

Spatial-temporal modeling of root zone soil moisture dynamics in a vineyard using machine learning and remote sensing

Kisekka, Isaya; Peddinti, Srinivasa Rao; Kustas, William P.; McElrone, Andrew J.; Bambach-Ortiz, Nicolas; McKee, Lynn; Bastiaanssen, Wim

DOI

[10.1007/s00271-022-00775-1](https://doi.org/10.1007/s00271-022-00775-1)

Publication date

2022

Document Version

Final published version

Published in

Irrigation Science

Citation (APA)

Kisekka, I., Peddinti, S. R., Kustas, W. P., McElrone, A. J., Bambach-Ortiz, N., McKee, L., & Bastiaanssen, W. (2022). Spatial-temporal modeling of root zone soil moisture dynamics in a vineyard using machine learning and remote sensing. *Irrigation Science*, 40(4-5), 761-777. <https://doi.org/10.1007/s00271-022-00775-1>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Spatial–temporal modeling of root zone soil moisture dynamics in a vineyard using machine learning and remote sensing

Isaya Kisekka^{1,2} · Srinivasa Rao Peddinti¹ · William P. Kustas³ · Andrew J. McElrone^{4,5} · Nicolas Bambach-Ortiz⁴ · Lynn McKee³ · Wim Bastiaanssen⁶

Received: 30 October 2021 / Accepted: 18 January 2022
© The Author(s) 2022

Abstract

High-resolution spatial–temporal root zone soil moisture (RZSM) information collected at different scales is useful for a variety of agricultural, hydrologic, and climate applications. RZSM can be estimated using remote sensing, empirical equations, or process-based simulation models. Machine learning (ML) approaches for evaluating RZSM across numerous spatial–temporal scales are less generalizable than process-based models. However, data-driven ML approaches offer a unique opportunity to develop complex models of soil moisture without making assumptions about the processes governing soil water dynamics in a given study region. In this study, comparisons were made between two models, pySEBAL and EFSOIL, which were based on evaporation fraction (EF) and soil properties, and a data-driven model based on the Random Forest (RF) ensemble algorithm. These approaches were evaluated to demonstrate their capabilities for RZSM estimation. The EF obtained from Landsat images was used after validation with eddy covariance measurements as the major input to all three models, along with other meteorological and soil physical properties. The RF model was trained using in situ soil moisture data from Time Domain Reflectometry (TDR) sensors installed in a vineyard from 2018 to 2020. The predictor variables comprised of meteorological, soil properties, EF, and a vegetation index. The results reveal that there was a strong correlation between the in situ measured soil moisture and the RF predicted soil moisture at all sensor locations. Due to the complexity of the physical processes involved in soil water flow, the empirical models pySEBAL and EFSOIL were unable to reliably predict RZSM values at all monitored locations. The high RZSM predicted by pySEBAL demonstrated the presence of possible bias in the model's algorithm used to estimate soil moisture. We also demonstrated that ML based on the RF algorithm may be used to predict spatially distributed RZSM when a few soil moisture ground measurements are combined with remote sensing to produce soil moisture maps.

✉ Isaya Kisekka
ikisekka@ucdavis.edu

¹ Department of Land, Air and Water Resources, University of California Davis, One Shields Avenue, Davis, CA 95616-5270, USA

² Department of Biological and Agricultural Engineering, University of California Davis, Davis, CA 95616, USA

³ Hydrology and Remote Sensing Laboratory, USDA-ARS, Building 007, BARC West, Room 104, Beltsville, MD 20705-2350, USA

⁴ Department of Viticulture and Enology, University of California, Davis, Davis, CA 95616, USA

⁵ Crops Pathology and Genetics Research Unit, US Department of Agriculture-Agricultural Research Service, Davis, CA 95618, USA

⁶ Delft University of Technology (TU Delft), Stevinweg 1, 2628 CN Delft, The Netherlands

Introduction

As the world's population continues to grow and water resources increasingly become scarce (Boretti and Rosa 2019) due to climate change, competition from various beneficial uses, and increased regulation of agriculture use, there is an urgent need for the development of more sustainable production practices. For example, in California, from the winter of 2011 to 2016, the state experienced extreme drought that impacted agricultural production and served as the catalyst for passing the Sustainable Groundwater Management Act (SGMA), which limits groundwater pumping for irrigation. Currently (as of October 2021), California is experiencing another major drought that started in 2020. The increased frequency of extreme drought and increased regulation of groundwater use are making irrigated agriculture challenging in California. Achieving precision irrigation

water management requires a thorough understanding of crop-specific biophysical processes including crop response to water at different growth stages as well as variability over space. Soil moisture is a critical and vital variable that can help us improve our understanding of the relationships between climate dynamics (D'Odorico and Porporato 2004), water (Gao et al. 2014), drought (Sheffield and Wood 2008), and food security (Sadri et al. 2020). Information on soil moisture is important for the development of appropriate irrigation systems to maximize crop yield, and long-term soil moisture information combined with climatic information provides insights into patterns, agricultural thresholds, and losses (Bastiaanssen et al. 2007; Lin et al. 2018). Thus, soil moisture information is required to achieve the benefits of precision agriculture and agricultural sustainability.

Soil moisture collected at the soil surface is referred to as surface soil moisture (SSM), and most remote-sensing methods are confined to the determination of SSM. Soil moisture obtained at deeper depths where plant root water uptake occurs is referred to as root zone soil moisture (RZSM) and is of much more relevance (Scott et al. 2003; Bauer-Marschallinger et al. 2019) to management and decision making. Atmospheric conditions have a more direct impact on SSM than on RZSM (Hirschi et al. 2014). A precise RZSM estimate is required for quantifying plant available water and for irrigation scheduling. In situ sensors deployed at various depths along the soil profile can provide direct RZSM data (Cosh et al. 2016). However, one disadvantage of the soil moisture sensors is that they require site-specific calibration to provide reliable measurements of volumetric water content. Because of the effects that variations in soil properties have on sensor output, it is necessary to calibrate the sensor at each site (Peddinti et al. 2020a). Also, the volume of soil sensed is somewhat limited. With the advent of new in situ and proximal sensors, new satellites, other sensing technologies, and increased modeling capabilities, soil moisture monitoring techniques are undergoing fast expansion and innovation (Peddinti et al. 2018, 2020b). As a result, an increasing variety of soil moisture data products are being developed.

Various theoretical and empirical models were developed over the last few decades to retrieve surface soil moisture at a depth of 0–5 cm (Petropoulos et al. 2015) by utilizing passive or active microwave sensors and establishing the relationship between the soil dielectric constant and water content (Jackson 1993). In spite of this, remote-sensing satellites are unable to provide direct soil moisture content in the root zone at depths of 30–60 cm because of technical limitations imposed by L-band and X-band characteristics within the microwave wavelength range (Engman 1991). In most cases, SSM data may be obtained more easily than RZSM data. In most instances, the RZSM can be extrapolated from SSM data, which can be obtained either in situ

or through satellites (Wigneron et al. 1999; Montaldo et al. 2001; Sabater et al. 2007). Spatially distributed RZSM can be a challenge, since the installation of a large number of sensors in the network within the subsurface is costly and time-consuming, and is likely to affect the soil characteristics (González-Teruel et al. 2019). It is, unfortunately, less common to measure RZSM at depths of 100 cm and beyond, such as is required for deep rooting crops, e.g., pistachio trees. The use of data-driven models that can efficiently relate the inputs to the desired output while being computationally efficient is required to precisely predict spatially distributed RZSM (Kornelsen and Coulibaly 2014; Carranza et al. 2021).

On a large-scale, detailed process-based models for simulating soil moisture dynamics based on the Richards equation require a lot of data for parameterization and can be computationally expensive. Data-driven prediction technologies (Kornelsen and Coulibaly 2014) such as artificial neural networks (ANNs) (Hassan-Esfahani et al. 2015), Random Forest (RF) (Carranza et al. 2021), and statistical learning tools, such as Support Vector Machines (SVMs) (Yu et al. 2012), are increasingly being used for SSM and RZSM estimations. They are designed to extract information from data by examining patterns of variability in the data and to stimulate responses that are being taught by the data in the process. A common prerequisite is that in situ data are properly calibrated and sufficient for model training. Specifically, data-driven techniques implicitly include and assess all of the interacting processes that result in the production of a specific RZSM state (Carranza et al. 2021). Advances in machine learning (ML) techniques have been mostly utilized in hydrology (Lange and Sippel 2020) and climate research (Huntingford et al. 2019) for the prediction and forecasting of environmental variables (Li et al. 2011), as well as the optimization of model parameters. Over the last few years, ML approaches have become more common in soil hydrology research to estimate model-derived RZSM using ANNs or satellite-derived SSM using SVMs (Yu et al. 2012; Adab et al. 2020; Carranza et al. 2021).

Yu et al. (2012) used SVMs and the ensemble particle filter (EnPF) to develop a multi-layer soil moisture prediction model for the Meilin watershed in China, which showed that SVMs are statistically significant and resilient for soil moisture prediction in both the surface and root zone layers. Using simulated soil moisture data from the Soil and Water Assessment Tool (SWAT), Al-Mukhtar (2016) determined which ANNs were most effective for modeling the RZSM up to 2 m depth. He found that layer recurrent network and feedforward network were the most effective estimators. Kornelsen and Coulibaly (2014) applied the process-based HYDRUS model and data-driven ANNs using surface soil moisture observations to predict RZSM. They demonstrated that ANNs were capable of accurately predicting soil

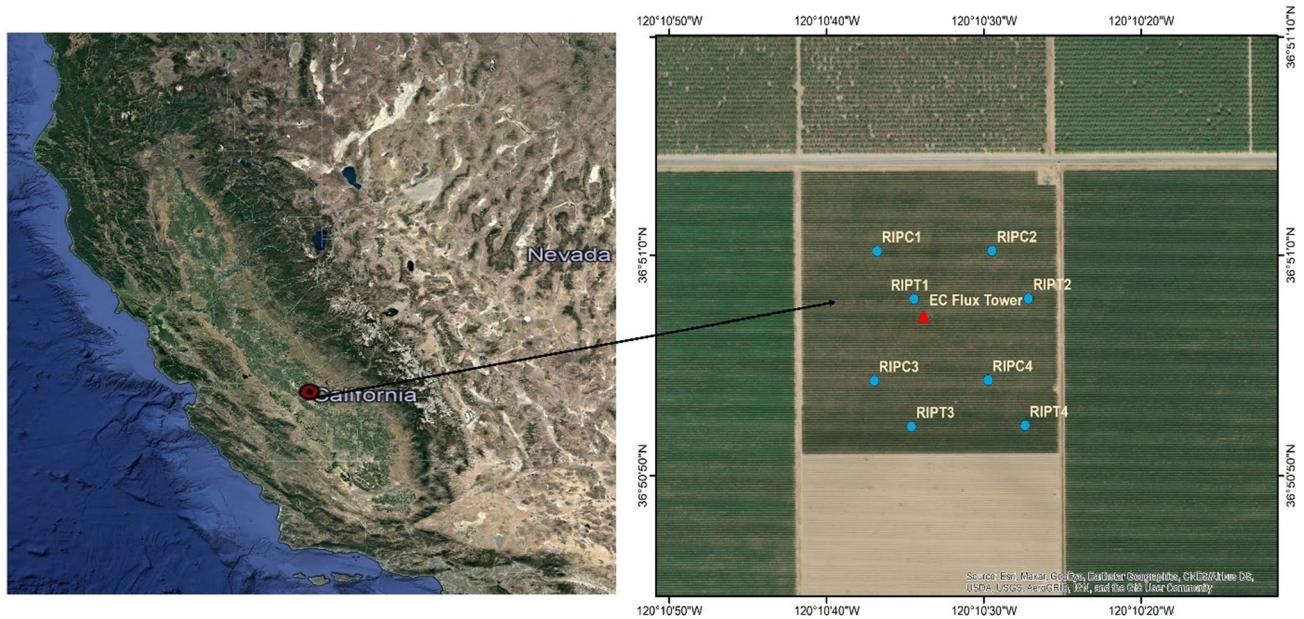


Fig. 1 The study area showing the eight soil moisture sensors (RIPC1, RIPC2, RIPC3, RIPC4, RIPT1, RIPT2, RIPT3, and RIPT4) and four flux tower (red triangle) locations at Ripperdan Ranch, near

Madera, was placed under the grapevines for which root zone soil moisture and both carbon and water flux data were collected

moisture as estimated by HYDRUS, but the performance was reduced when compared to in situ moisture observations outside the training conditions. According to the findings of a recent study by Carranza et al. (2021), in situations where adequate training data can be obtained from intense observing campaigns where soil hydraulic parameters are not accessible, the RF model was shown to be more favorable than the process-based HYDRUS model, particularly when the primary goal was to predict soil moisture content.

The purpose of this study was to compare the performance of two semi-empirical models, pySEBAL and EFSOIL to the machine learning-based RF model in combination with remote sensing data for predicting spatial-temporal changes in RZSM in a commercial vineyard. The in situ soil moisture data were acquired from eight TDR soil moisture sensors in a vineyard at Ripperdan Ranch, California during the growing seasons from 2018 to 2020 as part of the Grape Remote Sensing Atmospheric Profile and Evapotranspiration eXperiment (GRAPEX).

Materials and methods

Study area description and soil moisture data

In this study, in situ soil moisture measurements were acquired from Ripperdan Ranch near Madera, CA (38.8500 N, – 120.1768 W), as part of the GRAPEX project (Kustas et al. 2018; Alfieri et al. 2019). As depicted in

Fig. 1, soil moisture data were collected from eight TDR soil moisture sensors (Model: CS655, Campbell Sci. Inc., Logan, UT, USA¹) over the 2018–2020 grapevine crop growth seasons. Soil samples were collected in 2018 from eight different locations at three different depths: 30, 60, and 90 cm, for soil texture assessment and sensor calibration. Using the gravimetric calibration procedure (Peddinti et al. 2020a), the following calibration equation was developed to get calibrated volumetric water content from the sensors:

$$\theta_{\text{volumetric}} = (1.2713 \times (\theta_{\text{gravimetric}})^2) + (0.2914 \times \theta_{\text{gravimetric}}) + 0.0873;$$

here, $\theta_{\text{volumetric}}$ is volumetric water content obtained from the sensors and $\theta_{\text{gravimetric}}$ is the gravimetric water content obtained from the soil samples (Table 1).

The daily average RZSM refers to volumetric soil moisture content within the top 60 cm, calculated as the average of sensor readings at 30 and 60 cm or, in some cases, 40 and 60 cm depth. It is possible that the majority of the grapevine root system can be found deeper than 60 cm; however, on the basis of the available data, we have considered the data from 60 cm depth as the effective RZSM.

¹ Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Table 1 Soil properties within the sensor locations at the root zone in the study region

ID	Date	Depth (cm)	% Sand	% Silt	% Clay	Class	K_s (cm/day)	Porosity (%)	PWP ($\text{cm}^3 \text{cm}^{-3}$)	FC ($\text{cm}^3 \text{cm}^{-3}$)	AW ($\text{cm}^3 \text{cm}^{-3}$)	BD (g cm^{-3})
RIPC1	11/13/2018	30–60	66	22	12	Sandy loam	36.34	0.42	0.10	0.20	0.10	1.53
RIPC2	11/13/2018	30–60	58	28	14	Sandy loam	25.79	0.44	0.10	0.22	0.11	1.49
RIPC3	11/13/2018	30–60	59	24	17	Sandy loam	22.09	0.45	0.12	0.22	0.11	1.47
RIPC4	11/13/2018	30–60	66	22	12	Sandy loam	36.34	0.42	0.10	0.20	0.10	1.53
RIPT1	4/6/2018	30–60	56	28	16	Sandy loam	21.33	0.45	0.11	0.23	0.11	1.47
RIPT2	4/6/2018	30–60	56	30	14	Sandy loam	24.22	0.44	0.10	0.22	0.12	1.49
RIPT3	11/13/2018	30–60	60	30	10	Sandy loam	36.12	0.42	0.09	0.21	0.12	1.55
RIPT4	4/6/2018	30–60	64	22	14	Sandy loam	30.61	0.43	0.10	0.21	0.10	1.51

Here, K_s is soil hydraulic conductivity (cm day^{-1}), PWP is permanent wilting point ($\text{cm}^3 \text{cm}^{-3}$), FC is the field capacity ($\text{cm}^3 \text{cm}^{-3}$), AW is the available water ($\text{cm}^3 \text{cm}^{-3}$), and BD is the soil bulk density (g cm^{-3})

The majority of sensors were positioned beneath the vine row, where they sensed irrigation events during the growing season. In the root zone, the soil was classified as sandy loam, with 60% sand, 25% silt, and 15% clay. There was a difference in the soil bulk density ranging between 1.47 and 1.55 g cm^{-3} at all sensor locations. Field capacity was estimated as 0.21 $\text{cm}^3 \text{cm}^{-3}$, and the permanent wilting point was 0.10 $\text{cm}^3 \text{cm}^{-3}$ (Table 2). The soil moisture readings collected from all of the sensor locations over a three-year period ranged from 0.08 to 0.30 $\text{cm}^3 \text{cm}^{-3}$ (Fig. 2), with a saturated moisture content of 0.39 $\text{cm}^3 \text{cm}^{-3}$ and residual water content of 0.048 $\text{cm}^3 \text{cm}^{-3}$.

Remote sensing and meteorological data

To estimate the EF fraction from an energy balance model during the period 2018–2020 grapevine crop growing seasons, a total of 123 Landsat Thematic Mapper (TM)-Enhanced Thematic Mapper Plus (ETM+)/Operational Land Imager (OLI) images under clear sky conditions were collected from the USGS Earth Resources Observation and Science Center (<https://earthexplorer.usgs.gov/>). Image dates for Landsat 7 and 8 are presented in Fig. 3 along with the corresponding day of the year (DOY). The USGS EROS Center produced a high-resolution digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) with a resolution of 90 m, which was rescaled to a resolution of 30 m to match Landsat resolution.

The meteorological data, i.e., hourly and daily data, including solar radiation, air temperature, relative humidity, and wind speed required to run the pySEBAL model and energy balance components including net radiation, soil heat flux, and sensible heat flux, were collected from the eddy covariance (EC) flux tower located within the study region, as shown in Fig. 1. The ET data were also obtained from this EC tower which is one of the primary biophysical processes that governs root zone water dynamics. In an earlier publication, Alfieri et al. (2019) provide detailed information on the general sensor design of the GRAPEX eddy covariance flux towers with more specific details given by Knipper et al. (2019).

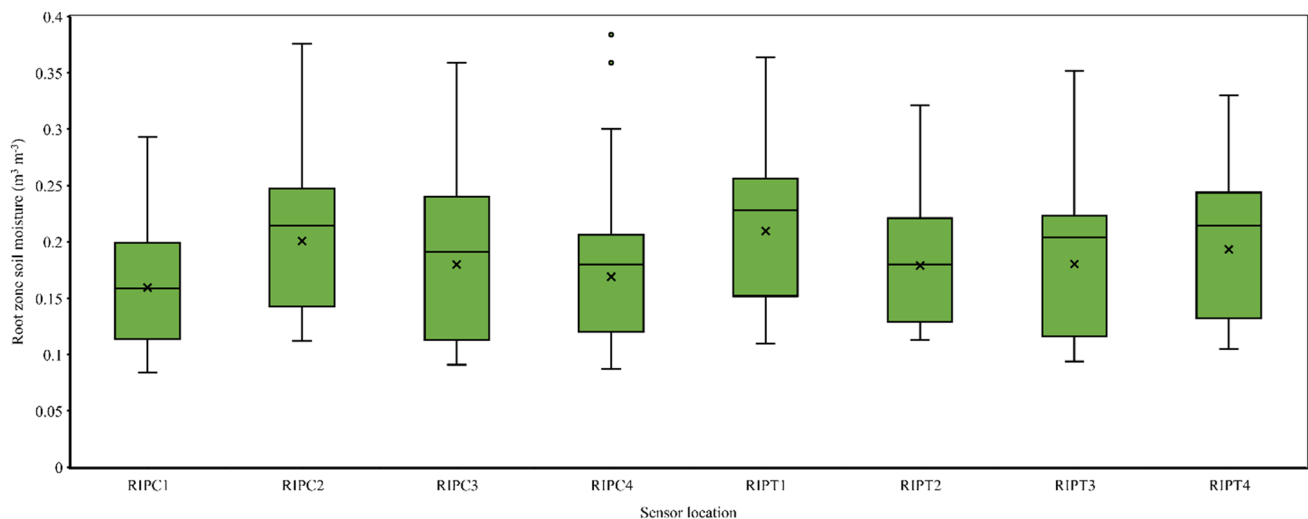
Root zone soil moisture estimation from pySEBAL

Traditional systems that use soil moisture sensors in conjunction with data from meteorological stations provide only point measurements, and are relatively expensive, and require frequent maintenance. IrriWatch (<https://www.irriwatch.com/en/>; IrriWatch, Maurik, The Netherlands) addresses this issue by combining root zone soil moisture and crop ET into a single, cost-effective technology commercial service that does not require hardware installation. IrriWatch is one of the first products to deliver comprehensive

Table 2 The goodness-of-fit statistics for the three models assessed for pySEBAL, EFSOIL, and RF models were compared to measured and predicted root zone soil moisture at eight sensor locations

Site	RMSE			Bias			R^2			d		
	pySEBAL	EFSOIL	RF	pySEBAL	EFSOIL	RF	pySEBAL	EFSOIL	RF	pySEBAL	EFSOIL	RF
RIPC1	0.12	0.07	0.031	0.06	0.034	0.028	0.16	0.19	0.83	0.27	0.12	0.86
RIPC2	0.09	0.06	0.012	0.02	0.003	0.001	0.04	0.09	0.94	0.31	0.22	0.98
RIPC3	0.12	0.07	0.025	0.049	0.03	0.014	0.23	0.3	0.83	0.23	0.12	0.92
RIPC4	0.11	0.06	0.036	0.052	0.027	0.031	0.06	0.11	0.85	0.27	0.14	0.86
RIPT1	0.11	0.06	0.019	0.033	-0.003	-0.008	0.04	0.09	0.89	0.29	0.21	0.96
RIPT2	0.09	0.05	0.021	0.043	0.017	0.008	0.05	0.11	0.84	0.31	0.21	0.95
RIPT3	0.12	0.07	0.024	0.052	0.032	0.004	0.39	0.42	0.77	0.18	0.09	0.93
RIPT4	0.12	0.07	0.023	0.035	0.009	0.008	0.07	0.12	0.84	0.27	0.19	0.95
Average	0.11	0.06	0.024	0.043	0.019	0.011	0.13	0.18	0.85	0.27	0.16	0.93

Here, RMSE is the root-mean-square error, bias is the mean bias error, R^2 is the coefficient of determination, and d is the index of agreement between in situ and predicted RZSM at eight sensor locations

**Fig. 2** Box plots depicting the distribution of root zone soil moisture at eight sensor locations on the Ripperdan Ranch, near Madera, California

soil water potential and soil moisture data in the root zone to the farmers that is derived from thermal, multispectral, and optical-based remote sensing. In addition, estimates of ET and crop production among others are provided at the spatial resolution of 10 m for any field around the globe. The Surface Energy Balance Algorithm for Land (SEBAL) is the core algorithm behind IrriWatch, and it was implemented in a python environment, which is referred to as pySEBAL (python Surface Energy Balance Algorithm for Land) (Hessels et al. 2017; Jaafar and Ahmad 2020). The following section includes the major equations for estimating RZSM using the pySEBAL algorithm. Readers are referred to the following articles (Bastiaanssen et al. 1998b, a; Laipelt et al. 2021) for a more detailed explanation of the algorithm.

The soil moisture content is computed as a function of the evaporation fraction (EF) using the pySEBAL package

(Waters et al. 2002). Bastiaanssen et al. (1997) and Scott et al. (2003) were the first studies to identify a relationship between soil moisture and the EF, leading to the formulation of Eq. (1)

$$\theta = \theta_{\text{sat}} \times \exp\left(\frac{\text{EF} - a}{b}\right), \quad (1)$$

where θ_{sat} is saturated soil water content ($\text{cm}^3 \text{cm}^{-3}$), EF is evaporation fraction from pySEBAL, and a and b are curve fitting parameters set to 1 and 0.421. Note that Eq. (1) directly relates EF to volumetric soil water content. Nutini et al. (2014) evaluated this equation in semi-arid rangeland ecosystems in Niger and Chad, and Petropoulos et al. (2020) investigated the relationship in a Mediterranean environment in Spain.

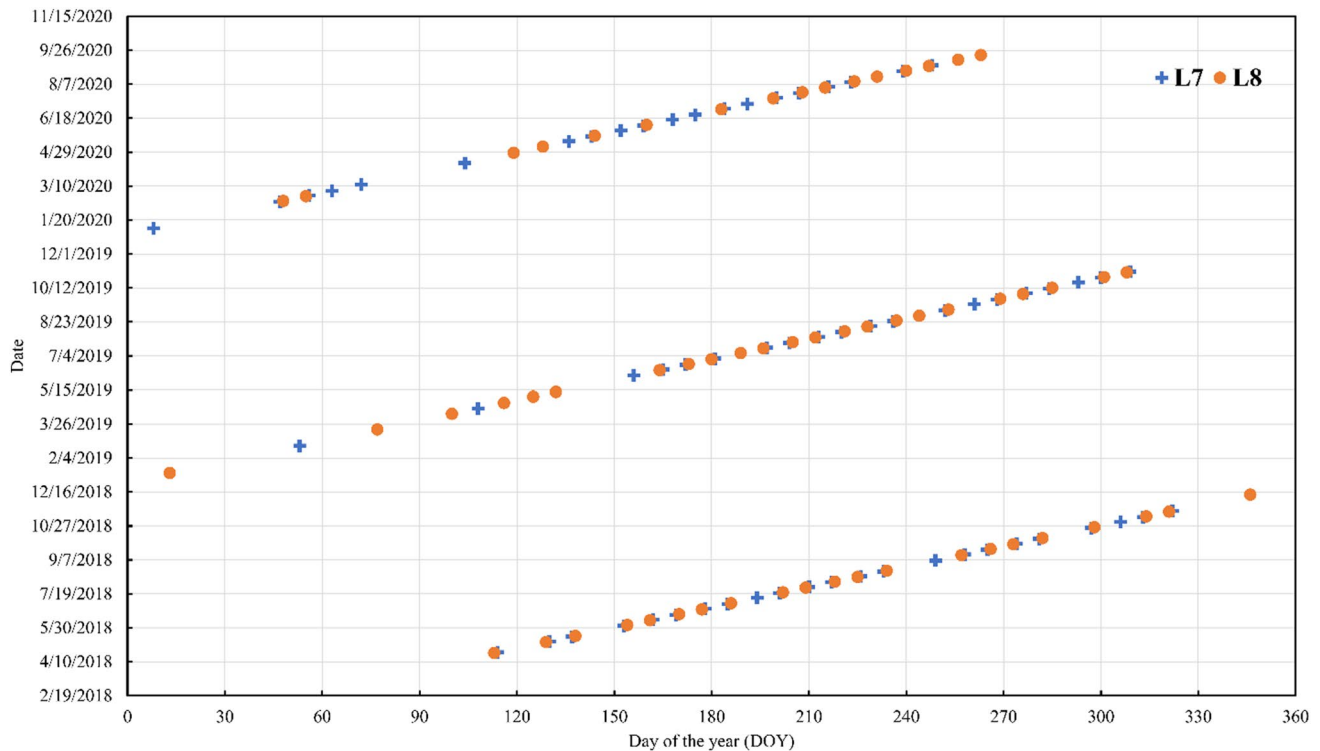


Fig. 3 Cloud free Landsat scene imagery (path: 042/043 row:034) was used to estimate evaporation fraction

The following formula is used to calculate the EF from the energy balance equation:

$$EF = \frac{R_n - G - H}{R_n - G}, \quad (2)$$

where R_n is the net radiation (W m^{-2}) at the surface, H is the sensible heat flux (W m^{-2}), and G is the soil heat flux (W m^{-2}).

The net radiation is the first and the most important computing step in the pySEBAL method. According to Eq. (2), R_n is computed by subtracting all outgoing radiation fluxes from all incoming radiation fluxes by Eq. (3)

$$R_n = (1 - \alpha)R_{s\downarrow} + \varepsilon_0 R_{L\downarrow} - \varepsilon_0 \sigma T_s^4, \quad (3)$$

where $R_{s\downarrow}$ is the incoming shortwave radiation measured at the time of satellite overpass (W m^{-2}), $R_{L\downarrow}$ is the incoming longwave radiation (W m^{-2}), α is the surface albedo, ε_0 is the surface emissivity calculated by a semi-empirical relationship involving Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI) (Xue et al. 2020), which can be retrieved from the red and near-infrared bands of the electromagnetic spectrum, σ is the Stephen–Boltzmann constant denoted as 5.67×10^{-8} ($\text{W m}^{-2} \text{K}^4$), and T_s is the temperature of the land surface (K).

G is expressed as a fraction of R_n , and pySEBAL employs the empirical formula (Eq. 4) for calculation of G established by (Bastiaanssen et al. 1997)

$$G = T_{s,\text{reference}} \times (0.0038 + 0.007\alpha) \times (1 - 0.98\text{NDVI}^4) \times R_n, \quad (4)$$

where $T_{s,\text{reference}}$ is the corrected land surface temperature (T_s) based on the DEM of the area of interest (AOI), considering the slope and aspect of the land surface and NDVI.

When using pySEBAL, an internal calibration of H is implemented, eliminating the requirement for an additional atmospheric adjustment of T_s to be performed. While calculating H , pySEBAL makes use of the bulk aerodynamic resistance equation, expressed as Eq. (5)

$$H = \frac{\rho \times C_p \times dT}{r_{\text{ah}}}, \quad (5)$$

where ρ is the air density (kg m^{-3}), C_p is the specific heat of air at constant pressure and it is equal to $1004 \text{ J}/(\text{kg K})$, r_{ah} is the aerodynamic resistance to heat transfer between z_1 and z_2 (s/m), and dT is the temperature difference between two near-surface height ($z_1 = 0.1 \text{ m}$ and $z_2 = 2 \text{ m}$) above the canopy layer (K) (Xue et al. 2020), which is estimated as a linear function of corrected surface temperature $T_{s,\text{reference}}$ (Eq. 4), being a major assumption for estimating sensible

heat flux (Bastiaanssen 1995; Allen et al. 2005b). Readers are referred to (Xue et al. 2020; Jaafar and Ahmad 2020) for a more detailed description of the pySEBAL algorithm and the automated selection of hot and cold pixels in the calculation of sensible heat fluxes. After estimating ET, IrriWatch uses a soil water balance and site-specific information on soil physical characteristics such as field capacity, wilting point, and effective crop root depth to estimate soil water content within the root zone. Soil physical characteristics are obtained from public databases such as gSSURGO. Validation of the EF data was carried out using EC tower data from the study region's source footprint area to ensure accuracy.

Root zone soil moisture estimation from evaporation fraction (EF) and soil properties

In the absence of ground-based auxiliary measurements, Pradhan (2019) proposed a method for estimating soil moisture content from the satellite-derived EF and soil physical properties. In this study, the relationship between satellite-based EF and soil properties was used to derive RZSM. In this study, this approach was defined as EFSOIL (Evapotranspiration fraction and soil properties-based RZSM).

Budyko and Zubenok (1961) defined the ratio of actual crop evapotranspiration (AET) to plant-specific reference evapotranspiration (ET_r) as a function of actual available soil moisture (SM) and plant available soil moisture (PAM), which can be written as

$$\frac{AET}{ET_r} = \frac{SM}{PAM} \quad (6)$$

This is a simplified equation for a complex physical process that requires additional testing for various soil and climatic conditions. In Eq. (6), the right-hand side term can be taken as the relative saturation or soil wetness index defined as

$$\frac{SM}{PAM} = \frac{(\theta - \theta_{wp})}{(\theta_{fc} - \theta_{wp})} \quad (7)$$

where θ is the actual soil water content within the root zone ($\text{cm}^3 \text{cm}^{-3}$), and θ_{fc} and θ_{wp} are site-specific field capacity and permanent wilting point ($\text{cm}^3 \text{cm}^{-3}$), respectively.

By rearranging the terms from Eqs. (6) and (7), Eq. (8) can be written as

$$\frac{AET}{ET_r} = \frac{(\theta - \theta_{wp})}{(\theta_{fc} - \theta_{wp})} \quad (8)$$

Similarly, in Eq. (8), the left-hand side term is equal to the crop coefficient (Kc). As reported by Trezza 2002; Tasumi 2003; and Allen et al. 2007, Kc can be similar to the EF under certain conditions (Allen et al. 1998). While the

parameter EF takes into consideration water stress, the actual ET is the variable that accounts for environmental stresses (Allen et al. 2005a). By assuming that EF is equal to Kc and that it is related to soil moisture fraction defined in Eq. (8), the following equation can be written as:

$$EF = \frac{(\theta - \theta_{wp})}{(\theta_{fc} - \theta_{wp})} \quad (9)$$

Equation (9) can be used to derive spatially distributed soil moisture content (θ_i) at any spatial location denoted as

$$\theta_i = EF(\theta_{fci} - \theta_{wpi}) + \theta_{wpi} \quad (10)$$

This is the fundamental equation that was used in the EFSOIL model to retrieve the RZSM within the vineyard at the Ripperdan Ranch by utilizing the Landsat-derived EF from pySEBAL algorithm and soil characteristics at each sensor location.

Root zone soil moisture estimation using Random Forest

To estimate RZSM at fine spatial resolution, the RF machine learning algorithm was trained using in situ soil moisture measurements from eight sensors in combination with EF derived from Landsat imagery, meteorological, soil, and topography data as predictor variables. RF is an ensemble-based machine learning approach that uses multiple classifications and regression trees in sample selection utilizing the bootstrapping method, which is referred to as “bagging” (Breiman 1996, 2001). The detailed RF flowchart used in this study is shown in Fig. 4. With the bootstrap in the various decision trees, the selection of the variables is randomized in this method, with only a portion of the samples being selected in each of the multiple trees. RF is capable of performing both classification and regression processes (Fig. 4). The number of trees (tree) and the number of features (mtry) are the two most influential factors in the RF algorithm. According to its general definition, it is a method used to improve the precision of models when compared to linear regression, because it is resistant to multicollinearity and is capable of solving complex interactions between the predictor and explanatory variables (Drobníč et al. 2020). During the generation of RF samples, the input data are grouped into rows (called samples) and columns (called features) with respect to row sampling; the technique is to use replacement sampling, which means that some samples may occur several times in the training set of a tree or may never exist at all in the training set (Meyer et al. 2019).

The RF approach was implemented using the R package CAST developed by Hanna Meyer (<https://cran.r-project.org/web/packages/CAST/index.html>). To estimate the RZSM

using RF, a random interpolation of randomly selected points from each sensor location within the time-series data from 2018 to 2020 was used. With data from all eight TDR soil moisture sensor stations on the Ripperdan Ranch integrated, a single RF model was developed and applied to the spatial prediction of soil moisture over the entire vineyard. Random samples were taken from the daily time-series data at each site to perform RF interpolation (Fig. 4). The samples from each station were generated using a proportion of 70–80% of the daily time-series measurements at each site. These were then integrated into a single training set for the purpose of developing an RF model. For interpolation, the value of the *n*tree option was set to 500 trees. The optimization of the RF model was carried out by modifying the *m*try value from 2 to 10 for each training set proportion tested. Covariates or predictor variables utilized in the construction of an RF regression model were meteorological data, soil properties, and EF derived from Landsat imagery. During the training phase of the model, daily linear interpolated EF and normalized difference vegetation index (NDVI) values between two satellite dates from 123 images taken over 3 years were utilized as predictor variables. Also, meteorological factors, such as daily average solar radiation, reference evapotranspiration, wind speed, maximum, minimum, and daily average air temperature, and relative humidity respectively, were utilized as covariates. Additionally, soil attributes such as spatial distribution of bulk density, soil temperature (three temperature sensors were installed at the flux tower site at a depth of 10 cm), and a digital elevation model derived were also employed as predictor variables.

The validation of the RF model was accomplished by the use of *k*-fold cross-validation (CV). CV is widely used to estimate the performance of a model in the context of data that has not been utilized for model training (Meyer et al. 2018, 2019). The CV procedure involves training models on a large number of occasions (*k* models), and in each model run, the data from onefold are set aside and used not for model training, but for model validation instead. This way, the model's performance can be assessed using data that were not used in the model's training (Meyer et al. 2018, 2019). We used the index argument to account for data dependencies by leaving the entire dataset from one sensor location out. A random *k*-fold CV contains data points from each sensor site that are contained in each of the folds with the maximum degree of certainty.

Performance evaluation criteria

Different statistical goodness-of-fit indicators were employed to assess the errors between modeled and observed

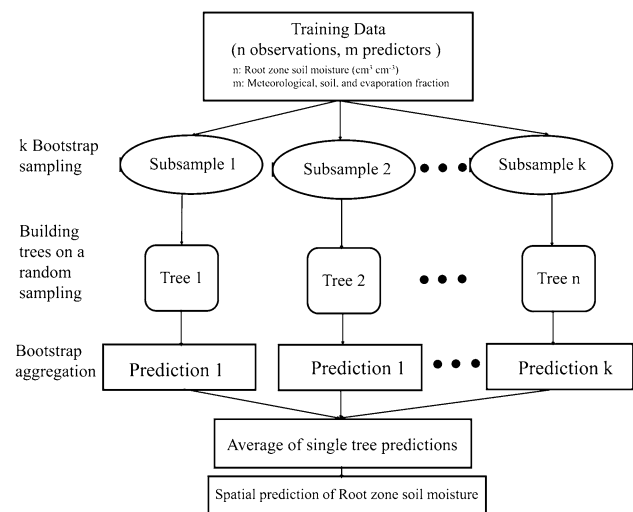


Fig. 4 Workflow diagram for the Random Forest (RF) model. The construction of regression trees is based on a large number of bootstrap samples. Each tree is formed by picking the datasets from each subsample and putting them together. Each tree's predictions were averaged to provide a single value for the purpose of building a model, which was then used to predict the spatial distribution of root zone soil moisture

soil moisture values from the three models. The difference in RZSM values between observed (in situ) and modeled values was quantified using root-mean-square error (RMSE), index of agreement (*d*), and coefficient of determination (R^2) (Huryna et al. 2019); see Eqs. (11)–(14)

$$\text{RMSE} = \left[\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \right]^{1/2} \quad (11)$$

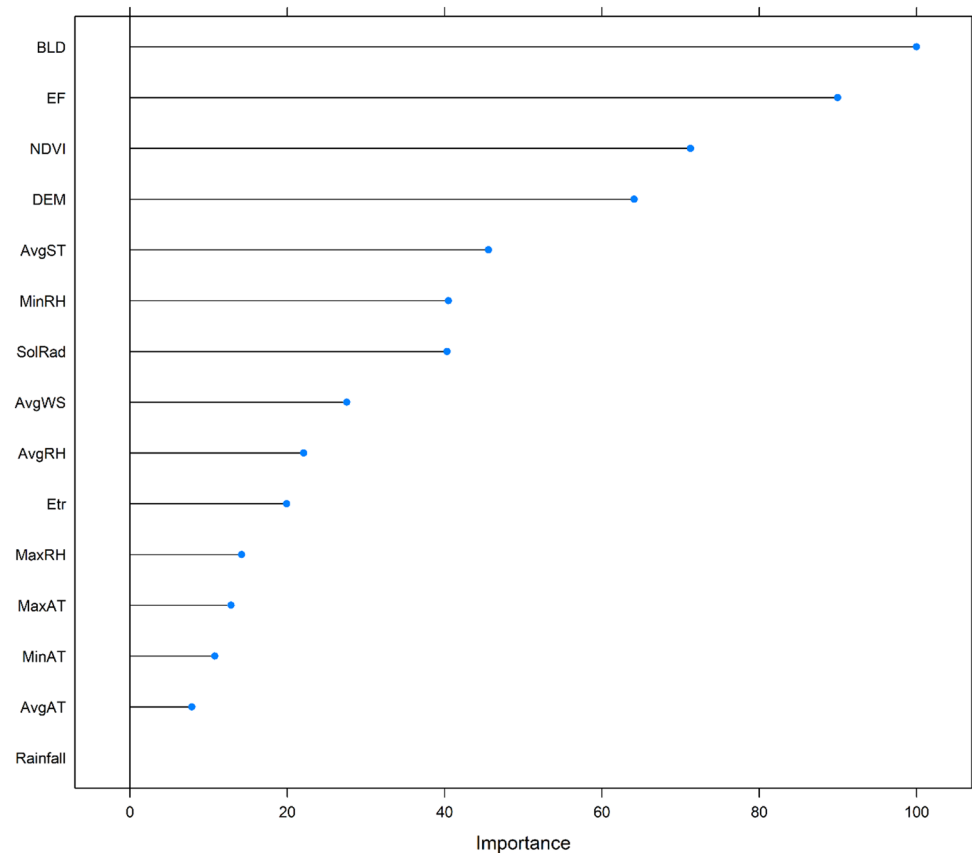
$$d = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}, 0 \leq d \leq 1 \quad (12)$$

$$R^2 = 1 - \frac{\sum (P_i - O_i)^2}{\sum (P_i - \bar{O})^2} \quad (13)$$

$$\text{bias} = \frac{\sum_{i=1}^n (P_i - O_i)}{n}, \quad (14)$$

where P_i is modeled value, O_i is the observed value, \bar{O} is the average of the observed values, and *n* is the number of observations.

Fig. 5 Variable importance of RF model used in this study. Here, BLD is soil bulk density at each sensor location, AvgST is the daily average soil temperature, EF is the evaporation fraction, NDVI is normalized difference vegetation index, DEM is the digital elevation model, SolRad is the daily solar radiation, MinRH, MaxRH, and AvgRH are the minimum, maximum, and daily average relative humidity respectively, MaxAT, MinAT, and AvgAT are minimum, maximum, and daily average air temperatures, respectively, Etr is reference evapotranspiration, AvgWS is daily average wind speed, and Rainfall is the sum of precipitation in that particular day



Results and discussion

Model tuning and variable significance in the random forest model

The RF model built with varying amounts of training data sets showed the lowest RMSE when 70–80% of the total data was used, with R^2 of 0.84 and RMSE of $0.02 \text{ cm}^3 \text{ cm}^{-3}$. As a result, by setting ntree equal to 500, the 80% training set was chosen for further evaluation of the model. The results of the tenfold cross-validation demonstrated that, compared to meteorological factors, soil bulk density, EF, NDVI, and average soil temperature had the greatest impact on the RF model accuracy (Fig. 5). The soil bulk density and the soil temperature were the most critical parameters that impact the amount of moisture present in the soil within the root zone. Water movement in the root zone was influenced by the dominant sandy soil found at the study site, which has a high bulk density and thus a low porosity due to the coarse texture that affects the saturated water content. Additionally, at low temperatures, root water uptake may be decreased due to lower evaporative demand, and that is also accompanied by reductions in the photosynthetic rate in the grapevines. Topography and meteorological variables had minor effect on the RF model accuracy (Fig. 5), which is not surprising

given that the study vineyard is relatively flat. The EF variable is strongly connected to crop ET that drives RZSM. Reference ET was found to significantly influence the RF model predictions (Fig. 5). Even though the NDVI is one of the variables used to derive EF, when used as an independent covariate, it had a major impact on the RF model, which was not surprising given that canopy size affects light interception and consequently crop water use. In light of the fact that precipitation has a direct impact on soil moisture, but the rainfall parameter did not show a high rank on the list of important variables, it is possible that the Mediterranean rainfall pattern in which there is negligible rainfall during the summer growing season explains this observation.

Validation of the evaporation fraction and evapotranspiration

The evaporation fraction (EF) was determined using the energy balance components from the EC tower, which were then correlated with the EF derived from the pySEBAL model at the available dates (Fig. 6). The correlations revealed that the EC tower measured EF agreed well with remote-sensing-derived EF. This was critical, since EF is one of the important covariates (Fig. 5) in the training of the RF model for soil moisture predictions within the study

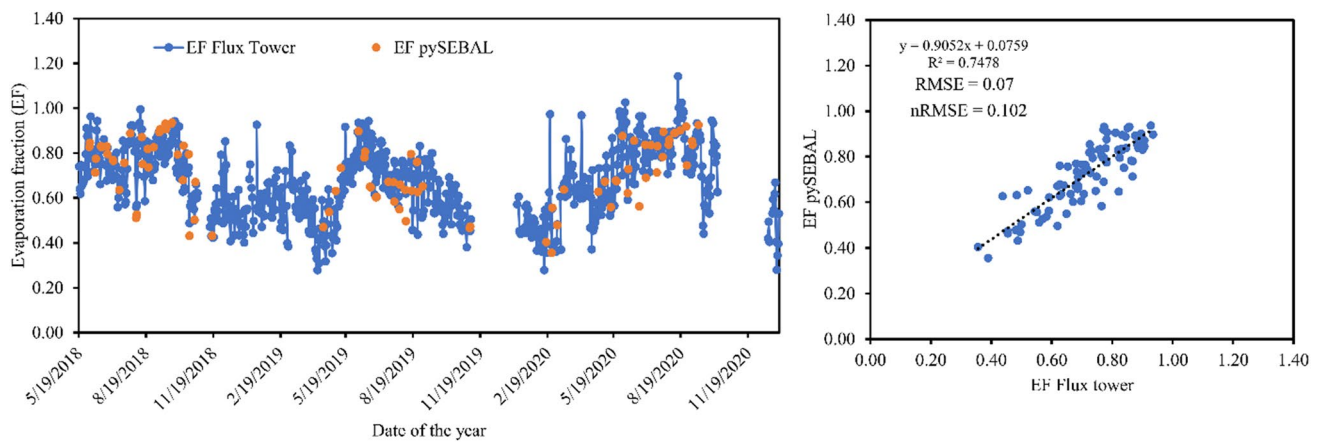


Fig. 6 Evaporation fraction (EF) values obtained from the eddy covariance flux tower and those derived using pySEBAL were compared for the three crop seasons, with scatter plots reflecting the root-

mean-square error (RMSE), normalized RMSE (nRMSE), and coefficient of determination (R^2)

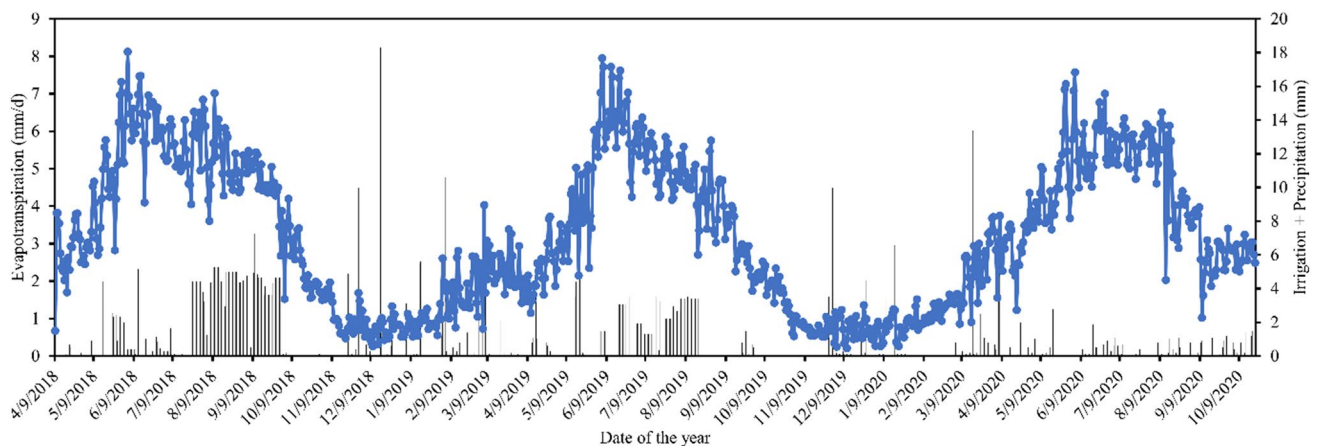


Fig. 7 Daily evapotranspiration from eddy covariance flux tower and combined irrigation and precipitation for the three seasons from 2018 to 2020 in a vineyard at the Ripperdan Ranch near Madera, CA

area. Furthermore, the trained model was used to make spatial predictions by providing satellite-derived spatial EF data. The RMSE and nRMSE (normalized root-mean-square error) between measured and modeled EF were found to be 0.07 and 0.102, respectively, with a coefficient of determination (R^2) of 0.75. The amount of irrigation was implemented at each sensor location separately in the research block. The daily dynamics of ET with combined irrigation and precipitation at the flux tower location are depicted in Fig. 7. The ET patterns indicate that ET is low at the beginning of the crop season and gradually increases with crop growth as transpiration rates increase, and then begin to decrease at the end of the crop season. In addition, we can detect a relationship between soil moisture, the amount of water applied through irrigation and precipitation, and ET through Figs. 7

and 8, respectively. For example, during the middle part of the season, when ET is at its peak, RZSM is depleted at a faster rate and reaches its lowest levels. Also, soil moisture reaches its peak during the winter, when ET is low and precipitation is high (California has a Mediterranean climate with dominant winter rainfall). Additionally, it is reasonable to infer that the EF contains indirect information on irrigation, which has an impact on soil water dynamics.

Root zone soil moisture dynamics

The daily average soil moisture in the root zone considered as the top 60 cm in this study was acquired from soil moisture sensors at eight sensor locations. These data were used to train the RF model and to validate predicted soil

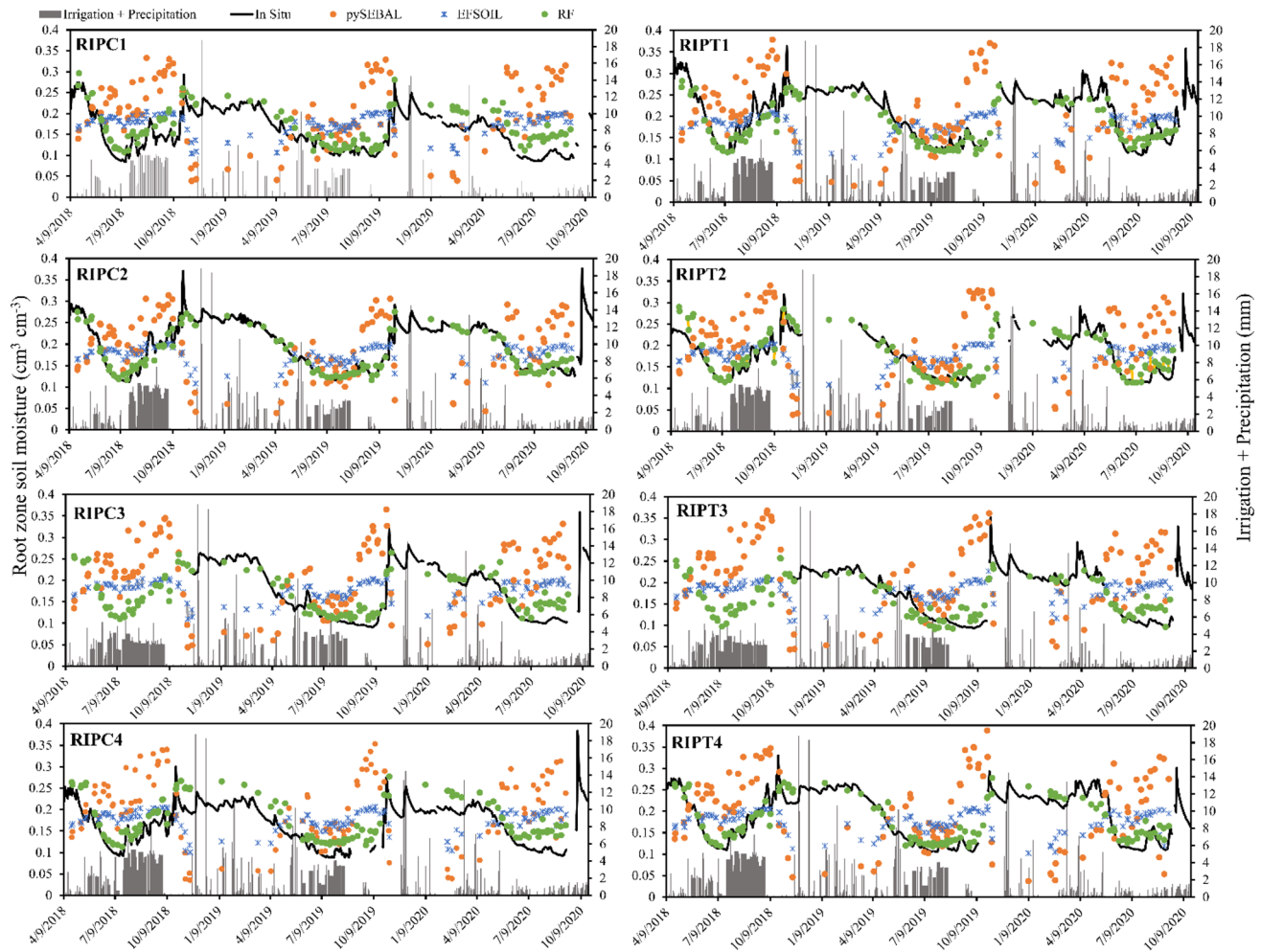


Fig. 8 Time-series plots of root zone soil moisture estimates from pySEBAL (orange dots), EFSOIL (blue stars), RF models (green dots), and in situ measurements (black solid lines) at eight sensor locations (RIPC1, RIPC2, RIPC3, RIPC4, RIPT1, RIPT2, RIPT3,

and RIPT4) and combined irrigation and precipitation (histogram) at each sensor location are shown on secondary axis for the three growing seasons of grapevines

moisture from the three models to measured values. The daily observed and predicted RZSM dynamics from the three models pySEBAL, EFSOIL, and RF are shown in Fig. 8 for the vineyard growing seasons from 2018 to 2020. It should be noted that the RZSM correlations were limited to the times when the EF data were available over the 3-year studied period (i.e., 123 Landsat-derived EF data were considered in 3-year time frame). When comparing the three models with the observed RZSM, the soil moisture demonstrated significant temporal variability. During each of the three growing seasons, the RF model showed high correlations with in situ RZSM data. However, soil moisture estimated by the pySEBAL model either overestimated or underestimated soil moisture when compared to in situ data from all sensor locations. Similar results were found for the EFSOIL model at all the eight sensor locations. As expected, the in situ RZSM effectively tracked precipitation;

for example, when the amount of rainfall was high, e.g., in 2019 and 2020, soil moisture content within the vineyard was high. The RF model was able to capture these wetting and drying cycles better than the pySEBAL and EFSOIL models (Fig. 8). Under dry conditions, the performance of both pySEBAL and EFSOIL models agreed closely with in situ sensor data, but the RF model produced more accurate predictions across all conditions. It was encouraging to see that, while in situ data for the 2018 growth period were not available for the RIPC3 and RIPT3 sensor locations, the RF model predictions showed extremely consistent and precise dynamics, similar to the dynamics observed at the other sensor locations during this period. This observation provided confidence in predicting the RZSM at this vineyard, which was useful for future implementations of the RF model.

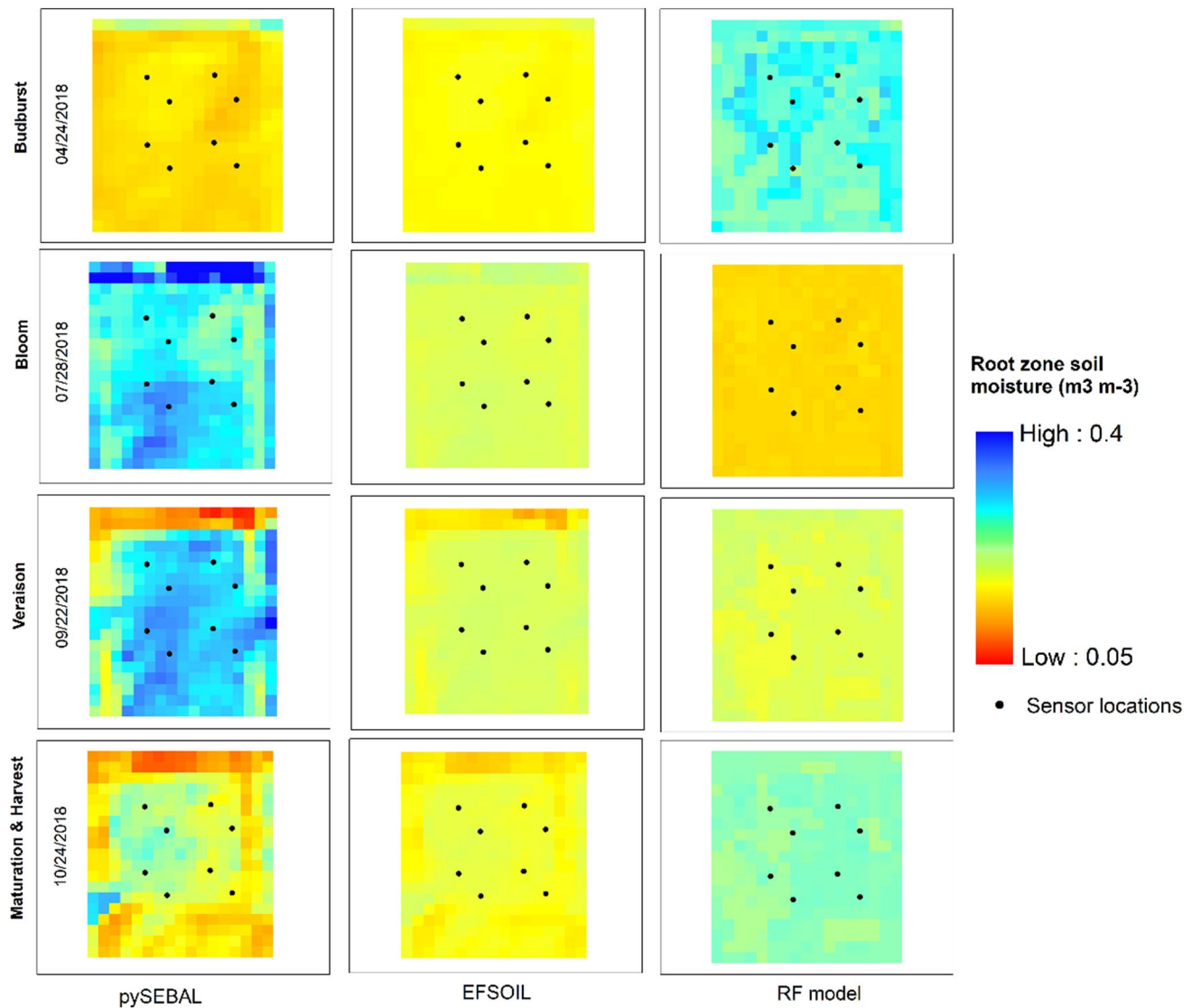


Fig. 9 On the growing season of 2018, an example of the spatial distribution of root zone soil moisture from the three models pySEBAL, EFSOIL, and RF for a certain day of the crop season. The crop

growth stages are divided as Budburst, Bloom, Veraison, and Maturation & Harvest

The goodness-of-fit statistical indicators obtained from the comparison of observed and predicted root zone soil moisture at each sensor location are reported in Table 2. The results from the RF model had a high R^2 (>0.80), low RMSE ($0.012\text{--}0.036\text{ cm}^3\text{ cm}^{-3}$), low mean bias error (-0.008 to 0.031), and a high index of agreement value (>0.86) from all eight sensor locations, indicating that a data-driven ML-based method such as RF is capable of accurately predicting RZSM across the study this site. The pySEBAL model had the weakest performance, with low R^2 (<0.04), high RMSE (>0.09 to $0.12\text{ cm}^3\text{ cm}^{-3}$), high bias (>0.02 to 0.052), and a low index of agreement value ($0.18\text{--}0.31$) from all of the monitoring locations. However, when compared to pySEBAL, the EFSOIL model predictions were

only marginally better, with low R^2 (<0.09), high RMSE ($0.05\text{--}0.07\text{ cm}^3\text{ cm}^{-3}$), high mean bias (-0.003 to 0.034), and a very low index agreement value ($0.09\text{--}0.22$) at all eight soil moisture sensor locations. Within the vineyard, soil moisture patterns were controlled by a variety of factors including soil evaporation and crop root water uptake. Compared to the data-driven RF-based approach, the semi-empirical models based on pySEBAL and EFSOIL were unable to accurately predict the RZSM values. As discussed earlier, the RF model was developed using a number of covariates such as meteorological data and soil characteristics that have high correlations with the in situ soil moisture data. The findings from this study indicate that the RF model can be used to accurately predict soil water status to guide

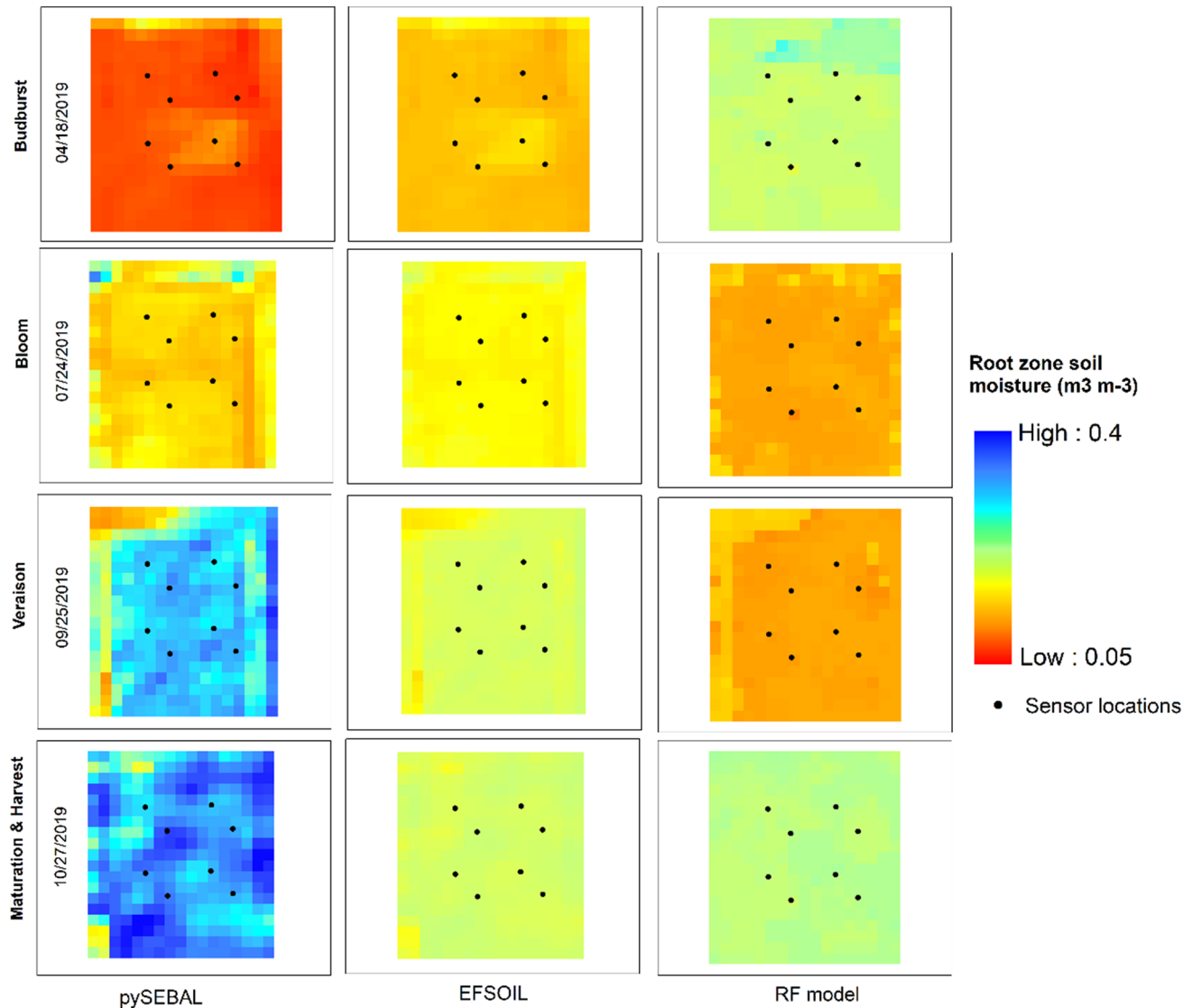


Fig. 10 On the growing season of 2019, an example of the spatial distribution of root zone soil moisture from the three models pySEBAL, EFSOIL, and RF for a certain day of the crop season. The crop growth divided as Budburst, Bloom, Veraison, and Maturation & Harvest stages

irrigation scheduling decisions or in evaluating root zone soil water balance.

Spatial root zone soil moisture dynamics

The spatial distribution pattern of soil moisture predicted by the pySEBAL, EFSOIL, and RF models for each grapevine crop growth stage (single day for each stage) from 2018 to 2020 was evaluated at a 30 m spatial resolution, respectively, as shown in Figs. 9, 10 and 11. The spatial patterns were evaluated in relation to the growth stages of the grapevine. Several processes and events occur during the annual growth cycle of grapevines; however, the major ones are classified into four categories: budburst, bloom, veraison, fruit maturation, and harvest. In each of the 3 years, the

spatial variability in RZSM differed from the three models in terms of magnitude and crop stage. The pySEBAL model revealed consistent patterns at the beginning (budburst) and end (harvest) of the crop season. However, the pySEBAL model exhibited very high soil moisture levels that were not practical (compared to soil moisture sensor values) during the Veraison stage, given the fact that the farmer applied deficit irrigation to allow light stress on crops to increase grape quality at this growth stage, as shown in Fig. 6. It is worth noting that the pySEBAL modeling framework was developed for predicting ET over large spatial scales and soil is derived as an auxiliary output. This might explain why it did not do very well at the vineyard scale. The sensor measurements of RZSM revealed a decrease in the observed RZSM over the months of June–July across all three crop

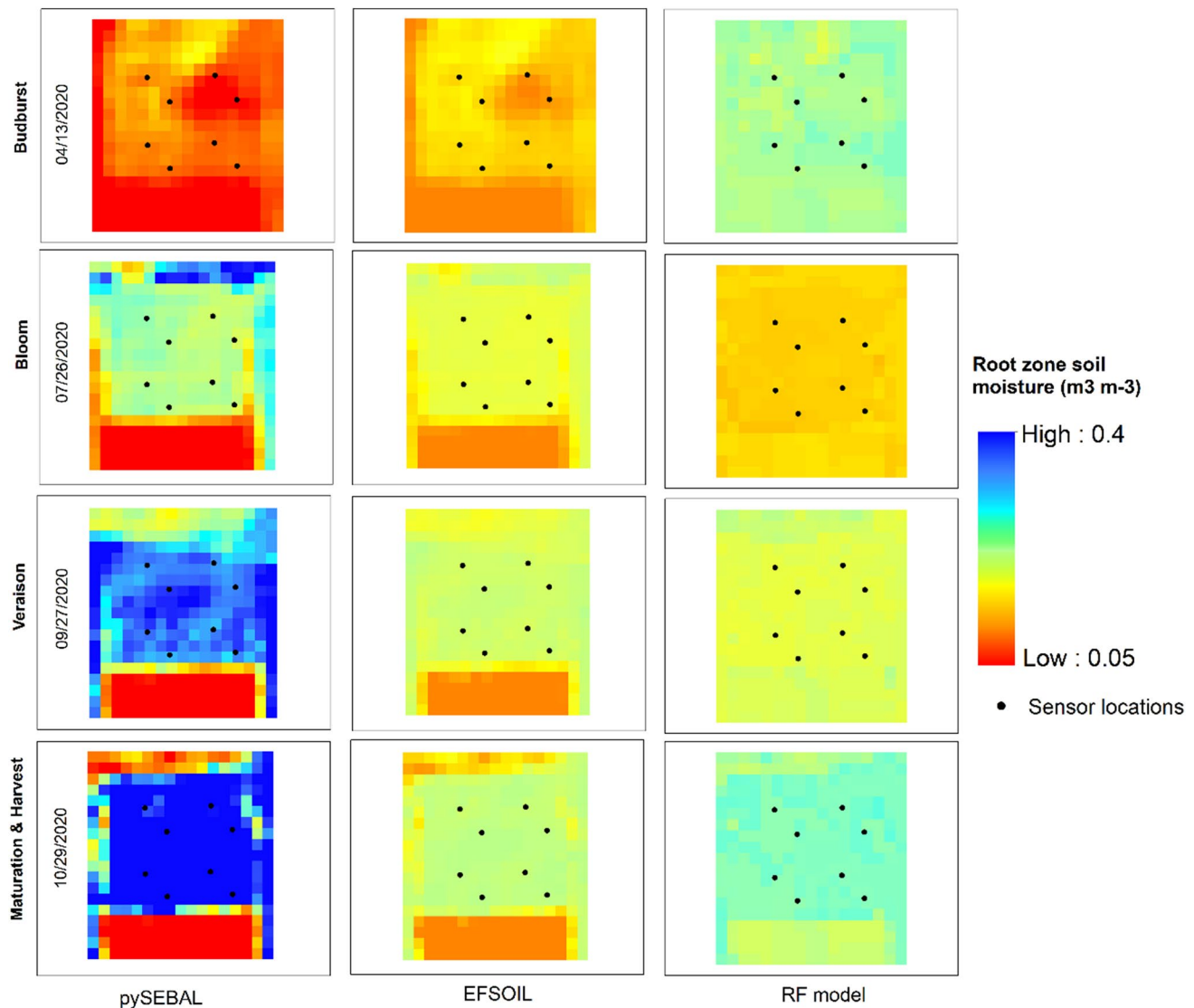


Fig. 11 On the growing season of 2020, an example of the spatial distribution of root zone soil moisture from the three models pySEBAL, EFSOIL, and RF for a certain day of the crop season. The crop growth divided as Budburst, Bloom, Veraison, and Maturation & Harvest stages

seasons (Fig. 8). Furthermore, the RF model has enhanced stability during both well-watered and crop stress growth stages, and performed well both spatially and temporally (Figs. 8, 9, 10 and 11). The RF-based data-driven model, which was trained solely on point location data from the TDR soil moisture sensors, provided accurate predictions at the sensor locations. The spatial distributions of RZSM in the EFSOIL model had constant soil moisture values throughout the crop's growing season, which may be attributed to the fact that the field capacity and wilting point do not vary from season to season. Despite the fact that a spatial resolution of 30 m was insufficient to reliably predict soil moisture in response to crop phenological phases, spatially distributed estimates of soil moisture from the RF model can be used to refine site-specific irrigation scheduling in

vineyards and other high-value crops by considering the location in addition to irrigation timing and amount.

Future possibilities of Random Forest model

Using random forest modeling for spatial prediction, regression, and classification with complex data sets in a variety of subjects has become increasingly popular in recent years. This is due to the fact that it has a computational advantage over other regression models and is simple to implement. Because this model is heavily reliant on the dependent variable and covariates, the RF predictions are always within a reasonable range of the observed data, and the values of the tuning parameters are insensitive to the model's parameters. The drawback of RF approaches is they require extremely

high densities of in situ data within a given study region to be properly trained; appropriate data sets are not always available in some regions. For example, obtaining high spatial and temporal data on soil moisture is extremely difficult in some regions due to the high cost of the sensors and the need for more frequent and proper maintenance of the sensors. Aside from that, when creating a large number of trees to train the model, requires significantly more computational power and resources, as well as a significant amount of time to train the model to make decisions based on the majority of tree votes. Future research should place an emphasis on the spatial predictions of different variables in data-poor regions, as well as the development of more accurate validation methods. In addition, future work should explore developing Cyberphysical infrastructures that combine low cost sensors, and Internet of Things technologies with predictive power of RF-based machine learning models to enhance technology adoption among users.

Conclusions

This study compared two semi-empirical approaches that use the evaporation fraction (EF) from remote sensing to a machine learning data-driven approach that uses Random Forest (RF) for predicting spatial-temporal root zone soil moisture distribution in a vineyard. When sufficient observed data for training are available, data-driven models based on machine learning can be developed that accurately predict RZSM and are less computationally expensive compared to process-based models. RF showed the highest agreement with observed soil moisture compared to the semi-empirical models. Soil bulk density, soil temperature, and EF were the most influential covariates for predicting spatially distributed root zone soil moisture within the vineyard. The semi-empirical analytical models pySEBAL and EFSOIL, which were tested in this work, were unable to accurately predict the root zone soil moisture dynamics at the soil moisture sensor locations; depending on the wet and dry conditions, they were either overestimating or underestimating the RZSM. During the crop stress period, the pySEBAL model predicated very high spatial RZSM values which were not comparable to sensor data, but during budburst and harvesting growth stages, the model predicted reasonable spatial distributions throughout the studied area. With some algorithmic tweaks or parameterization, the pySEBAL may potentially be enhanced to produce accurate spatially distributed RZSM predictions, which is extremely valuable for site-specific irrigation scheduling and soil water balance evaluations, because it is based on Landsat imagery that are freely available. In summary, the RF approach that produced good predictions of RZSM though demonstrated in a vineyard, the framework can be applied to other cropping

systems or conditions where accurate spatially distributed predictions of root zone soil moisture are needed as long as there is adequate soil moisture monitoring data over the range of soil textures within the field to train the machine learning algorithm. Combining remote sensing with machine learning techniques has the potential to enhance precision agricultural water management.

Acknowledgements Funding and logistical support for the GRAPEX project were provided by E. & J. Gallo Winery and from the NASA Applied Sciences-Water Resources Program (Grant No. NNH17AE39I). In addition, this research was supported in part by the U.S. Department of Agriculture, Agricultural Research Service. We also want to thank the staff of Viticulture, Chemistry and Enology Division of E. & J. Gallo Winery for the logistical support in helping maintain field instrumentation and the cooperation of the vineyard management staff at the Ripperdan Ranch for coordinating field operations during this study. This work was also partially funded by USDA NIFA (Award # 2021-68012-35914). USDA is an equal opportunity provider and employer.

Funding This article was funded by National Institute of Food and Agriculture, Arizona Space Grant Consortium (Grant No. NNH17AE39I).

Declarations

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adab H, Morbidelli R, Saltalippi C et al (2020) Machine learning to estimate surface soil moisture from remote sensing data. *Water (switzerland)* 12:1–28. <https://doi.org/10.3390/w12113223>
- Al-Mukhtar M (2016) Modelling the root zone soil moisture using artificial neural networks, a case study. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-016-5929-2>
- Alfieri JG, Kustas WP, Prueger JH et al (2019) A multi-year intercomparison of micrometeorological observations at adjacent vineyards in California's Central Valley during GRAPEX. *Irrig Sci* 37:345–357. <https://doi.org/10.1007/s00271-018-0599-3>
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56
- Allen RG, Tasumi M, Morse A et al (2007) Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—applications. *J Irrig Drain Eng*. [https://doi.org/10.1061/\(asce\)0733-9437\(2007\)133:4\(395\)](https://doi.org/10.1061/(asce)0733-9437(2007)133:4(395))

- Allen RG, Tasumi M, Morse A, Trezza R (2005a) A landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning. *Irrig Drain Syst* 19:251–268. <https://doi.org/10.1007/S10795-005-5187-Z/EMAIL/CORRESPONDENT/C1/NEW>
- Bastiaanssen WGM (1995) Regionalization of surface flux densities and moisture indicators in composite terrain: a remote sensing approach under clear skies in Mediterranean climates
- Bastiaanssen WGM, Allen RG, Droogers P et al (2007) Twenty-five years modeling irrigated and drained soils: State of the art. *Agric Water Manag* 92:111–125. <https://doi.org/10.1016/J.AGWAT.2007.05.013>
- Bastiaanssen WGM, Menenti M, Feddes RA, Holtslag AAM (1998a) A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. *J Hydrol.* [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4)
- Bastiaanssen WGM, Pelgrum H, Droogers P et al (1997) Area-average estimates of evaporation, wetness indicators and top soil moisture during two golden days in EFEDA. *Agric for Meteorol.* [https://doi.org/10.1016/S0168-1923\(97\)00020-8](https://doi.org/10.1016/S0168-1923(97)00020-8)
- Bastiaanssen WGM, Pelgrum H, Wang J et al (1998b) A remote sensing surface energy balance algorithm for land (SEBAL): 2. Validation. *J Hydrol.* [https://doi.org/10.1016/S0022-1694\(98\)00254-6](https://doi.org/10.1016/S0022-1694(98)00254-6)
- Bauer-Marschallinger B, Freeman V, Cao S et al (2019) Toward global soil moisture monitoring with Sentinel-1: harnessing assets and overcoming obstacles. *IEEE Trans Geosci Remote Sens* 57:520–539. <https://doi.org/10.1109/TGRS.2018.2858004>
- Boretti A, Rosa L (2019) Reassessing the projections of the World Water Development Report. *npj Clean Water* 2:1–6. <https://doi.org/10.1038/s41545-019-0039-9>
- Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140. <https://doi.org/10.1007/bf00058655>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Budyko MI, Zubenok LI (1961) Determination of evaporation from the land surface. *Izv Akad Nauk SSSR, Ser Geogr* 6:3–17
- Carranza C, Nolet C, Pezij M, van der Ploeg M (2021) Root zone soil moisture estimation with Random Forest. *J Hydrol* 593:125840. <https://doi.org/10.1016/j.jhydrol.2020.125840>
- Cosh MH, Ochsner TE, McKee L et al (2016) The soil moisture active passive Marena, Oklahoma, IN SITU SENSOR TESTBED (SMAP-MOISST): test bed design and evaluation of in situ sensors. *Vadose Zo J* 15:1–11. <https://doi.org/10.2136/vzj2015.09.0122>
- D'Odorico P, Porporato A (2004) Preferential in soil moisture and climate dynamics. *Proc Natl Acad Sci U S A* 101:8848–8851. <https://doi.org/10.1073/pnas.0401428101>
- Drobníč F, Kos A, Pustišek M (2020) On the interpretability of machine learning models and experimental feature selection in case of multicollinear data. *Electron* 9:761. <https://doi.org/10.3390/electronics9050761>
- Engman ET (1991) Applications of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sens Environ* 35:213–226. [https://doi.org/10.1016/0034-4257\(91\)90013-V](https://doi.org/10.1016/0034-4257(91)90013-V)
- Gao X, Wu P, Zhao X et al (2014) Effects of land use on soil moisture variations in a semi-arid catchment: Implications for land and agricultural water management. *L Degrad Dev* 25:163–172. <https://doi.org/10.1002/ldr.1156>
- González-Teruel JD, Torres-Sánchez R, Blaya-Ros PJ et al (2019) Design and calibration of a low-cost SDI-12 soil moisture sensor. *Sensors (switzerland)*. <https://doi.org/10.3390/s19030491>
- Hassan-Esfahani L, Torres-Rua A, Jensen A, McKee M (2015) Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sens* 7:2627–2646. <https://doi.org/10.3390/rs70302627>
- Hessels T, van Opstal J, Trambauer P et al (2017) pySEBAL Version 3.3. 7.
- Hirschi M, Mueller B, Dorigo W, Seneviratne SI (2014) Using remotely sensed soil moisture for land-atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. *Remote Sens Environ* 154:246–252. <https://doi.org/10.1016/j.rse.2014.08.030>
- Huntingford C, Jeffers ES, Bonsall MB et al (2019) Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ Res Lett* 14:124007. <https://doi.org/10.1088/1748-9326/ab4e55>
- Huryňa H, Cohen Y, Karnieli A et al (2019) Evaluation of TsHARP utility for thermal sharpening of Sentinel-3 satellite images using Sentinel-2 visual imagery. *Remote Sens.* <https://doi.org/10.3390/rs11192304>
- Jaafar HH, Ahmad FA (2020) Time series trends of Landsat-based ET using automated calibration in METRIC and SEBAL: The Bekaa Valley, Lebanon *Remote Sens Environ* 238:111034. <https://doi.org/10.1016/j.rse.2018.12.033>
- Jackson TJ (1993) III. Measuring surface soil moisture using passive microwave remote sensing. *Hydrol Process* 7:139–152. <https://doi.org/10.1002/hyp.3360070205>
- Knipper KR, Kustas WP, Anderson MC et al (2019) Using high-spatiotemporal thermal satellite ET retrievals for operational water use and stress monitoring in a California vineyard. *Remote Sens* 11:2124. <https://doi.org/10.3390/rs11182124>
- Kornelsen KC, Coulibaly P (2014) Root-zone soil moisture estimation using data-driven methods. *Water Resour Res* 50:2946–2962. <https://doi.org/10.1002/2013WR014127>
- Kustas WP, Anderson MC, Alfieri JG et al (2018) The grape remote sensing atmospheric profile and evapotranspiration experiment. *Bull Am Meteorol Soc* 99:1791–1812. <https://doi.org/10.1175/BAMS-D-16-0244.1>
- Laipelt L, Henrique Bloedow Kayser R, Santos Fleischmann A et al (2021) Long-term monitoring of evapotranspiration using the SEBAL algorithm and Google Earth Engine cloud computing. *ISPRS J Photogramm Remote Sens* 178:81–96. <https://doi.org/10.1016/j.isprs.2021.05.018>
- Lange H, Sippel S (2020) *Machine Learning Applications in Hydrology*. Springer, Cham, pp 233–257
- Li J, Heap AD, Potter A, Daniell JJ (2011) Application of machine learning methods to spatial interpolation of environmental variables. *Environ Model Softw* 26:1647–1659. <https://doi.org/10.1016/j.envsoft.2011.07.004>
- Lin BB, Egerer MH, Liere H et al (2018) Soil management is key to maintaining soil moisture in urban gardens facing changing climatic conditions. *Sci Rep* 8:1–9. <https://doi.org/10.1038/s41598-018-35731-7>
- Meyer H, Reudenbach C, Hengl T et al (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ Model Softw* 101:1–9. <https://doi.org/10.1016/j.envsoft.2017.12.001>
- Meyer H, Reudenbach C, Wöllauer S, Nauss T (2019) Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecol Modell* 411:108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
- Meyers JN, Kisekka I, Upadhyaya SK, Michelon G (2019) Development of an artificial neural network approach for predicting plant water status in almonds. *Trans ASABE* 62:19–32. <https://doi.org/10.13031/trans.12970>
- Montaldo N, Albertson JD, Mancini M, Kiely G (2001) Robust simulation of root zone soil moisture with assimilation of surface soil moisture data. *Water Resour Res* 37:2889–2900. <https://doi.org/10.1029/2000WR000209>

- Nutini F, Boschetti M, Candiani G et al (2014) Evaporative fraction as an indicator of moisture condition and water stress status in semi-arid rangeland ecosystems. *Remote Sens* 6:6300–6323. <https://doi.org/10.3390/rs6076300>
- Peddinti SR, Hopmans JW, Najm MA, Kisekka I (2020a) Assessing effects of salinity on the performance of a low-cost wireless soil water sensor. *Sensors (switzerland)* 20:1–14. <https://doi.org/10.3390/s20247041>
- Peddinti SR, Kambhammettu BVNP, Lad RS et al (2020b) A macroscopic soil-water transport model to simulate root water uptake in the presence of water and disease stress. *J Hydrol* 587:124940. <https://doi.org/10.1016/j.jhydrol.2020.124940>
- Peddinti SR, Kambhammettu BVNP, Ranjan S et al (2018) Modeling soil-water-disease interactions of flood-irrigated mandarin orange trees: Role of root distribution parameters. *Vadose Zo J*. <https://doi.org/10.2136/vzj2017.06.0129>
- Petropoulos GP, Ireland G, Barrett B (2015) Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Phys Chem Earth* 83–84:36–56
- Petropoulos GP, Sandric I, Hristopulos D, Carlson TN (2020) Evaporative fluxes and surface soil moisture retrievals in a mediterranean setting from sentinel-3 and the “simplified triangle.” *Remote Sens* 12:1–20. <https://doi.org/10.3390/rs12193192>
- Pradhan NR (2019) Estimating growing-season root zone soil moisture from vegetation index-based evapotranspiration fraction and soil properties in the Northwest Mountain region, USA. *Hydrol Sci J* 64:771–788. <https://doi.org/10.1080/02626667.2019.1593417>
- Sabater JM, Jarlan L, Calvet JC et al (2007) From near-surface to root-zone soil moisture using different assimilation techniques. *J Hydrometeorol* 8:194–206. <https://doi.org/10.1175/JHM571.1>
- Sadri S, Pan M, Wada Y et al (2020) A global near-real-time soil moisture index monitor for food security using integrated SMOS and SMAP. *Remote Sens Environ* 246:111864. <https://doi.org/10.1016/j.rse.2020.111864>
- Scott CA, Bastiaanssen WGM, Ahmad M, ud D, (2003) Mapping root zone soil moisture using remotely sensed optical imagery. *J Irrig Drain Eng* 129:326–335. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2003\)129:5\(326\)](https://doi.org/10.1061/(ASCE)0733-9437(2003)129:5(326))
- Sheffield J, Wood EF (2008) Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J Clim* 21:432–458. <https://doi.org/10.1175/2007JCLI1822.1>
- Tasumi M (2003) Progress in operational estimation of regional evapotranspiration using satellite imagery
- Trezza R (2002) Evapotranspiration using a satellite-based surface energy balance with standardized ground control
- Waters R, Allen R, Tasumi M et al (2002) SEBAL; advanced training and users manual. 1–98
- Wigneron JP, Olioso A, Calvet JC, Bertuzzi P (1999) Estimating root zone soil moisture from surface soil moisture data and soil-vegetation-atmosphere transfer modeling. *Water Resour Res* 35:3735–3745. <https://doi.org/10.1029/1999WR900258>
- Xue J, Bali KM, Light S et al (2020) Evaluation of remote sensing-based evapotranspiration models against surface renewal in almonds, tomatoes and maize. *Agric Water Manag*. <https://doi.org/10.1016/j.agwat.2020.106228>
- Yu Z, Liu D, Lü H et al (2012) A multi-layer soil moisture data assimilation using support vector machines and ensemble particle filter. *J Hydrol* 475:53–64. <https://doi.org/10.1016/j.jhydrol.2012.08.034>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.