinel-1	Time-varying backscatter, incident angle, and time-constant temporal statistics (min, mean, max, standard deviation)	5×20 m	6-12 days/N.A.
GMTED2010 digital elevation model	Slope, aspect, flow direction, topographic position index, topographic roughness index	500 m	N.A.
SoilGrids & Openlandmap	Clay, silt, sand contents, soil organic carbon content, bulk density, field capacity, permanent witling point at depth 0–0.05 m	250 m	N.A.
MODIS (MCD12Q1.006)	Land cover types	500 m	N.A.
Station soil moisture monitoring networks	Soil water content at depth 0–0.05 m	N.A.	30-min

Table 2 Summary statistics of land cover types and surface soil moisture (SSM) at various regional-scale (HOBE, OZNET, REMEDHUS) and continental-scale (SCAN, USCRN) soil moisture monitoring networks for training and validation datasets.

Network	Country	No. Stations	Land cover	SSM (Training)						SSM (Validation)					
				No. stations	Min.	Mean	Median	Max	SD	No. stations	Min.	Mean	Median	Max	SD
HOBE	Denmark	24	Cropland (54%), Savanna (27%), Forest (18%)	18	0.03	0.19	0.18	0.49	0.08	6	0.00	0.18	0.17	0.38	0.09
OZNET	Australia	19	Cropland (52%), Grassland (48%)	14	0.00	0.14	0.14	0.40	0.09	5	0.00	0.15	0.15	0.40	0.08
REMEDHUS	Spain	12	Cropland (79%), Grassland (11%), Shrubland (10%)	8	0.00	0.10	0.09	0.33	0.07	4	0.05	0.20	0.18	0.45	0.09
SCAN	USA	147	Cropland (27%), Grassland (53%), Savanna (9%), Shrubland (5%), Forest (2%), Barren (4%)	109	0.00	0.16	0.13	0.58	0.12	38	0.00	0.18	0.16	0.55	0.12
USCRN	USA	76	Cropland (11%), Grassland (48%), Savanna (16%), Shrubland (13%), Forest (10%), Barren (3%)	56	0.00	0.16	0.14	0.55	0.11	20	0.00	0.21	0.19	0.61	0.11

Table 3 Comparison between measured surface soil moisture (SSM) with the predicted SSM from the quantile random forest (QRF) and SMAP based on the validation dataset. Note: *r*, Pearson's correlation coefficient; ME, mean error; RMSE, root mean squared error; values inside the brackets are the minimum and maximum values calculated among all validation stations and values outside the brackets are overall values calcualted by merging measurments from all the stations within a same land cover type.

Network	Land cover	Station names	No. Stations	No. measurements	QRF			SMAP		
					r	ME ($m^3 m^{-3}$)	RMSE ($m^3 m^{-3}$)	*r*	ME ($m^3 m^{-3}$)	RMSE ($m^3 m^{-3}$)
HOBE	Cropland	1.09, 3.04, 3.09	3	393	0.79 [0.65, 0.88]	0.01 [0.00, 0.01]	0.03 [0.03, 0.03]	0.74 [0.60, 0.84]	0.04 [0.03, 0.06]	0.06 [0.05, 0.07]
	Savanna	1.01, 3.06	2	431	0.62 [0.55, 0.58]	-0.01 [-0.08, 0.05]	0.08 [0.06, 0.11]	0.30 [0.54, 0.56]	0.04 [-0.05, 0.11]	0.11 [0.09, 0.12]
	Forest	1.04	1	135	0.58	0.52	0.58	0.52	0.58	0.52
OZNET	Cropland	Kyeamba_Mouth, Spring_Bank, Uri_Park, Wollumbi	4	159	0.60 [0.46, 0.86]	0.01 [-0.01, 0.05]	0.06 [0.04, 0.08]	0.63 [0.41, 0.83]	0.08 [0.06, 0.12]	0.13 [0.11, 0.14]
	Grassland	Cheverelis	1	29	0.88	0.90	0.88	0.90	0.88	0.90
REMEDHUS	Cropland	Canizal, Guarrati, Las_Bodegas	3	394	0.46 [0.29, 0.82]	-0.10 [-0.13, -0.03]	0.13 [0.05, 0.15]	0.42 [0.29, 0.81]	-0.10 [-0.13, -0.02]	0.13 [0.06, 0.16]
	Grassland	Las_Arenas	1	102	0.82	0.79	0.82	0.79	0.82	0.79
SCAN	Cropland	Abrams, Fort Reno #1, Molly Caren #1, North Issaquena, Perthshire, Princeton #1, Rock Springs Pa, Scott, Tidewater #1, Tunica, Uapb Dewitt, Uapb Point Remove, Uapb-Earle	13	817	0.66 [0.00, 0.79]	-0.01 [-0.07, 0.07]	0.08 [0.05, 0.11]	0.61 [0.04, 0.78]	0.02 [-0.05, 0.15]	0.10 [0.06, 0.15]
	Grassland	Alcalde, Bodie Hills, Crossroads, Jordan, Lindsay, Nephi, Stephenville, Torrington #1, Vermillion, Vernon, West Summit, Mandan #1, Price, Reese Center, Sheldon, Tule Valley, Violett, Walnut Gulch #1	18	1,878	0.60 [0.07, 0.89]	-0.01 [-0.09, 0.07]	0.07 [0.04, 0.11]	0.64 [0.07, 0.87]	-0.01 [-0.08, 0.07]	0.07 [0.03, 0.10]
	Savanna	Pee Dee, Powell Gardens, Morris Farms	3	199	0.79 [0.54, 0.71]	-0.03 [-0.08, 0.00]	0.07 [0.04, 0.10]	0.41 [0.57, 0.70]	0.06 [-0.04, 0.14]	0.11 [0.06, 0.15]
	Shrubland	Spooky	1	126	0.16	0.08	0.08	0.19	0.05	0.05
	Forest	Reynolds_Homestead	1	45	0.60	0.02	0.08	0.56	0.14	0.15
	Barren	Death Valley Jct., Lovelock NNR	2	222	0.79 [0.26, 0.70]	0.04 [0.04, 0.05]	0.05 [0.05, 0.06]	0.78 [0.22, 0.67]	0.05 [0.05, 0.05]	0.06 [0.05, 0.06]
USCRN	Cropland	IA_Des_Moines_17_E, KY_Versailles_3_NNW,	4	364	0.68	-0.03	0.07	0.57	-0.02	0.08

		NE_Lincoln_8_ENE, MO_Joplin_24_N			[0.62, 0.82]	[-0.07, 0.00]	[0.05, 0.09]	[0.63, 0.77]	[-0.09, 0.05]	[0.06, 0.11]
	Grassland	MT_Dillon_18_WSW, MT_Wolf_Point_34_NE, NC_Asheville_13_S, NE_Whitman_5_ENE, OK_Stillwater_2_W, OR_John_Day_35_WNW, SD_Aberdeen_35_WNW, SD_Pierre_24_S, TX_Muleshoe_19_S, MT_Lewistown_42_WSW, OR_Riley_10_WSW	11	1,022	0.65 [0.48, 0.91]	-0.01 [-0.08, 0.07]	0.07 [0.03, 0.10]	0.70 [0.47, 0.92]	-0.01 [-0.09, 0.06]	0.07 [0.03, 0.10]
	Savanna	IN_Bedford_5_WNW, MN_Goodridge_12_NNW	2	149	0.50 [0.45, 0.83]	-0.07 [-0.14, -0.04]	0.12 [0.08, 0.19]	0.67 [0.50, 0.81]	-0.03 [-0.07, -0.01]	0.08 [0.06, 0.13]
	Shrubland	CA_Fallbrook_5_NE	1	163	0.68	-0.04	0.05	0.87	-0.04	0.04
	Forest	CO_Boulder_14_W	1	152	0.58	-0.03	0.07	0.61	0.00	0.06
	Barren	ID_Arco_17_SW	1	39	0.85	0.05	0.07	0.87	0.03	0.05
Overall	Cropland	–	27	2,127	0.73 [0.00, 0.88]	-0.02 [-0.13, 0.07]	0.08 [0.03, 0.15]	0.64 [0.04, 0.84]	0.00 [-0.13, 0.15]	0.10 [0.05, 0.16]
	Grassland	–	31	3,031	0.63 [0.07, 0.91]	-0.01 [-0.09, 0.07]	0.07 [0.03, 0.11]	0.67 [0.07, 0.92]	-0.01 [-0.09, 0.07]	0.07 [0.03, 0.10]
	Savanna	–	7	779	0.73 [0.45, 0.83]	-0.02 [-0.14, 0.05]	0.09 [0.04, 0.19]	0.54 [0.50, 0.81]	0.03 [-0.07, 0.14]	0.11 [0.06, 0.15]
	Shrubland	–	2	289	0.22 [0.16, 0.68]	0.01 [-0.04, 0.08]	0.07 [0.05, 0.08]	0.78 [0.19, 0.87]	0.00 [-0.04, 0.05]	0.05 [0.04, 0.05]
	Forest	–	3	332	0.58 [0.58, 0.60]	-0.05 [-0.11, 0.02]	0.09 [0.07, 0.11]	0.63 [0.52, 0.61]	0.01 [-0.03, 0.14]	0.08 [0.06, 0.15]
	Barren	–	3	261	0.77 [0.26, 0.85]	0.04 [0.04, 0.05]	0.05 [0.05, 0.07]	0.77 [0.22, 0.87]	0.04 [0.03, 0.05]	0.06 [0.05, 0.06]