SAR2ET: End-to-End SAR-Driven Multisource ET Imagery Estimation Over Croplands

Samet Cetin, Berk Ülker, Esra Erten, *Senior Member, IEEE*, and Ramazan Gokberk Cinbis

***Abstract*—Evapotranspiration (ET) is a crucial parameter in agriculture as it plays a vital role in managing water resources, monitoring droughts, and optimizing crop yields across different ecosystems. Given its significance in crop growth, it is essential to measure ET accurately and continuously to conduct precise analyses in agriculture. However, the continuous monitoring of ET changes is very challenging: while in-situ measurements are costly and not feasible for covering a wide geography, remote sensing-based ET products are typically dependent on optical satellites that cannot operate and transmit data under certain weather conditions, especially in the presence of clouds. In this article, we present the first comprehensive study on predicting ET from synthetic aperture radar (SAR) imagery, which we refer to as SAR2ET. Our work is motivated by the fact that SAR has the critical advantages of being all-weather available and sensitive to crop and soil changes. In handling the SAR2ET problem, we additionally incorporate nonoptical meteorological and topographical input data from auxiliary data sources. We approach SAR2ET as a multimodal image-to-image translation task, for which we train a UNet-shaped network. To evaluate the effectiveness of SAR-based ET predictions, we construct a benchmark dataset over a large geographical region with image samples covering a whole agriculture season. Our experimental findings on this dataset suggest that first, the proposed approach leads to strong results, second, valuable information can be extracted from both SAR and auxiliary data sources, and finally, SAR2ET is overall a promising research direction toward obtaining data-driven year-round ET estimates. The benchmark dataset will be shared publicly upon publication to stimulate future work.**

***Index Terms*—Agriculture, climate change, disaggregation, Earth engine evapotranspiration flux (EEFlux), evapotranspiration (ET), Sentinel-1, weak supervision.**

I. Introduction

EVAPOTRANSPIRATION (ET), the evaporation from the surface of water-bodies, soil and vegetation, is a key parameter in agriculture, particularly for water management and for efforts to optimize yield. ET depends on complex interaction of meteorological variables, such as air temperature, humidity, solar radiation, and wind speed over vegetated surface, making its prediction challenging [1]. Typically, field-scale ground-based measurements are used to predict ET; however, these measurements cannot consider the spatial distribution characteristics of the soil and are not enough to provide necessary information at large scale agriculture. In contrast, conventional space-based remote sensing (RS) sensors are relevant for large scale monitoring, as they obviate the demanding ground-based surveys and provide detailed information about not only the characteristics of the soil and vegetated surfaces but also meteorological variables. Hence, there are RS-based global ET products covering different complex environment, such as the Global Land Evaporation Amsterdam Model (GLEAM), the MODerate Resolution Imaging Spectroradiometer-MOD16 (MODIS), the Global Land Data Assimilation System (GLDAS), and numerical terradynamic simulation group [2]. Nevertheless, when it comes to agriculture, their usage is limited due to their spatial resolution.

Given that ET plays a critical role in crop growth, the timely and continuous measurement of ET is essential for accurate analysis in agriculture [3], [4]. Indeed, the spatio-temporal resolution of RS-based ET products is restricted by the characteristics of the thermal infrared channel of the sensors. In order to obtain high resolution ET product, accordingly high resolution land surface temperature (LST), there has been a strong interest in data fusion techniques including spatial downscaling and spatiotemporal fusion [5], [6], [7], [8]. In spatiotemporal fusion tasks, the basic hypothesis is that high resolution thermal image can be reconstructed using medium resolution thermal image acquired under clear-sky. In addition to this, downscaling methods also take into account the assumption that the relationship between the LST and the auxiliary data, i.e., land cover, soil indices, digital elevation model (DEM), etc., is not varying along the spatial scale. For both approaches, machine learning methods, particularly supervised regression methods, have densely been used to establish a transition model for characterizing these multiscale nonlinear relationships. Xue et al. [9] underlined the importance of temporal sampling of high resolution Landsat images using other sources of medium-resolution thermal imaging (i.e., MODIS and ECOSTRESS) for rainfed agriculture. Similarly, Li et al. [5] designed a downscaling coupled with spatiotemporal fusion based framework for combining different resolution thermal images for having temporally more regular LST data. Dong et al. [10] analyzed 32 geographically diverse LST data with 35 different downscaling methods including

Received 30 January 2024; revised 25 April 2024, 4 August 2024, and 7 August 2024; accepted 17 August 2024. Date of publication 21 August 2024; date of current version 5 September 2024. This work was fully funded by the 2022 Climate Change AI Innovation Grants program, hosted by Climate Change AI with the additional support of Canada Hub of Future Earth, Grant Reference No. 182 and was partially also supported by the Research Fund of Istanbul Technical University (ITU-BAP). *(Corresponding author: Esra Erten.)*

Samet Cetin and Ramazan Gokberk Cinbis are with the Department of Computer Engineering, Middle East Technical University, Ankara 06800, Türkiye.

Berk Ülker is with the Department of Electronics Engineering, Eindhoven University of Technology, 5612, AE Eindhoven, The Netherlands, and also with Agcurate, 6825, MC Arnhem, The Netherlands.

Esra Erten is with the Civil Engineering Faculty, Istanbul Technical University, Istanbul 34469, Türkiye (e-mail: eerten@itu.edu.tr).

Data is available online at https://github.com/Agcurate/SAR2ET.

Digital Object Identifier 10.1109/JSTARS.2024.3447033

different types of scaling factors and regression models. In the study by Xiao et al. [11], RF model is redesigned to derive LST for cloudy pixels by mapping input vector including water index, solar radiation factor, vegetation index, topographic information, and latitude onto an output vector LST, where the samples are from clear sky pixels.

Recently, with the improvement of neural networks (NNs), spatiotemporal fusion networks have been proposed to generate fine resolution LST images. Chen et al. [12] have used a variational interference model based on a conditional variational autoencoder to learn the temporal and nonlinear relationship between the high and medium resolution LST images. Instead, Zhang et al. [13] have improved the NN-based networks and offered multi-information fusion network based on convolutional NN and attention mechanism using MODIS LST and ERA5 meteorological variables together. Mpakairi et al. [14] have also proposed similar methodology; strategically integrates data from multiple sources, including public repositories (e.g., cropland data, field data) using deep learning NN and RF. All these studies have showed that for high resolution ET product, multisource data-based framework is promising approach and deep learning-based methods are good at modeling complex relationship among these data.

Although with these fusion techniques, the new generation sensors provide high spatial resolution thermal imaging and its products, their availability remains uncertain due to the cloud coverage problem. In addition, the lack of temporally regular thermal sensor is one of the key challenges identified in the agriculture. In this regard, recent studies have sought to explore supervised and unsupervised domain adaptation and transfer learning techniques using microwave sensors [i.e., passive radiometers and active synthetic aperture radar (SAR) images] with the capability of operating in all weather conditions, day and night, to obtain high resolution temporally regular biophysical parameters (such as LAI and NDVI), which are typically derived from optical and thermal sensors [15], [16], [17], [18], [19].

Zhang et al. [20] proposed a RF-based model to understand the complex relationship between microwave radiometer and thermal LST imaging. The model improved the timeliness of the thermal LST product by including daily soil moisture (SM) information from low resolution radiometer, specifically; advanced microwave scanning radiometer for the Earth observing system. Zhu et al. [1] combined microwave-based SM data and MODIS LST data with more physical-based model. In all those studies, the main purpose of leveraging microwave radiometer information, even with the cost of low resolution, is making use of the sensitivity of microwave sensing to dielectric constant variations. In this context, Sentinel-1 mission allows for the first time to provide dense temporal (approximately weekly) and high spatial ($\sim$ 10 m) resolution SAR data, which is not possible with passive microwave sensors due to the practical restrictions on their spatial resolution. In addition, its free data policy has made Sentinel-1 accessible for regular mapping, which is essential for understanding all the other variables in dynamic agriculture [16], [21], [22], [23].

Amarzirh et al. [24] used Sentinel-1 data to characterize spatial variability of SM for fusion of high spatial Landsat-9 LST and medium resolution MODIS LST data. Their results when applied to two different agricultural sites underlined the importance of taking into account the Sentinel-1 as input to the disaggregation. Similarly, Chintala et al. [18] combined Sentinel-1 with Sentinel-2 and Sentinel-3 LST to directly estimate actual ET fluxes for croplands under high cloud cover. Besides its backscattering, Ouaadi et al. [25] also used Sentinel-1's weekly coherence information, which is linked to the vertical morphology of crops, for ET estimation. Similarly, for sugarcane ET, Alavi et al. [26] proposed using Sentinel-1in the case of unavailability of Landsat-8 LST data in the cloudy conditions. Whether using a pure data-driven or a physical-based methodology, all these recent studies aim to leverage the relationship between SAR and ET, which lies in the ability of SAR to provide information about the structure and water content of soil-vegetation volume.

Despite the advantage of using Sentinel-1 coupled with optical sensors for ET estimation, the requirement of sequential high spatial resolution optical data cannot be ignored. Specifically, for rainfed croplands, the performance of any model relying on optical sensor cannot satisfy the expectations, and the spatial resolution of any high temporal resolution optical data would not be enough for field-scale agriculture [14], [27], [28]. To address these difficulties, this work proposes a weakly supervised learning (WSL) framework to estimate ET as target and high spatial resolution Sentinel-1 and climate variables as inputs for having cloud-free regular ET product over croplands.

In this study, we first constructed a benchmark dataset to facilitate systematic investigations into SAR-driven ET estimation. This dataset includes both static and dynamic (time-dependent) data related to topography and meteorology, obtained from various sources with differing spatial and temporal resolutions, covering a vast geographical region. The dataset primarily comprises ET and SAR data, and multiple auxiliary weather, SM, topography, and soil texture variables. This benchmark dataset provides a foundation for our subsequent experimental analysis in later sections. Then, we proposed a SAR2ET[1] model that utilizes a UNet-like architecture that performs patch-level mapping, taking a combination of SAR and auxiliary data patches as inputs and predicting a corresponding ET patch, as summarized in Fig. 1. To train the model, we incorporated multiple data augmentation techniques and a customized mean squared loss error that emphasizes only high vegetation pixels. Then, we analyzed the effect of using auxiliary data from different sources, with or without SAR, on the overall accuracy of ET estimation. We assessed the accuracy using multiple evaluation metrics, and we further examined the performance of our proposed approach on samples grouped geographically and temporally and conducted additional analyses that are explained in detail in Section IV.

II. Data Sources and Preprocessing

To enable systematic studies on SAR-driven ET estimation, we carefully construct a benchmark dataset. To explore the potential value of auxiliary data sources, we also incorporate

[1]The SAR2ET dataset will be made available at:https://github.com/Agcurate/SAR2ET

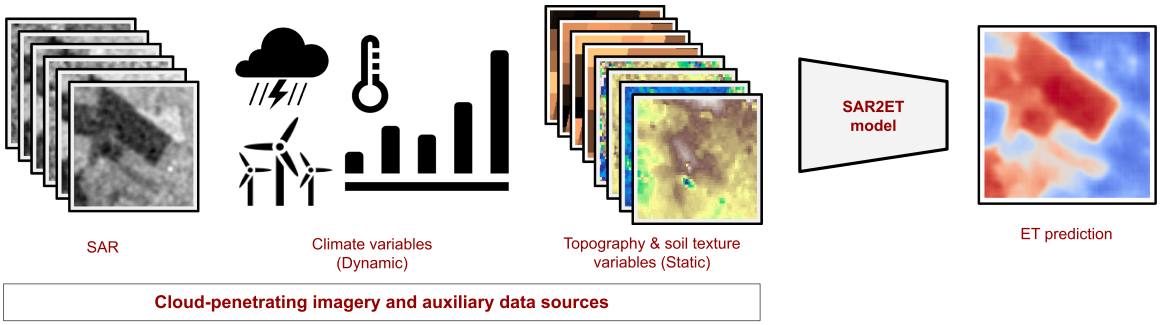

Fig. 1. Graphical representation of SAR-based ET estimation: SAR2ET.

TABLE I
DATA EMPLOYED IN THIS STUDY FOR WEAKLY SUPERVISED SAR-DRIVEN ET ESTIMATION

Data type	Number of channels (C)	Temporal resolution	Spatial resolution	Source
ET	1	16 days	30 m	Landsat
VV & VH backscatters, CR (both ASC and DESC)	6	6 days	10 m	Sentinel-1
Temperature, Dewpoint Temperature, Total Precipitation, U Component of Wind, V Component of Wind, Surface Net Solar Radiation Sum, Surface Pressure	7	daily	11132 m	ERA5-Land
SM	1	daily	9 km	SMAP
Elevation [Height (m), Slope (°), Aspect (°), Hillshade (°)]	4	N/A	30 m	ALOS DSM
Clay, Sand, Silt, bulk density (bdod)	4	N/A	250 m	SoilGrids
NDVI (only for analysis)	1	5 days	10 m	Sentinel-2
Cropland mask	1	annually	10 m	Dynamic World

various temporally static and dynamic data sources. In our selection of dynamic auxiliary data sources, we avoid options that would imply a dependency on optical sensors so as to maintain operability during weather events. Finally, the cropland masks are utilized to focus on agricultural regions, and Sentinel-2-based NDVI data only for experimental analysis.

A summary of all data sources, including temporal and spatial resolutions, are provided in Table I. The following paragraphs define the time period and geographical regions used in data collection, and the details of individual data types.

A. Time Period and Geographical Regions

To conduct this study, Landsat-based ET imagery is downloaded from the open source Earth engine evapotranspiration flux[2] (EEFlux) [29] for our selected area of interest between March 1 and September 30, 2021, and resampled to a 10-m resolution. Corresponding SAR (Sentinel-1) imagery and auxiliary data, such as meteorological variables, SM, topography, soil texture, cropland mask, and NDVI (Sentinel-2) are all downloaded from the Google Earth Engine (GEE) platform. In order to identify the study area, we first selected regions with significant agricultural activity. We then gathered all the available ET data in the vicinity of these regions, including any additional areas that may be covered, resulting in a large study area, as shown in Fig. 2. The time interval was selected to gather data that covers the majority of the agricultural season in the study area. This allows us to evaluate our approach on various crop development phases.

B. ET Data

EEFlux is used as the source for the actual ET data. EEFlux is a version of mapping evapotranspiration at high resolution with internalized calibration [30], a satellite-based image processing model. It operates on the GEE system and provides global ET products in 30-m spatial resolution based on Landsat data. The data was collected using Landsat with a revisit frequency of approximately 16 days. We retrieved a total of 121 ET data rasters based on 11 predefined distinct regions within a predefined date range. Although each raster covers a vast region of approximately 54 000 km^2, only a relatively small portion of all ET data are included in the final dataset after going through the process of patch extraction. EEFlux-derived ET serves as a viable alternative to field measurements in various studies. The efficacy of this estimation method has been corroborated

[2]EEFlux: https://eeflux-level1.appspot.com

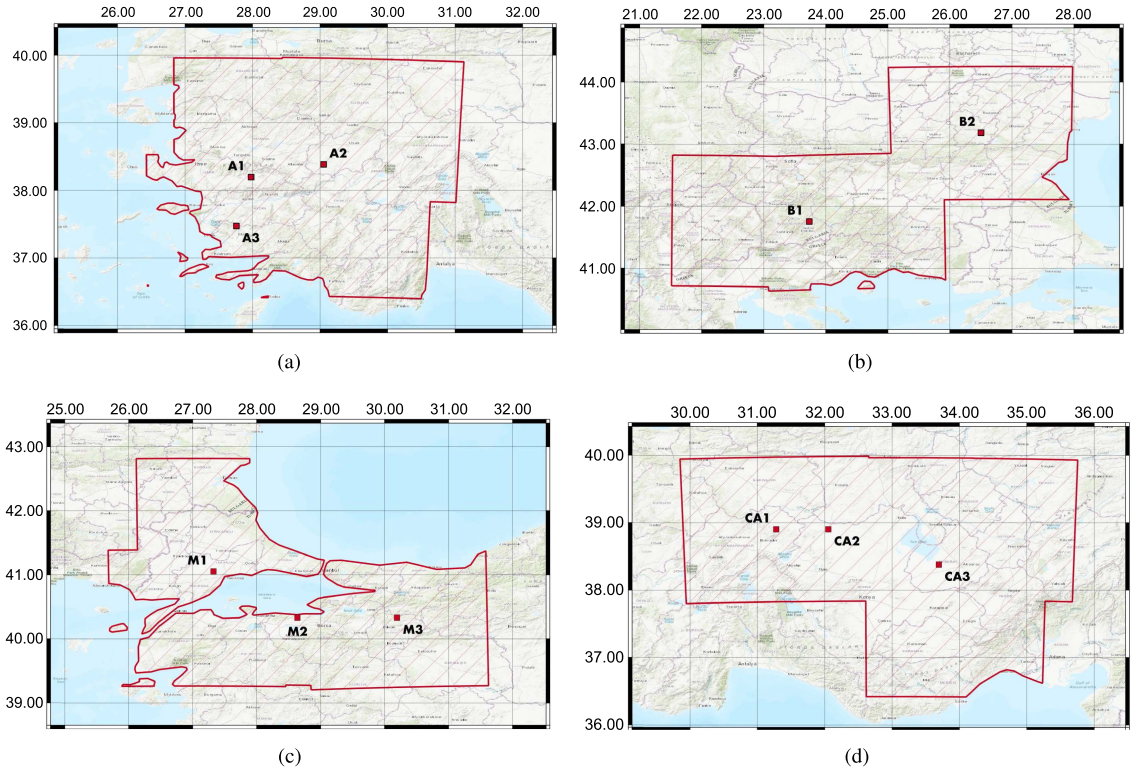

Fig. 2. Study area. The dataset consists of patches from specific regions with high agricultural activity, centered on 11 predefined areas. A1, A2, and A3 are chosen for the Aegean region (as shown in Section II-A, B1 and B2 for the The Balkans (Section II-A), M1, M2, and M3 for Marmara Region (Section II-A), and CA1, CA2, and CA3 for the Central Anatolia (Section II-A). (a) Aegean region. (b) Balkans. (c) Marmara region. (d) Central anatolia.

by Nisa et al. [31] and Kadam et al. [32], who demonstrated its satisfactory performance through comparison with ground observations in their respective works.

C. SAR Data

This study uses Sentinel-1 (S1) ground range detected products in 10-m resolution as the SAR data source. The interferometric wide swath mode of the radar satellite is used to obtain both log-scaled VV and VH backscatters. By default, the data undergoes preprocessing steps, such as thermal noise removal, radiometric calibration, and terrain correction. Backscatter values are separately collected according to the direction of the orbit, either descending (DSC) or ascending (ASC). They are subsequently subjected to incidence angle filtering (any backscatter values with an incidence angle below 30^o or above 42^o are excluded) due to the high impact of angle changes on agricultural products. In addition, to avoid redundant data points and reduce speckle effects in S1 scenes, edge filtering (by removing backscatter values lower than -30°) and spatial smoothing (by using a 5×5 average smoothing kernel) are applied. In order to acquire the final S1 data for every target ET data, all the available S1 data from a week prior to and a week after the date of ET data are gathered. The collected S1 data is then mosaicked based on their closeness to the date of ET data to obtain a single S1 raster that corresponds to the target ET and to fill in any gaps that may have been created due to incidence angle filtering and edge filtering. Finally, a cross ratio (CR) (VH-VV, in log scale) is included as an extra band to facilitate further ET estimation tasks.

D. Dynamic Auxiliary Data: Meteorological and SM

For the auxiliary meteorological (dynamic) data, we utilize global ERA5-Land Daily Aggregated [33] and soil moisture active passive (SMAP) [34] datasets, which offer daily data. The ERA5 reanalysis dataset provides a range of meteorological data directly influencing the rate of ET, such as temperature (measured 2 m above the surface), dewpoint temperature (measured 2 m above the surface), total precipitation, U and V components of wind (measured 10 m above the surface), surface net solar radiation, and surface pressure. These weather data variables have a spatial resolution of about 11 km. The SMAP dataset, on the other hand, provides SM data on a daily basis with a spatial resolution of 9 km.

E. Static Auxiliary Data: Topography and Soil Texture

For the auxiliary static data, we utilize ALOS World 3-D [35] and SoilGrids [36] datasets. The ALOS World 3-D global digital surface model provides static parameters of the height (m), slope (°), aspect (°), and hillshade (°) with a 30-m spatial resolution, which have impact on ET through their influence on temperature distribution and solar radiation exposure. The SoilGrids dataset offers comprehensive data on soil attributes worldwide, with a spatial resolution of 250 m. This study focuses on four vital parameters of SoilGrids for the ET: the bulk density (bdod), clay,

sand, and silt ratios of the fine Earth component at a depth of 0–5 cm beneath the surface, which are important for understanding water storage capability in the soil profile.

F. Third-Party Data: NDVI and Cropland Mask

In order to further analyze the benefits of our research with a focus on the vegetation activity, we gathered the two 10-m bands, red (band 4) and NIR (band 8), from the Sentinel-2 (S2) satellite (L2A product) to calculate the NDVI. Similar to the coregistration of SAR data, to obtain the final NDVI data for each target ET data, the available red and NIR bands of S2 from one week before and one week after the ET data's date are collected. These data are then combined based on their proximity to the ET data date through a process called mosaicking. This method is used to fill in any gaps that may have resulted from clouds since the products with a cloud probability greater than 10% are filtered out, and the remaining products are further filtered using a cloud mask based on the QA band.

As we focus on ET estimation on agricultural regions, cropland masks are essential. Rather than crafting the masks from scratch, we make use of the Dynamic World (DW) [37] dataset presented in 10-m resolution. We collected all available land cover masks for the targeted regions in the year of the study and then determined the crop pixels that are consistently indicated as crops throughout the year in each scene.

III. Method

In this section, we formally define our problem setting for end-to-end ET prediction, explain the details of the benchmark that we construct from the data sources defined in Section II, and give the details of the proposed multisource approach.

A. Problem Definition

We approach the multisource SAR2ET problem as a patch-level mapping task. We presume that the training set $\mathcal{D}_{\text{train}}$ consists of input and target ET image patch pairs (x, y). All data sources are resampled to 10-m resolution. Therefore, each input x is obtained by concatenating the input data tensors, i.e., $x \in \mathbb{R}^{H\times W\times|C|}$, where H and W refer to the height and width, and C indicates the input channels. The list C is obtained by selecting a combination of $\{C_{\text{SAR}}, C_{\text{ERA5}}, C_{\text{DEM}}, C_{\text{SMAP}}, C_{\text{SoilGrids}}\}$, corresponding to the data sources presented in Table I. Each target output is similarly of the same spatial dimensionality, i.e., $y \in \mathbb{R}^{H\times W}$.

The SAR2ET model is represented by the function $f(x;\theta) : \mathbb{R}^{H\times W\times|C|} \to \mathbb{R}^{H\times W}$, parameterized by the trainable parameters θ. Therefore, the modeling problem is essentially seen as learning an image-to-image translation model from the inputs to the target patch of corresponding ET values. We note, however, that the task involves a fundamentally WSL problem: while the primary input SAR imagery provides a native 10-m resolution, the model is constructed at this most detailed resolution level, the ET imagery originally has 30-m resolution. The fact that the targets come from a 3× lower-resolution sensor imply that a single supervisory signal is provided effectively every 9 output pixels.

We also highlight that the modeling problem is challenging at the input side as well. First, we observe that most auxiliary data sources have original resolutions different from, and sometimes much lower than the model's operating resolution, e.g., SM has 9 km resolution as opposed to the 10-m modeling resolution. Second, some of the input channels contain data with highly different modalities. Third, the sources are coarsely geo-registered, especially due to drastic resolution differences, pixels are only coarsely across the data channels. All these factors greatly increase the difficulty of the modeling task.

The model selection and final evaluation steps are carried out on held out validation and test datasets, referred to as $\mathcal{D}_{\text{val}}$ and $\mathcal{D}_{\text{test}}$, respectively. These datasets are presumed to contain input and target ET image patch pairs with the same spatial and spectral specifications.

B. Approach Details

In the following parts, we present the approach details, including the benchmark construction, network architecture, data augmentations, the loss function, the evaluation metrics and the model selection methodology.

Benchmark construction: Creating a benchmark dataset for studying SAR2ET predictions is a central part of our work. As mentioned in Section II, we collected matching rasters from various data sources to create a complete dataset that includes and is based on 121 target ET rasters collected in 2021. Each sample in the dataset consists of an ET patch and corresponding patches from different data sources that cover the same exact location and belong to the same or a closer date to ET's date. Once all the data sources, covering the same locations and dates, are gathered, we carefully extract patches that are 128×128 pixels in size. We take care to ensure that none of the patches from any data source contain any invalid pixel values that may have resulted from no data and nonterrain pixels, or cloud pixels if the data source is optic-satellite-based. Furthermore, patches that do not contain any crop pixels, as determined by cropland masks (refer to Section II-F), are also excluded. After completing the pairings of the patches, a major challenge was the uneven distribution of patches collected over a wide geographic area and time range. To this end, we ensured that the patches were split into sets in an equally distributed manner. Therefore, we split the data into train, validation, and test sets using a 70%, 15%, and 15% ratio, respectively, by adopting a splitting strategy that applies the same split ratio to each individual ET raster and its matching rasters of different data sources to ensure the patches were proportionally distributed from different regions and dates into each split. As a result, we obtained 447 951, 95 939, and 96 111 patches in the train, validation, and test sets, respectively.

Network architecture: The network architecture is a modified version of the fully convolutional UNet [38], featuring four layers that comprise both contracting and expansive paths, and double convolution operations are utilized throughout the forward pass. The contracting path follows a standard convolutional network approach, utilizing repeated 3×3 kernel convolutions.

Each convolution is succeeded by a rectified linear unit and a max pooling operation. The expansive pathway combines the extracted feature and spatial information through a sequence of up-convolutions. To maintain the original size of the features, padding is applied, as the size diminishes after each convolution step in the original UNet implementation. In addition, batch normalization layers are added to double convolution blocks for more stable and efficient training.

In order to prepare the SAR2ET model input, we combine SAR and auxiliary data patches and input them into the model. Channel-wise concatenation is applied depending on the SAR2ET model being trained. For example, if we are training a basic SAR2ET model that uses only SAR patches to predict ET, the input tensor will be shaped as $(6 \times 128 \times 128)$ since SAR data has six variables (see Table I). However, if we want to train a SAR2ET model that predicts ET using both static (like DEM) and dynamic (like ERA5) auxiliary data in addition to SAR, we must concatenate patches from all data sources. In this scenario, we concatenate patches of size $(5 \times 128 \times 128)$ for DEM, $(7 \times 128 \times 128)$ for ERA5, and $(6 \times 128 \times 128)$ for SAR, to obtain a final patch of size $(18 \times 128 \times 128)$, which is fed into the model as input.

Data augmentations: In order to prevent overfitting and artificially increase the number of training examples for the model, we use a data augmentation pipeline that involves various geometric augmentations, such as 90° rotations (with a uniform probability of rotating between 0 to 3 times), and horizontal and vertical flipping (with a 50% probability), applied to each training example. During training, we applied the same augmentations equivalently to each input patch from different data sources (SAR, ERA5, DEM, etc.) and target ET patch.

Masked loss: We approach the problem of estimating ET as a regression task, and simply use the mean squared error (MSE) loss for model training. To ensure that we focus on lands with high agricultural activity, we have utilized the binary cropland masks corresponding to each patch, using the corresponding mask source explained in Section II-F. Consequently, we incorporate the cropland masks into the final loss function, ℓ, via pixel-wise masking. The masked loss function is defined as follows:

$$\ell(\hat{y}, y) = MSE(\hat{y}, y) \odot M \tag{1}$$

where y and $\hat{y}$ represent the ground truth and predicted ET, respectively, both having the same size $(1 \times 128 \times 128)$. The binary cropland mask, denoted by M, has the same size $(1 \times 128 \times 128)$ but of boolean type and indicates the pixels with high vegetation activity. The masking provides accumulating gradient signals only over the cropland pixels.

Evaluation and model selection: To evaluate the quality of ET predictions, MSE, mean absolute error (MAE), and coefficient of determination (R^2) metrics are used. During the training process, a separate validation set is utilized to fine-tune the models and prevent over-fitting to the training data. This is achieved by calculating batch-averaged MSE, MAE, and R^2 on the validation set. Once the hyperparameters are fine-tuned according to their batch-averaged validation R^2, their final performance is evaluated on the separate test set. All reported quantitative and qualitative results in the experiments section are based on the performance of the models on the test set.

In order to speed up the training process, instead of using all training examples, 85% of randomly selected examples are iterated over in each *epoch*, which affects the learning rate schedule and makes the entire training process more efficient without any significant performance drop.

Summary: A summary of the model is provided as an illustration in Fig. 3. Once the model is trained, we estimate pixel-wise ET values based on SAR data as well as static data, such as topographical and soil texture, and dynamic data, such as meteorological and SM, which do not rely on optical satellites. The influence of each input data type, as well as the overall behaviour of the model, is evaluated and analyzed in the next section.

IV. Experiments

In this section, we thoroughly evaluate various SAR2ET models that estimate ET using different combinations of static and dynamic data sources. We first provide a summary of important implementation details. Then, in Section IV-B, we provide a detailed quantitative comparison of the models, highlighting their strengths and weaknesses, and showing the impact of different data sources on ET estimation accuracy. More specifically, we measure model performance using the evaluation metrics explained in Section III-B, and analyze seasonal and regional differences by evaluating these models across different study areas to identify any regional disparities and analyze R^2 scores on a monthly basis for the year 2021. In Section IV-C, we provide qualitative results to further examine the models' performance. This includes presenting visual examples of patches from the test set and their corresponding predictions by the models. Finally, in Section IV-D we evaluate the importance of SAR as a data source for ET estimation in comparison auxiliary data sources, and analyze the impact of soil texture and vegetation on estimation quality.

A. Implementation Details

Instead of training multiple region-specific SAR2ET models, we trained a single SAR2ET model using the train data constructed from crops in all four regions, collectively. With this approach, we aim to teach the model a generalizable representation that can be applied to any unseen data from any of the included regions in the inference stage.

To optimize the training of deep learning models, tuning the learning rate is crucial, as it is one of the most sensitive hyperparameters. To achieve this, we utilized Adam optimizer with default parameters except for weight decay and adopted the 1-cycle learning rate scheduling policy [39] since it offers a principled and efficient learning rate tuning scheme, and high learning rate values in the middle of the model training show a regularization effect. With this approach, the learning rate gradually increases from the initial rate to a maximum rate and then decreases further to a minimum rate that is significantly lower than the initial rate. We set the maximum learning rate to 1e-4. Initially, the (initial) learning rate is set to 1/25 of

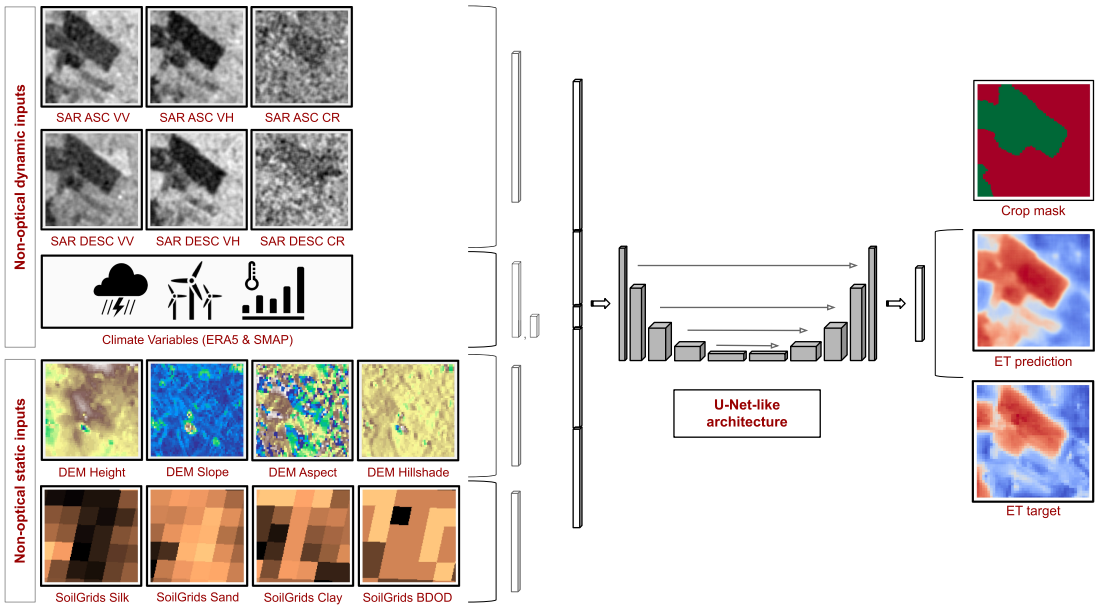

Fig. 3. Illustration of the proposed SAR-based ET estimation approach. A U-Net-like SAR2ET model takes a combination (channel-wise concatenation) of data patches from multiple nonoptical dynamic and static sources and predicts ET. With a low-resolution ground truth ET and a binary crop mask, the model learns to optimize its prediction strategy with a focus on the pixels with high vegetation activity.

TABLE II
COMPARISON OF MULTIPLE SAR2ET MODELS TRAINED WITH DATA FROM VARIOUS STATIC AND DYNAMIC DATA SOURCES ALONG WITH SAR FOR ESTIMATING ET, IN TERMS OF MSE (MM/DAY), MAE (MM/DAY), AND R^2 ON THE TEST SET

Model inputs	MSE	MAE	R^2
SAR	0.94	0.72	0.63
SAR, SoilGrids	0.86	0.68	0.66
SAR, DEM	0.79	0.65	0.69
SAR, DEM, SoilGrids	0.74	0.63	0.71
SAR, SMAP	0.85	0.68	0.67
SAR, ERA5	0.51	0.52	0.80
SAR, ERA5, SMAP	0.51	0.52	0.80
SAR, ERA5, DEM	**0.46**	**0.49**	**0.82**

this maximum value. It gradually increases and reaches the maximum value at around 10 epochs. Afterward, the learning rate gradually decreases to 1/1000 of the maximum value until the end of the learning process. Apart from that, for all the experiments, the weight decay is set to 1e-6, the batch size is set to 128, and the models are trained for 50 epochs. All experiments are performed on a system with a single GPU of NVIDIA GeForce RTX 3090 Ti on the CUDA 12.0 platform using PyTorch 2.0.1.

B. Main Results

In this section, we present the quantitative results and provide a comprehensive understanding of the measured variables in our study.

Table II shows the performance metrics of various SAR2ET models that are trained using different combinations of static and dynamic data sources, on the test set. The baseline SAR model achieves an MSE of 0.94 mm/day, an MAE of 0.72 mm/day, and a R^2 of 0.63. However, as auxiliary static data sources are added to the models, their performance further improves. For example, when combined with SoilGrids data, the SAR model's R^2 increases to 0.66. The inclusion of DEM data further improves the model's performance, resulting in an even higher R^2 of 0.69. Even better results are achieved when combining SAR, DEM, and SoilGrids data altogether, with an MSE of 0.74 mm/day, an MAE of 0.63 mm/day, and a R^2 of 0.71 (an 8-point improvement over the baseline), which shows the significance of integrating multiple static data sources. However, dynamic data sources, especially ERA5-based meteorological data, also have a significant impact on enhancing model accuracy. To illustrate, the SAR model, when combined with ERA5 data, achieves an outstanding R^2 of 0.80, a significant 17-point improvement over the baseline, indicating the importance of meteorological data in SAR-based ET estimation. The best model in this comparison incorporates SAR, ERA5, and DEM data, achieving an MSE of 0.46 mm/day, an MAE of 0.49 mm/day, and a R^2 of 0.82 (a 19-point improvement over the baseline). This underscores the synergistic benefits of combining both static and dynamic data sources. Expanding this combination further with SoilGrids data yields the nearly the same results across all evaluation metrics, therefore, in the interest of brevity, we have chosen to omit it. Similarly, although SM is directly related to ET, the added value of SMAP-based SM data to SAR signal could be overlooked in the presence of meteorological reanalysis data. This is attributed to the spatial resolution of the SMAP dataset, which is insufficient to capture field-level SM variations. Instead, the meteorological parameters do not vary as much as SM across the spatial domain. Overall, Table II emphasizes the importance of integrating SAR, DEM, and ERA5 data to enhance the accuracy of SAR2ET models for estimating ET.

Table III provides a detailed comparison of multiple SAR2ET models, each trained with a combination of static and dynamic

TABLE III
COMPARISON OF MULTIPLE SAR2ET MODELS TRAINED WITH DATA FROM VARIOUS STATIC AND DYNAMIC DATA SOURCES ALONG WITH SAR FOR ESTIMATING ET BY MONTHS OF THE STUDY YEAR 2021, IN TERMS OF R^2 (HIGHER IS BETTER) ON THE TEST SET

Model inputs	Mar	Apr	May	Jun	Jul	Aug	Sep	Overall
SAR	0.37	0.31	0.49	0.48	0.63	0.68	0.49	0.63
SAR, SoilGrids	0.43	0.35	0.54	0.51	0.66	0.70	0.55	0.66
SAR, DEM	0.51	0.49	0.58	0.55	0.69	0.72	0.58	0.69
SAR, DEM, SoilGrids	0.56	0.54	0.61	0.57	0.70	0.74	0.61	0.71
SAR, SMAP	0.52	0.43	0.59	0.52	0.65	0.69	0.55	0.67
SAR, ERA5	0.72	0.67	0.76	0.69	0.77	0.82	0.75	0.80
SAR, ERA5, SMAP	0.72	0.67	0.77	0.69	0.77	0.83	0.75	0.80
SAR, ERA5, DEM	**0.75**	**0.69**	**0.78**	**0.71**	**0.79**	**0.85**	**0.78**	**0.82**

TABLE IV
COMPARISON OF MULTIPLE SAR2ET MODELS TRAINED WITH DATA FROM VARIOUS STATIC AND DYNAMIC DATA SOURCES ALONG WITH SAR FOR ESTIMATING ET BY REGIONS CENTERED ON STUDY POINTS (SEE FIG. 2), IN TERMS OF R^2 (HIGHER IS BETTER) ON THE TEST SET

	Aegean (A)				Balkans (B)			Marmara (M)				Central Anatolia (CA)				
Model inputs	A1	A2	A3	All	B1	B2	All	M1	M2	M3	All	CA1	CA2	CA3	All	Overall
SAR	0.56	0.55	0.59	0.56	0.57	0.77	0.71	0.61	0.52	0.53	0.59	0.58	0.63	0.63	0.62	0.63
SAR, SoilGrids	0.59	0.59	0.61	0.59	0.63	0.80	0.75	0.63	0.53	0.56	0.61	0.63	0.64	0.65	0.65	0.66
SAR, DEM	0.62	0.62	0.63	0.63	0.65	0.80	0.75	0.67	0.58	0.61	0.65	0.66	0.67	0.68	0.68	0.69
SAR, DEM, SoilGrids	0.63	0.64	0.64	0.64	0.68	0.83	0.78	0.68	0.60	0.63	0.66	0.68	0.69	0.71	0.70	0.71
SAR, SMAP	0.58	0.57	0.60	0.58	0.59	0.79	0.73	0.65	0.53	0.56	0.62	0.70	0.66	0.66	0.68	0.67
SAR, ERA5	0.76	0.74	0.78	0.75	0.80	0.87	0.85	0.75	0.70	0.71	0.74	0.82	0.79	0.80	0.81	0.80
SAR, ERA5, SMAP	0.76	0.74	0.78	0.75	0.80	0.87	0.85	0.76	0.70	0.71	0.74	0.83	0.79	0.80	0.81	0.80
SAR, ERA5, DEM	**0.79**	**0.77**	**0.81**	**0.78**	**0.81**	**0.87**	**0.86**	**0.78**	**0.73**	**0.74**	**0.76**	**0.84**	**0.81**	**0.82**	**0.83**	**0.82**

data sources, for the purpose of estimating ET over the course of the study year 2021. The primary focus of this analysis is the variation in R^2 scores across different time intervals (months of the year), shedding light on the models' performance throughout the year. Examining the R^2 scores by month, we observe distinct patterns. The SAR-only model starts with relatively lower R^2 values in March (0.37) and April (0.31) but gradually improves its performance as the year progresses, achieving its highest R^2 scores in July (0.63) and August (0.68). This suggests that the SAR model becomes more accurate in capturing ET variations as vegetation and environmental conditions change during the warmer months. As additional data sources are incorporated into the models, such as SoilGrids, DEM, SMAP, and ERA5 data, the R^2 scores generally show improvement across all months. The model trained with SAR, ERA5, and DEM data consistently outperforms others, with the highest R^2 scores across all months, notably achieving a remarkable R^2 of 0.85 in August. This demonstrates that combining SAR with dynamic (meteorological) and static (topographical) data significantly enhances the model's ability to estimate ET throughout the year. Overall, Table III illustrates the importance of data integration in SAR2ET modeling and highlights how the accuracy of ET estimation varies by month, with SAR-based models becoming more effective in capturing ET dynamics during the warmer months of the year when the crops have sufficient developed leaves for exchanges of mass and energy. In other words, with vegetation growth the model is getting more sensitive to Sentinel-1 parameters explaining depolarization based on volume scattering. The clear relationship between SAR and ET based on the crop's volume gradually diminishes towards the late season of crops (around September). This trend is attributed to the radar signals' high penetration into dry crop volume.

Table IV demonstrates notable improvements in the overall performance of SAR2ET models across various geographic regions. Initially, the SAR model exhibits moderate accuracy in estimating ET, with R^2 scores ranging from 0.56 to 0.71 across different regions. However, as additional static data sources, such as SoilGrids and DEM are integrated into the models, a consistent enhancement in R^2 scores is evident. The model trained with SAR, DEM, and SoilGrids data, for instance, achieves a R^2 value of 0.64 in the Aegean region (an 8-point improvement over the baseline), illustrating the positive influence of static data, topographical and soil texture information, on ET estimation. This trend continues in other regions, including the Balkans (B), Marmara (M), and Central Anatolia (CA), where the integration of data sources like DEM and SoilGrids consistently improve R^2 scores by 7–8 points. The most remarkable performance improvement is observed in the model trained with SAR, ERA5, and DEM data, where R^2 scores reach their maximum when both the most significant static (DEM) and dynamic (ERA5) data sources are present. In areas, such as the Balkans (B) and Central Anatolia (CA), this model has shown remarkable R^2 values of up to 0.86 and 0.83, respectively. These values represent a significant improvement of 15 and 21 points over the baseline and indicate a robust correlation between model predictions and observed ET data. These findings emphasize the importance of

TABLE V
COMPARISON OF SAR VERSUS AUXILIARY DATA SOURCES

Model inputs	MSE	MAE	R^2
SAR-only	0.94	0.72	0.63
Auxiliary-only	1.06	0.77	0.58
SAR + auxiliary sources	**0.46**	**0.49**	**0.82**

Auxiliary refers to the combination of ERA5, SMAP, DEM and SoilGrids.

combining topographical data (DEM) and meteorological data (ERA5) in SAR-based ET estimation, especially in areas where these variables significantly influence ET dynamics.

C. Qualitative Results

In this section, we share and examine the qualitative results of our study, providing a detailed analysis and discussions of the insightful and contextually based information gathered from the experiences.

Fig. 4 illustrates the comparison between ground truth ET and predicted ET's made by different SAR2ET models. Each row shows a different example belonging to the test set. The first column displays true-colored satellite imagery derived from S2, while the next column shows their corresponding ground truths ET's. The remaining columns show the predictions made by multiple SAR2ET model that are trained with different sets of data that are indicated at the top of each column. Overall, the figure suggests that the SAR2ET model, which harnesses the synergy of meteorological data, topographical features, and SAR, emerges as the most robust and accurate predictor of ET. This combined approach allows the model to effectively capture micro-seasonal weather patterns and local topographical structures. Consequently, it consistently delivers ET predictions that closely align with the target values. In contrast, the SAR2ET model exclusively relying on SAR data exhibits limitations in accurately representing the subtle nuances of micro-seasonal variations in weather conditions and local terrain characteristics. It tends to yield ET predictions that deviate from the expected target, sometimes underestimating or overestimating the ground truth ET. Meanwhile, the SAR2ET model that only relies on a combination of auxiliary static and dynamic data sources performs poorly in its ET predictions. The primary challenge lies in the relatively lower resolution of the auxiliary data, which, although capable of approximating average target ET values more faithfully in certain instances, fails to capture the finer details of ET variations across the study area (see Appendix for additional results).

Furthermore, as demonstrated in the last row of Fig. 4, the SAR2ET model has successfully proven its capability to accurately estimate ET by capturing ground terrain signals through the use of cloud-penetrating SAR images, even in the presence of clouds.

Fig. 5 shows a comparison of ground truth and predicted mean ET by date on four study areas (see Fig. 2). Each study area is explained with two subfigures. In the subfigure on top, the ground truth mean ET (green circle) is shown as well as the mean ET predicted by two SAR2ET models. The first model was trained with SAR only (red triangle), while the second model was trained with SAR, ERA5, and DEM data combined (blue square). In the bottom subfigure, the residual error which refers to the difference in errors obtained using both models, are presented. Based on the figure, it can be concluded that the SAR2ET model trained with SAR, ERA5, and DEM outperforms the model trained with only SAR in almost all cases and regions by better capturing the trend in ET change throughout the year and making accurate predictions matching the ground truth mean ET's with smaller deviations. It is also clear from the residual error that the model trained with a combination of SAR, ERA5, and DEM consistently achieves a significantly smaller loss compared to the model trained with SAR only. Another observation is that as the sample size increases (indicated by larger circles), the mean of the predictions tends to align with the mean of the target values. Conversely, with a smaller number of examples, the variance tends to be higher. Overall, Fig. 5 indicates that the SAR2ET model, which combines meteorological and topographical features with SAR, produces more accurate predictions, consistent with the quantitative and qualitative results presented.

D. Analysis

In this section, we provide two important additional analysis on 1) the importance of SAR incomparison to other data sources, and 2) the impact of soil texture and vegetation on SAR2ET results.

SAR versus auxiliary data sources: The results so far show that the best results are obtained when SAR is used in combination with auxiliary data sources. An important question is the role of SAR in overall achievements, considering that the auxiliary sources already provide a rich set of meteorological and topographical indicators. To address this question, Table V presents a comparison of the models using only SAR, only auxiliary sources and their combination. Here, auxiliary sources refer to a full combination of ERA5, SMAP, DEM, and SoilGrids. We can observe that auxiliary-only model yields a larger MSE (1.06 mm/day) and MAE (0.77 mm/day) and a lower R^2 (0.58) when compared to the SAR-only model. This highlights the fact the inclusion of SAR itself is a crucial component for the construction of the SAR2ET model. The significant achievements of the final result, combining SAR and auxiliary sources, highlight the complementarity of these sources of very different nature. Consistent observations can similarly be made in the montly comparison presented in Table VI for these three main model variations. These results again highlight the value of combining SAR with dynamic and static data sources in SAR2ET modeling.

Impact of soil texture and vegetation: Fig. 6 presents an analysis of the best-performing SAR2ET model, which leverages a combination of static and dynamic data sources, DEM and ERA5, in addition to SAR. We first group all test examples according to the soil texture and NDVI features, then, calculate the R^2 scores for each group and plot them as line graphs. This allows us to observe the trends in model performance in relation to the the soil texture indicators and NDVI values. Each plot corresponds to a soil type, and each plot curve corresponds to an NDVI range, as indicated in the colorbars.

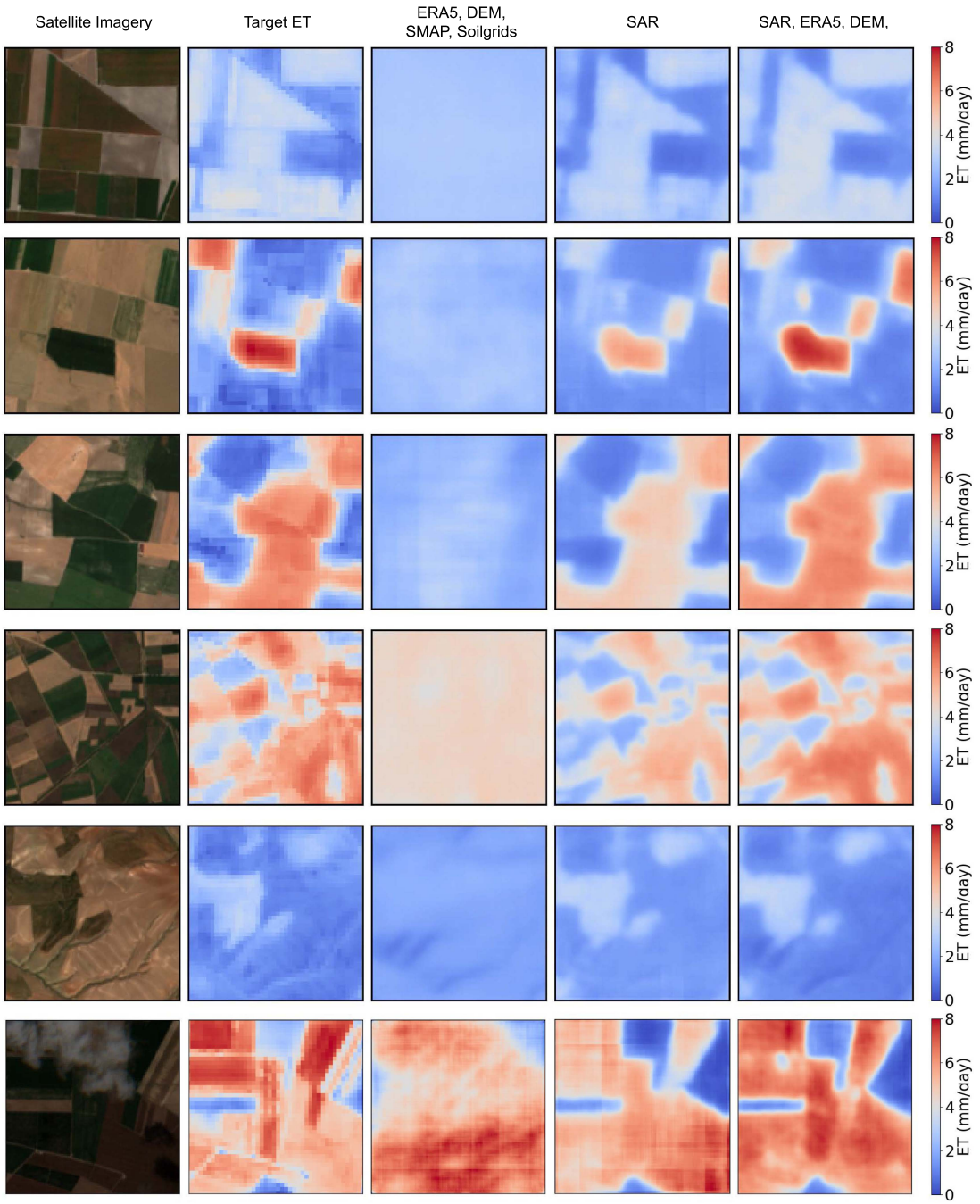

Fig. 4. Multiple examples of test patches alongside their corresponding ET predictions from different SAR2ET models. Each row represents a different example, with the columns presenting an RGB image (captured using S2) of the patch area, the ground truth ET of the same area, and the ET predictions from SAR2ET models trained using auxiliary data sources only, SAR-only, and the highest performing model trained with SAR, ERA5, and DEM. (Best viewed in color. Red indicates higher ET while blue indicates lower ET in terms of mm/day.).

TABLE VI
COMPARISON OF SAR VERSUS AUXILIARY DATA SOURCES WITH PER-MONTH RESULT, IN TERMS OF R^2

Model inputs	Mar	Apr	May	Jun	Jul	Aug	Sep	Overall
SAR-only	0.37	0.31	0.49	0.48	0.63	0.68	0.49	0.63
Auxiliary-only	0.60	0.35	0.53	0.47	0.48	0.61	0.52	0.58
SAR + auxiliary sources	**0.75**	**0.70**	**0.79**	**0.72**	**0.79**	**0.85**	**0.79**	**0.82**

Auxiliary refers to the combination of ERA5, SMAP, DEM and SoilGrids.

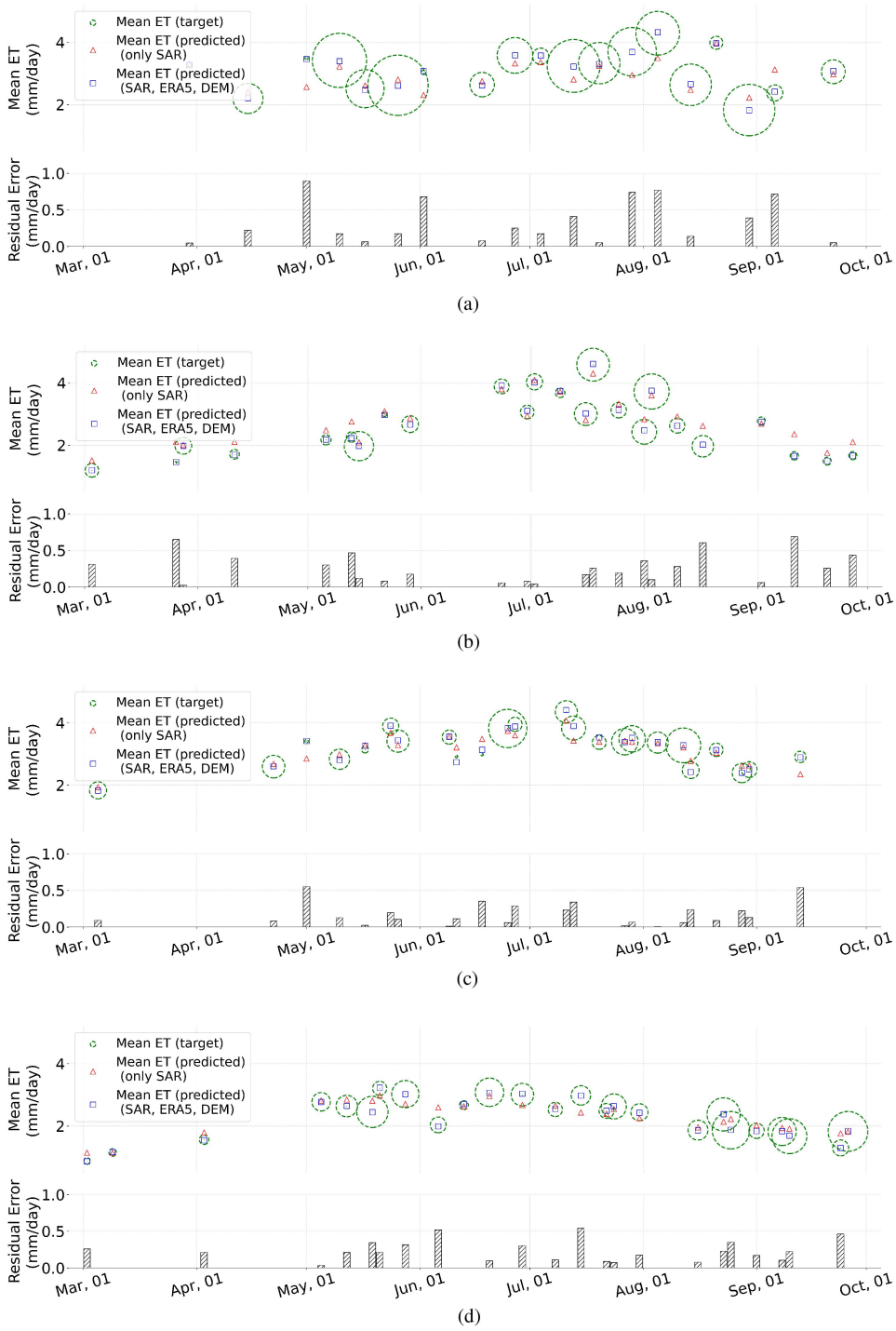

Fig. 5. Plots comparing ground truth and predicted mean ET (mm/day) by date on four study areas. In each figure, the top subfigure shows ground truth and predicted mean ET's by dates (predicted by two SAR2ET models; the one trained with SAR only and the one trained with SAR, ERA5, and DEM combined and the bottom subfigure shows the residual error which refers to the difference in errors obtained using the model trained with SAR, DEM, ERA combined versus the model trained with SAR only. The size of the green circles (indicator of ground truth mean ETs) indicates the number of samples collected; bigger circles represent larger sample sizes. (Predictions that are closer to the center of the circle is better.) (a) Aegean region (A). (b) Balkans (B). (c) Marmara region (M). (d) Central anatolia (CA).

One prominent trend that we spot in Fig. 6 is the model's sensitivity to vegetation activity. It can be seen that as NDVI values increase, the SAR2ET model's predictions for ET improve consistently. This relation suggests that the model's accuracy is positively correlated with the level of vegetation cover, irrespective of the soil texture information, which is in line with the observations in Table III. Furthermore, when examining specific soil texture parameters, such as clay, silt, and bulk density (bdod), a notable trend emerges. The SAR2ET model significantly performs better in predicting ET in highly

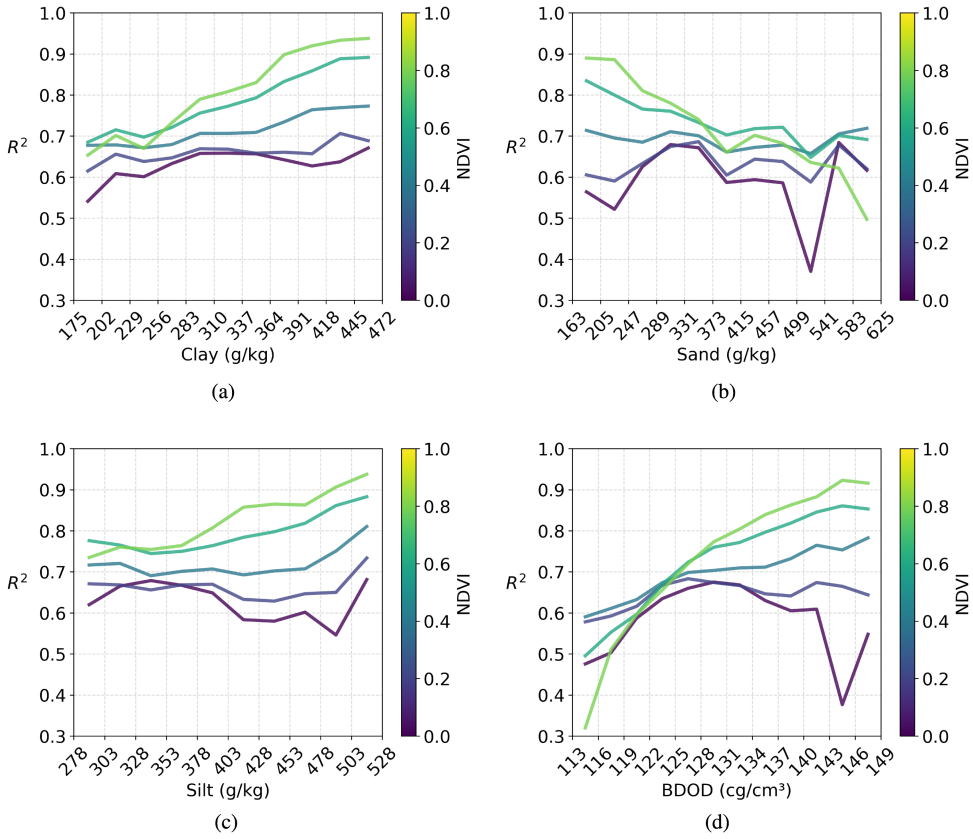

Fig. 6. Performance of the best-performing SAR2ET model with varying soil texture parameters and NDVI in terms of R^2 on the test set. (Best viewed in color.) (a) Clay. (b) Sand. (c) Silt. (d) Bulk density.

vegetated areas when there is a higher concentration of these soil components. Conversely, the model's performance is notably enhanced when the soil contains a lower proportion of sand in highly vegetated regions. This suggests that, for these areas, a decrease in the amount of sand content and an increase in the amount of clay content contribute to improved ET predictions due to the soil's water holding capacity.

V. Conclusion

The aim of this research was to understand how Sentinel-1 learns ET variability over croplands. To do this, we have explored end-to-end deep learning-based models for the estimation of ET values by harnessing the capabilities of SAR imagery. Toward this goal, we have proposed a carefully constructed benchmark to study SAR2ET in a systemic way based on SAR and auxiliary nonoptical meteorological and topographical data sources. We have also proposed a deep learning-based model that allows studying a variety of data source combinations.

The quantitative results highlight that SAR provides essential information for predicting ET values over croplands, and the potential is boosted by combining auxiliary data sources. The analyses over a time period and a set of regions show the consistency of model behavior across time and regions. Overall, we believe that the proposed SAR2ET model can be a valuable tool for agricultural monitoring, especially in areas with challenging weather conditions that limit the use of optical satellite data. We also believe the presented benchmark dataset provide a medium to accelerate the research in advancing SAR-based ET estimation.

Although the study provides valuable guidance for having regular ET over croplands, specifically for rainy regions, and emphasizes the critical role of comprehensive data integration in advancing our understanding of water-related processes in the environment, it is important to note that the accuracy of the model is directly proportional to the quality of its target variable, EEFlux. In this case, under clear sky conditions, high spatial resolution LST-based ET data are more accurate, as LST provides direct information about the amount of energy available for evaporation and transpiration processes. In addition, the good performance of the model is directly linked to the SAR signal's inherent sensitivity to the structure and water content of the soil-vegetation volume. The model may have limitations when applied to other land uses, such as urban-ET studies.

Inspired by the performance of the SAR2ET model, our next research will consider further improving the performance of the model by adding crop type information and SAR coherence as additional data modalities [25], [40], [41].

Appendix

A. Data Preprocessing Steps

In this section, we present all data preprocessing steps applied to data employed in this study (see Section-I), collectively.

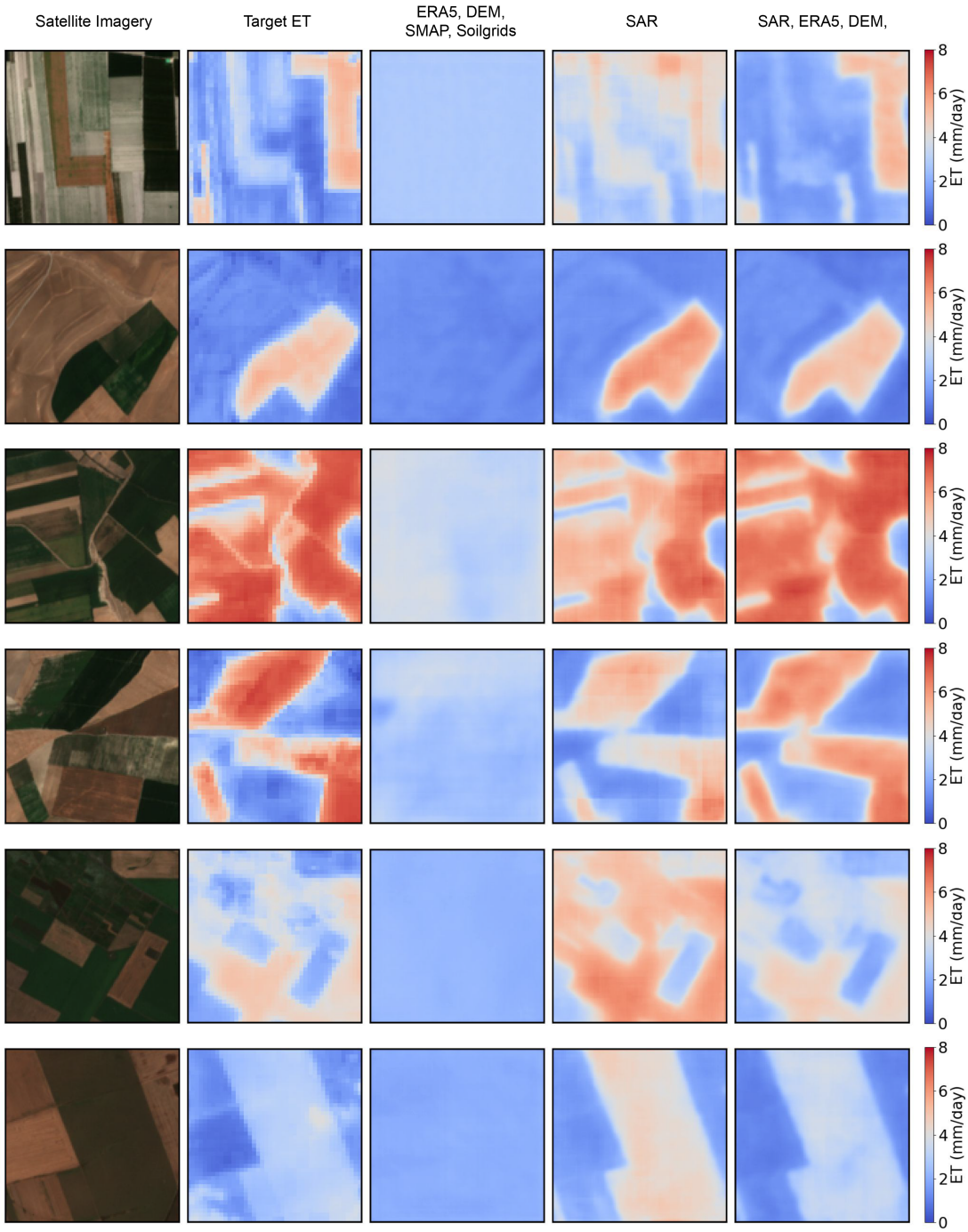

Fig. 7. Additional examples of test patches alongside their corresponding ET predictions from different SAR2ET models. Each row represents a different example, with the columns presenting an RGB image (captured using S2) of the patch area, the ground truth ET of the same area, and the ET predictions from SAR2ET models trained using auxiliary data sources only, SAR-only, and the highest performing model trained with SAR, ERA5, and DEM. (Best viewed in color. Red indicates higher ET while blue indicates lower ET in terms of mm/day.).

Downloading ground truth ET: We began by downloading all the available ET rasters from EEFlux that covered the year 2021 (March 1 to September 30, more specifically), for study regions shown in Fig. 2. In total, we downloaded 121 ET rasters for this year. We then upsampled the rasters to 10 m, which was a higher resolution than the original 30 m. These ET rasters served as our reference data points, or target ETs, for the rest of the dataset creation process; we obtained SAR and other auxiliary data from various sources and aligned and coregistered them for each of these 121 ET rasters.

Downloading SAR: We used GEE platform to obtain SAR rasters from Sentinel-1 (S1) ground range detected products with a 10-m resolution. First, we downloaded all the SAR rasters available within the time frame covering one week before and after the sensing date of each target ET. We selected the parameters of log-scaled VV, log-scaled VH, and CR for both

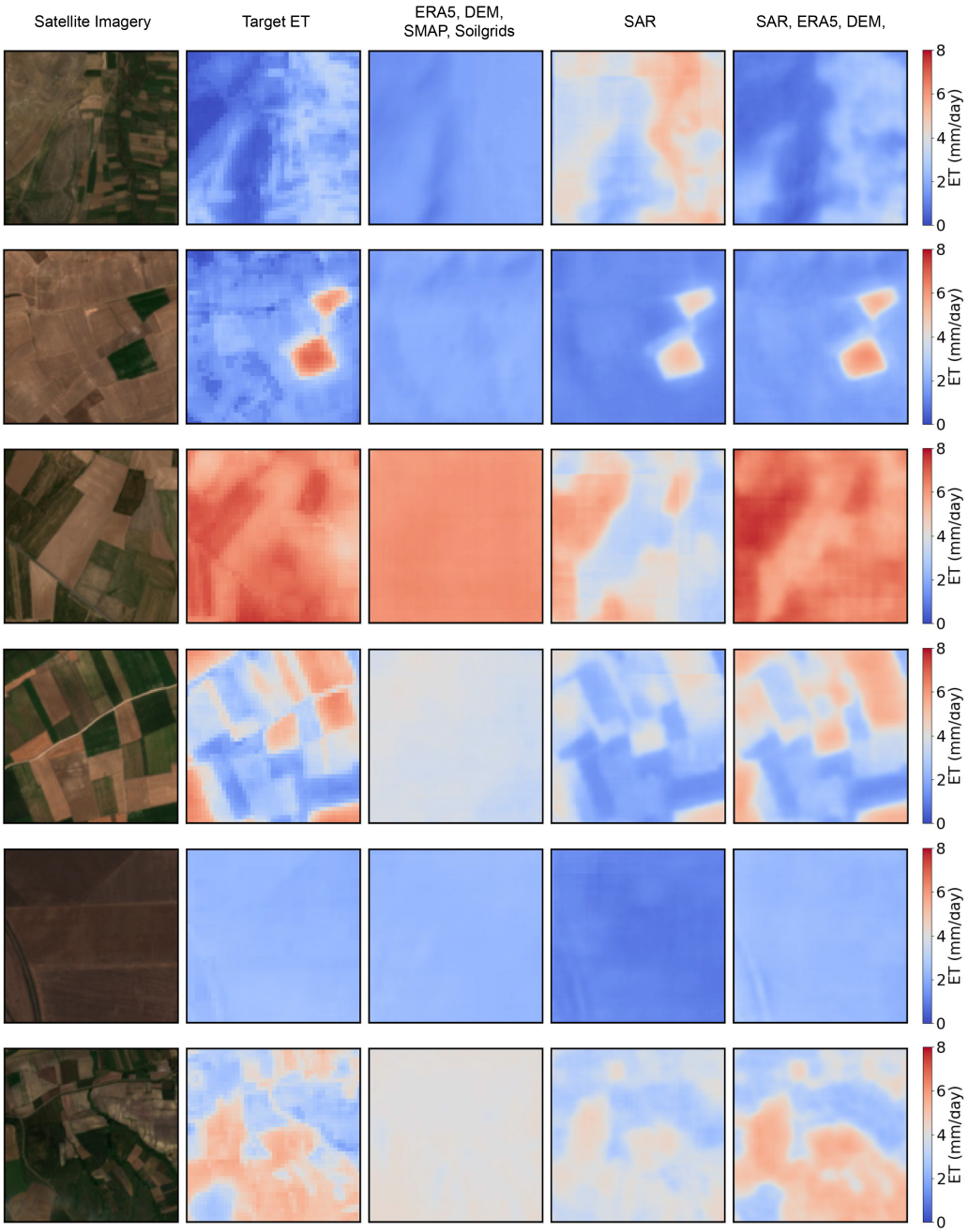

Fig. 8. Additional examples of test patches alongside their corresponding ET predictions from different SAR2ET models. Each row represents a different example, with the columns presenting an RGB image (captured using S2) of the patch area, the ground truth ET of the same area, and the ET predictions from SAR2ET models trained using auxiliary data sources only, SAR-only, and the highest performing model trained with SAR, ERA5, and DEM. (Best viewed in color. Red indicates higher ET while blue indicates lower ET in terms of mm/day.).

the ASC and DSC orbit directions, resulting in data images that have six channels in total. Then, we filtered out any backscatter values with an incidence angle below 30° or above 42°. We also removed backscatter values lower than −30° through edge filtering and applied spatial smoothing using a 5×5 average smoothing kernel. Finally, the SAR rasters are mosaicked based on their proximity to the date of ET data to obtain a single data image at the end to fill in any gaps that may have been created due to incidence angle filtering and edge filtering.

Downloading ERA5: We utilized GEE platform to download meteorlogical data rasters from ERA5-Land Daily Aggregated dataset corresponding to the date of target ET, in order to obtain meteorological data for each target ET. We then upsampled the rasters to 10-m resolution (original resolution was around 11 km). Our selection of meteorological parameters for the project included temperature (measured 2 m above the surface), dewpoint temperature (measured 2 m above the surface), total precipitation, U and V components of wind (measured 10 mabove the surface), surface net solar radiation, and surface pressure.

Downloading SMAP: We utilized GEE platform to download SM rasters from SMAP dataset corresponding to the date of

target ET, in order to obtain SM data for each target ET. We then increased the resolution of these rasters to 10 m, as the original resolution was around 9 km.

Downloading DEM: We used GEE platform to download DEM rasters from the ALOS World 3-D dataset. Later, we upsampled the rasters to 10-m resolution, as the original resolution was 30 m. For our project, we selected various topographical parameters, including height, slope, aspect, and hillshade. We converted the aspect parameter into two components, "cos" and "sin," to ensure numerical stability.

Downloading SoilGrids: We utilized GEE platform to acquire soil properties rasters from the SoilGrids dataset. Subsequently, we increased the resolution of the rasters to 10 m, since the original resolution was 30 m. For our project, we chose various soil property parameters, such as bulk density (bdod), clay, sand, and silt ratios of the fine earth component, all at a depth of 0–5 cm below the surface.

Downloading NDVI: We used GEE platform to obtain NDVI rasters from Sentinel-2 with a 10 m resolution. First, we downloaded all the available NDVI rasters within a time window of one week before or after the sensing date of the target ET. Then, we applied a cloud filtering method based on the QA band to discard rasters with more than 10% cloud coverage. Finally, we created a single data image by mosaicking NDVI rasters based on their proximity to the ET data date to prevent missing pixel values after applying cloud filtering.

Downloading cropland masks: We utilized the GEE platform to obtain cropland masks from the DW dataset in 10-m resolution. Initially, we downloaded all the available land cover masks covering the year 2021. Subsequently, we performed a pixel-wise mode operation in the time direction and accepted all the pixels classified as cropland consistently throughout the year as cropland pixels. Since the cropland mask is a binary mask, the remaining pixels were considered noncropland pixels.

Dataset construction: After we obtained 121 rasters from different data sources, which were spatially and temporally aligned, we then proceeded to extract patches of size 128×128 pixels from these rasters. During the extraction process, we made sure that the patches did not contain any invalid pixel values, such as no data and nonterrain pixels, or cloud pixels if the data source was optic satellite-based. We also excluded patches that did not contain any crop pixels, as determined by cropland masks. Once we completed the extraction process, we divided the data into three sets: train, validation, and test sets. We used a random splitting strategy that employed a 70%, 15%, and 15% ratio for the train, validation, and test sets, respectively. We ensured that the splitting strategy was applied uniformly to each individual ET raster and its corresponding rasters from different data sources. This ensured that the patches were distributed proportionally across different regions and dates for each split.

B. Additional Qualitative Results

In this section, we present additional figures, Figs. 7 and 8, illustrating the comparison between ground truth ET and predicted ET's made by different SAR2ET models. Each row shows a different example belonging to the test set, in the same format as in Fig. 4. The first column shows true-colored satellite imagery obtained from S2, while the next column shows the corresponding ground truth ET's. The remaining columns show predictions made by multiple SAR2ET models that are trained with different sets of data, as indicated at the top of each column.

Acknowledgment

The research presented in this article constitutes a part of the Samet Cetin Ph.D. thesis study at the Graduate School of Middle East Technical University.

References

[1] W. Zhu, L. Fan, and S. Jia, "Integration of microwave satellite soil moisture products in the contextual surface temperature-vegetation index models for spatially continuous evapotranspiration estimation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 203, pp. 211–229, 2023.

[2] G. Kayan, U. Türker, and E. Erten, "A fuzzy logic framework to handle uncertainty in remote sensing-based hydrological data for water budget improvement across mid- and small-scale basins," *Hydrological Processes*, vol. 36, no. 11, 2022, Art. no. e14740.

[3] J. Wang et al., "Enhancing land cover mapping in mixed vegetation regions using remote-sensing evapotranspiration," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4407322.

[4] M. F. Celik, M. S. Isik, G. Taskin, E. Erten, and G. Camps-Valls, "Explainable artificial intelligence for cotton yield prediction with multisource data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 8500905.

[5] Y. Li, H. Wu, H. Chen, and X. Zhu, "A robust framework for resolution enhancement of land surface temperature by combining spatial downscaling and spatiotemporal fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4404514.

[6] K. Sara, R. Eswar, and B. K. Bhattacharya, "The utility of simpler spatial disaggregation models for retrieving land surface temperature at high spatiotemporal resolutions," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6004305.

[7] P. Wu et al., "Spatially continuous and high-resolution land surface temperature product generation: A review of reconstruction and spatiotemporal fusion techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 112–137, Sep. 2021.

[8] M. Pieri et al., "Estimation of actual evapotranspiration in fragmented mediterranean areas by the spatio-temporal fusion of NDVI data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5108–5117, Dec. 2019.

[9] J. Xue et al., "Improving the spatiotemporal resolution of remotely sensed ET information for water management through Landsat, Sentinel-2, ECOSTRESS and VIIRS data fusion," *Irrigation Sci.*, vol. 40, no. 4, pp. 609–634, 2022.

[10] P. Dong et al., "Global comparison of diverse scaling factors and regression models for downscaling Landsat-8 thermal data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 44–56, 2020.

[11] Y. Xiao et al., "An integrated method for the generation of spatiotemporally continuous LST product with MODIS/Terra observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5001614.

[12] Y. Chen, Y. Yang, X. Pan, X. Meng, and J. Hu, "Spatiotemporal fusion network for land surface temperature based on a conditional variational autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5002813.

[13] Y. Zhang, Y. Yang, X. Pan, Y. Ding, J. Hu, and Y. Dai, "Multiinformation fusion network for mapping gapless all-sky land surface temperature using thermal infrared and reanalysis data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5001915.

[14] K. S. Mpakairi, T. Dube, M. Sibanda, and O. Mutanga, "Fine-scale characterization of irrigated and rainfed croplands at national scale using multi-source data, random forest, and deep learning algorithms," *ISPRS J. Photogrammetry Remote Sens.*, vol. 204, pp. 117–130, 2023.

[15] Z. Zhang, Y. Xu, Q. Cui, Q. Zhou, and L. Ma, "Unsupervised SAR and optical image matching using siamese domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5227116.

[16] N. Efremova, M. E. A. Seddik, and E. Erten, "Soil moisture estimation using Sentinel-1/-2 imagery coupled with CycleGAN for time-series gap filing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4705111.

[17] P. K. Musyimi, G. Sahbeni, G. Timár, T. Weidinger, and B. Székely, "Actual evapotranspiration estimation using Sentinel-1 SAR and Sentinel-3 SLSTR data combined with a gradient boosting machine model in Busia County, Western Kenya," *Atmosphere*, vol. 13, no. 11, 2022, Art. no. 1927. [Online]. Available: https://www.mdpi.com/2073-4433/13/11/1927

[18] S. Chintala, T. S. Harmya, B. Kambhammettu, S. Moharana, and S. Duvvuri, "Modelling high-resolution evapotranspiration in fragmented croplands from the constellation of Sentinels," *Remote Sens. Appl.: Soc. Environ.*, vol. 26, 2022, Art. no. 100704.

[19] X. Huang, W. Ju, Z. Xu, and J. Li, "A novel method for mapping moso bamboo forests using remote sensing data with the consideration of phenological status," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4406618.

[20] Q. Zhang, J. Cheng, and N. Wang, "Fusion of all-weather land surface temperature from AMSR-E and MODIS data using random forest regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2502705.

[21] A. Villarroya-Carpio, J. M. Lopez-Sanchez, and M. E. Engdahl, "Sentinel-1 interferometric coherence as a vegetation index for agriculture," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113208.

[22] M. F. Celik, M. S. Isik, O. Yuzugullu, N. Fajraoui, and E. Erten, "Soil moisture prediction from remote sensing images coupled with climate, soil texture and topography via deep learning," *Remote Sens.*, vol. 14, no. 21, 2022, Art. no. 5584.

[23] T. Jagdhuber et al., "On the potential of active and passive microwave remote sensing for tracking seasonal dynamics of evapotranspiration," in *2023 IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 2610–2613.

[24] A. Amazirh, O. Merlin, and S. Er-Raki, "Including Sentinel-1 radar data to improve the disaggregation of MODIS land surface temperature data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 11–26, 2019.

[25] N. Ouaadi et al., "Are the C-band backscattering coefficient and interferometric coherence suitable substitutes of NDVI for the monitoring of the FAO-56 crop coefficient?," *Agricultural Water Manage.*, vol. 282, 2023, Art. no. 108276.

[26] M. Alavi, M. Albaji, M. Golabi, A. Ali Naseri, and S. Homayouni, "Estimation of sugarcane evapotranspiration from remote sensing and limited meteorological variables using machine learning models," *J. Hydrol.*, vol. 629, 2024, Art. no. 130605.

[27] G. Kaplan, L. Fine, V. Lukyanov, N. Malachy, J. Tanny, and O. Rozenstein, "Using Sentinel-1 and Sentinel-2 imagery for estimating cotton crop coefficient, height, and leaf area index," *Agricultural Water Manage.*, vol. 276, 2023, Art. no. 108056.

[28] P. K. Musyimi, G. Sahbeni, G. Timár, T. Weidinger, and B. Székely, "Actual evapotranspiration estimation using Sentinel-1 SAR and Sentinel-3 SLSTR data combined with a gradient boosting machine model in Busia County, Western Kenya," *Atmosphere*, vol. 13, no. 11, 2022, Art. no. 1927.

[29] R. G. Allen et al., "Eeflux: A landsat-based evapotranspiration mapping tool on the Google Earth engine," in *2015 ASABE/IA Irrigation Symp.: Emerg. Technol. Sustain. Irrigation-A Tribute to Career Terry Howell, Sr. Conf. Proc.*, American Society of Agricultural and Biological Engineers, 2015, pp. 1–11.

[30] R. G. Allen, M. Tasumi, and R. Trezza, "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (metric)–model," *J. Irrigation Drainage Eng.*, vol. 133, no. 4, pp. 380–394, 2007.

[31] Z. Nisa et al., "Evaluation of SEBS, METRIC-EEFLUX, and qwatermodel actual evapotranspiration for a mediterranean cropping system in southern Italy," *Agronomy*, vol. 11, no. 2, 2021, Art. no. 345.

[32] S. A. Kadam, C. O. Stöckle, M. Liu, Z. Gao, and E. S. Russell, "Suitability of Earth engine evaporation flux (EEFlux) estimation of evapotranspiration in rainfed crops," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3884.

[33] J. Muñoz-Sabater et al., "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth Syst. Sci. Data*, vol. 13, no. 9, pp. 4349–4383, 2021.

[34] P. O'Neill et al., "Smap enhanced l3 radiometer global and polar grid daily 9 km ease-grid soil moisture version 5," NASA National Snow and Ice Data Center DAAC, Boulder, CO, USA, 2021.

[35] T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto, "Precise global DEM generation by ALOS prism," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 71–76, 2014.

[36] L. Poggio et al., "Soilgrids 2.0: Producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, 2021.

[37] C. F. Brown et al., "Dynamic world, near real-time global 10 m land use land cover mapping," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 251.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Interv. 2015: 18th Int. Conf., Munich, Germany*, 2015, pp. 234–241.

[39] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, Bellingham, WA, USA: SPIE, 2019, pp. 369–386.

[40] A. Chakir et al., "Analysis of C-band radar temporal coherence over an irrigated olive orchard in a semi-arid region," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3635–3647, 2024.

[41] S. Cetin, B. Ulker, R. G. Cinbis, and E. Erten, "Image-to-image translation networks for estimating evapotranspiration variations: SAR2ET," in *2024 IEEE Int. Geosci. Remote Sens. Symp.*, 2024, pp. 329–332.

Samet Cetin received the B.S. and M.S. degrees in computer engineering from Middle East Technical University, Ankara, Türkiye, in 2019 and 2022, respectively.

His research interests include the domains of weakly supervised learning and meta-learning for image classification, and image understanding from Earth observation imagery for agricultural practices.

Berk Ülker recieved the bachelor's and master's degrees from Electrical-Electronics Department, Middle East Technical University, Ankara, Türkiye, in 2012 and 2015, respectively. He is currently working toward the Ph.D. degree in electrics and electronics engineering with the Middle East Technical University, Eindhoven University of Technology, Eindhoven, The Netherlands.

His research interests include Earth observation, domain adaptation for machine learning models, and efficient deep learning techniques.

Esra Erten (Senior Member, IEEE) received the Ph.D. degree in computer vision and remote sensing from the Department of Computer Engineering and Microelectronics, Technische Universität Berlin, Berlin, Germany, in 2010.

She was with the High-Frequency Institute, German Aerospace Center (DLR), Cologne, Germany, during her Ph.D. study. After that, she was with the Chair of Earth Observation and Remote Sensing, Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland, where she worked on applied radar remote sensing for environmental parameter estimation. In 2018, she was a Visiting Professor with Satellite Applications Catapult, Oxford, U.K. She held a Sabbatical Professorship with the School of Engineering, The Open University, Milton Keynes, U.K., during 2018–2020 academic period. Currently, she is a Professor with the Faculty of Civil Engineering, Istanbul Technical University, Istanbul, Türkiye. Her recent research interests include information extraction and image understanding from Earth Observation imagery, in particular uncertainty qualification, multivariate statistics, and multidimensional synthetic aperture radar data, with application to agriculture.

Dr. Erten was an Associate Editor for IEEE Geoscience and Remote Sensing Letters and some special issues on remote sensing journals. She was the President of the IEEE GRSS Türkiye Chapter for the period 2015–2019 (she served as vice-president in 2013–2015).

Ramazan Gokberk Cinbis received the graduation degree in computer engineering from Bilkent University, Ankara, Türkiye, in 2008, the M.A. degree in computer science from Boston University, Boston, MA, USA, in 2010, and the Ph.D. degree from Université de Grenoble, Grenoble, France, in 2014.

He was a doctoral student with INRIA Grenoble, Grenoble, between 2010 and 2014. He is currently an Associate Professor with Middle East Technical University, Ankara, Türkiye. His research interests include data-efficient machine learning with minimal supervision, generative models, meta-learning, and vision-language integration.