Soil Moisture Estimation using Sentinel-1/-2 Imagery Coupled with cycleGAN for Time-series Gap Filing

Natalia Efremova, Mohamed El Amine Seddik, Esra Erten

***Abstract*—Fast soil moisture content (SMC) mapping is necessary to support water resource management and to understand crops' growth, quality and yield. Thereby, Earth Observation (EO) plays a key role due to its ability of almost real-time monitoring of large areas at a low cost. This study aimed to explore the possibility of taking advantage of freely available Sentinel-1 (S1) and Sentinel-2 (S2) EO data for the simultaneous prediction of SMC with cycle-consistent adversarial network (cycleGAN) for time-series gap filling. The proposed methodology, first, learns latent low-dimensional representation of the satellite images, then learns a simple machine learning model on top of these representations. To evaluate the methodology, a series of vineyards, located in South Australia's Eden valley are chosen. Specifically, we presented an efficient framework for extracting latent features from S1 and S2 imagery. We showed how one could use S1 to S2 feature translation based on Cycle-GAN using S1&S2 time series when there are missing images acquired over an area of interest. The resulting data in our study is then used to fill gaps in time series data. We used the resulting latent representations to predict SMC with various ML tools. In the experiments, cycleGAN and the autoencoders were trained with data randomly chosen around the site of interest, so we could augment the existing dataset. The best performance was demonstrated with random forest algorithm, whereas linear regression model demonstrated significant overfitting. The experiments demonstrate that the proposed methodology outperforms the compared state-of-the-art methods if there are missing optical and synthetic-aperture radar (SAR) images.**

***Index Terms*—Agriculture, Sentinel-1, Sentinel-2, Generative adversarial networks (GANs), unsupervised domain adaptation, Soil Moisture, Machine Learning**

I. Introduction

TODAY, more than ever, new technologies are released to increase efficiency and productivity in agriculture due to increasing food demands and decreasing freshwater sources. One of the many industries embracing precision agriculture solutions using big data analytics is the viticulture industry, which is growing rapidly and steadily. For this branch of horticulture, improving water efficiency is one of the most profound problems. The recent studies on water efficiency in viticulture, then, seek cost-effective ways to monitor soil moisture (SM) content. Specifically, during the last decades, a lot of work has documented the potential of Earth Observation (EO) data for soil moisture monitoring in agriculture due to their potential to supply spatio-temporal information over large areas and being complementary to *in-situ* data [1].

Soil moisture content (SMC) studies in agriculture using EO data can roughly be categorised under two different approaches. The first approach, dating back to the 1980's with Landsat, considers land surface temperature (LST) mapping using thermal information, which is directly related to local moisture variations [2]. Although many studies have demonstrated the benefits of using LST in agriculture, the temporal and spatial resolution of the EO satellite data are too coarse for field scale studies, specifically for horticulture. When it comes to the operational orchard monitoring, the implementation of LST method created from a thermal camera is limited to the air-borne and unmanned aerial vehicle acquisitions [3], [4]. The second approach takes into account spectral and backscattering changes in visible/near-infrared and microwave domain, respectively, and relates this information to water stress by data-driven and physical models. While there are successful applications with high-spatial satellite acquisitions, one of the main limitations for their operational usage was their cost coupled with the limited temporal resolution. In this context, ESA Copernicus mission supplies free accessible radar Sentinel-1 (S1) and optical Sentinel-2 (S2) images with approximately weekly temporal resolution, with large amounts of data available for regular monitoring [5], [6], [7], [8]. For SMC estimation on field scale, C-band S1 data coupled with Landsat thermal data was successfully utilised [9], [10] using data-diriven machine learning (ML) techniques. Similarly, a neural network (NN) inversion [11] and backscattering change analysis [12] were implemented to estimate SMC by considering only S1 data. Recently, [5] underlines that regression based approaches have better accuracy for SMC estimation than those by the semi-empirical SAR and optical models over farmlands. However, none of the studies covered the direct measurement of SMC on the orchard scale and more importantly they applied the standard ML on the stacked images, thus, neglecting the temporal dependencies of features extracted from EO data.

There are several overarching methodologies that would recur when applying machine learning for agricultural tasks, particularly the use of convolutional neural networks (CNN)

This work was supported by the SPRINT project. Corresponding author: Esra Erten

N. Efremova is now with the University of Queen Mary, London and was with University of Oxford, Oxford, UK (e-mail: n.efremova@qmul.ac.uk)

Mohamed El Amine Seddik is with LIX Ecole Polytechnique, Palaiseau, France (e-mail: mohamed-el-amine.seddik@polytechnique.edu).

E. Erten is with the Civil Engineering Faculty of Istanbul Technical University, Istanbul, Turkey and was with the Faculty of Science, Technology, Engineering & Mathematics, The Open University, Milton Keynes, UK (e-mail: eerten@itu.edu.tr).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2021.3134127, IEEE Transactions on Geoscience and Remote Sensing

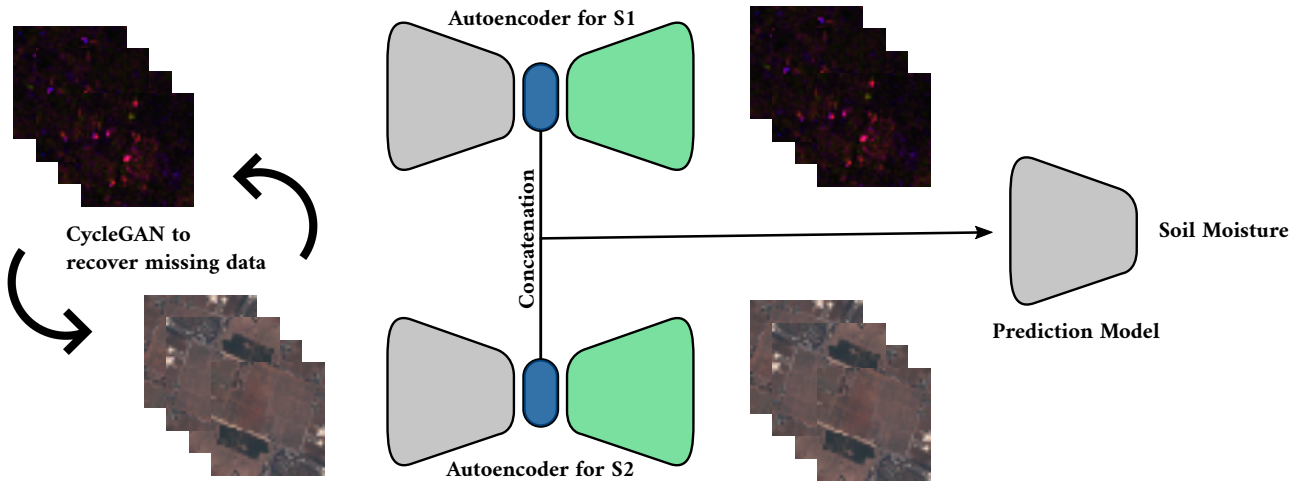

Fig. 1. The proposed architecture relies on building a latent representation of each domain (S1 and S2) based on autoencoders and then predicting Soil Moisture (through an additional prediction model) based on these representations. A cycleGAN is also trained to recover missing data from S1 to S2 and vice-versa.

for image classification and the usage of normalized difference vegetation index (NDVI) for vegetation health. For example, [13] calculated NDVI and then predicted it into the future using a Long-Short Term Memory (LSTM), performing per-pixel predictions to help minimize the impact of droughts. [14] used conditional Generative Adversarial Nets (GANs) for modelling cloud reflectance fields using Conditional Generative Adversarial Networks. Clouds are one of the biggest problems when it comes to doing analysis on remote sensing data, often making entire data periods unusable. Methods of classifying and removing clouds are still in the early stages, but could revolutionize satellite imagery analysis. In this work, authors used generative adversarial networks to generate simulated clouds with good reflectance values which could be used for future training data or other tasks. In [15], authors used LSTM networks to predict soil moisture interpolations into the future using EO data.

Across all the research domains, there is a big problem of lack of data to train ML models, since training requires a lot of sequential data [16], [17]. Such models have already taken attention in EO community in classification and global scale - low resolution - SMC data analysis [18], [19], [20], [21]. Utilising S1 or S2 imagery alone is often not enough for this purpose, therefore we aim to extract features from these two sources to combine them in time series. We use GANs to extract latent representation from both of these imagery sources [22]. Once latent low-dimensional representation of the satellite images is learned, missing optical features are reconstructed by temporal and spatial dependencies. GANs and autoencoders approaches are extensively applied for image to image translation [23], [24], [25]. Recently, its great potential in EO domain, specifically on data fusion between optical and radar images, has also been shown [26], [27], [28]. We believe that this approach can be applied also in biophysical parameter estimation using EO data.

Following this line, the contributions of this work can be formulated as following:

- we explore the potential of various ML architectures, which consider to spatial and time dependencies among the EO measurements, specifically S1/S2, to estimate SMC;
- we explore the feasibility of using GANs for data augmentation for training ML models;
- we propose an efficient framework for unsupervised deep domain adaptation for S1 and S2 satellite imagery with cycleGANs [29].

II. Proposed Architecture Formulation & Objective

Variational autoencoders (VAEs) and GANs are effective for image-to-image translation, where pairs of images are not readily available [30]. This is true for the case where we have pairs of images from S1 and S2, where the difference between image acquisition can vary from a few hours up to a few days. We employ this property of GANs to extract the features, meaningful for prediction of SMC from both sources of imagery.

Therefore, the main goal of this study is to investigate the problem of predicting SMC based on the combination of satellite images for S1 and S2. As illustrated in Figure 1, since the amount of available *in-situ* measurements might be considerably limited in practice, we propose to learn latent (low-dimensional) representations based on autoencoders for S1 and S2 images respectively. The autoencoders are specifically learned through a reconstruction problem with ℓ_2-loss for building such representations. These representations are then concatenated and fed into a machine learning model for predicting soil moisture. A technical difficulty arises then from the fact that some SM measurements might have corresponding images from either S1 or S2, and so we end up having missing data. To leverage this technical issue, we make use of the cycleGAN model to recover the missing data through an unsupervised manner.

The cycleGAN method is an unsupervised deep learning model which consists in learning two mappings $F : \mathcal{X} \rightarrow \mathcal{Y}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2021.3134127, IEEE Transactions on Geoscience and Remote Sensing

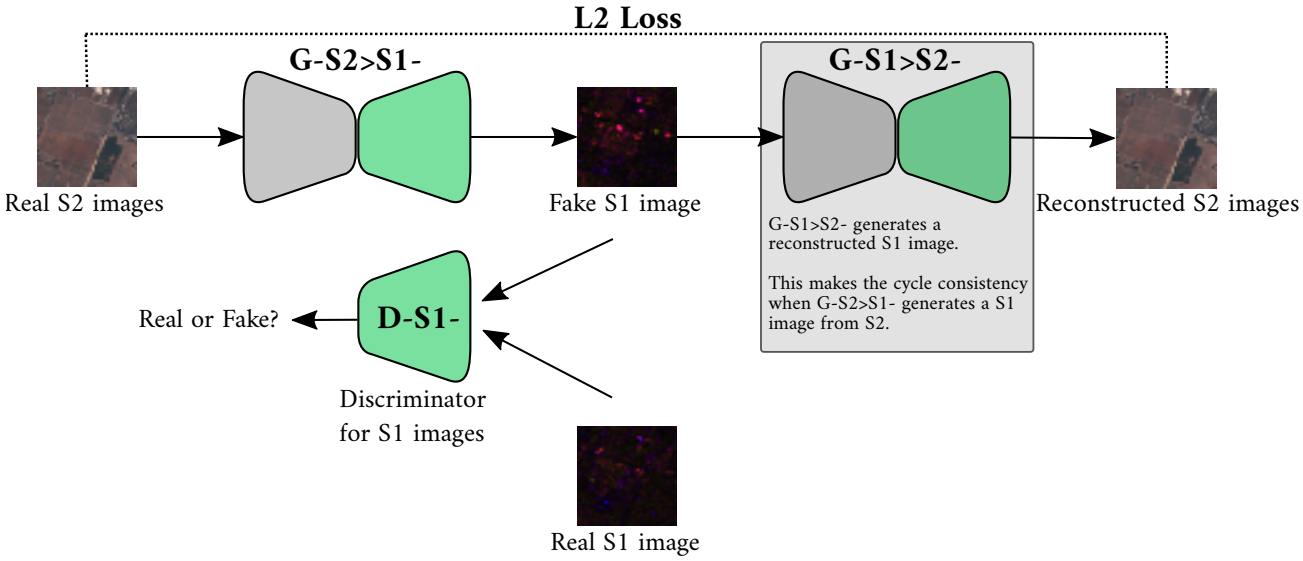

Fig. 2. Architecture of cycleGAN.

and $G : \mathcal{Y} \rightarrow \mathcal{X}$ in order to translate images from a domain $\mathcal{X}$ to a domain $\mathcal{Y}$ and vice-versa. Denoting by $p_{\mathcal{X}}$ and $p_{\mathcal{Y}}$ the probability distributions of the domains $\mathcal{X}$ and $\mathcal{Y}$ respectively, the cycleGAN objective function [29] is given by:

$$\mathcal{L}(F, G, D_{\mathcal{X}}, D_{\mathcal{Y}}) = \mathcal{L}_{gan}(F, D_{\mathcal{Y}}) + \mathcal{L}_{gan}(G, D_{\mathcal{X}}) + \lambda \mathcal{L}_{cyc}(F, G), \quad (1)$$

where $\mathcal{L}_{gan}(G, D)$ denotes the classical GAN [31] loss function involving a generator G and a discriminator D, whereas $\mathcal{L}_{cyc}(F, G)$ stands for the cycle loss which is given by

$$\mathcal{L}_{cyc}(F, G) = \mathbb{E}_{x \sim \mathcal{X}} \left[\|G(F(x)) - x\|_1 \right] + \mathbb{E}_{y \sim \mathcal{Y}} \left[\|F(G(y)) - y\|_1 \right], \quad (2)$$

and λ is a hyper-parameter. Both the mappings F and G are trained simultaneously adding a cycle consistency loss [32].

Figure 2 depicts the architecture of the cycleGAN model, an ℓ_2 reconstruction loss is applied to the mappings $G_{S2}(G_{S1})$ and $G_{S1}(G_{S2})$, while a discriminator for domain D_{S1} tries to fit the images distribution. In the following sections, we demonstrate the application of this method to translate satellite images from S1 to S2 and vice versa.

III. Study Area and Problem Definition

A series of Sentinel-1/2 images were used from early 2017 to April 2020 to cover SM measurements for 200 acres of vineyards located in the Upper Hunter Valley region in Australia. Ground Range Detected (GRD) Sentinel-1 backscattering data at HH and HV polarizations were collected during both ascending and descending orbits. 10 m and 20 m spatial resolution Sentinel-2 bands were downloaded in Level 2A, which provides a shadow and cloud mask and top of canopy reflectance. In addition to the backscattering and spectral reflectance values, backscattering ratio and vegetation index, namely Normalized Difference Vegetation Index (NDVI) were formed. All these continuous different resolution variables were resampled to 10×10 m resolution according to the modelling process.

As *in-situ* input data, we use SM measurements installed on 200 hectares of land. Input measurements are taken from the embedded soil moisture sensors, each installed at 10 centimetres depth up until 120 cm depth, rainfall and temperature data. Fig. 3 shows the in-situ SM measurements acquired simultaneously (on the same day) with S1 (top) and S2 (bottom). Despite high temporal resolution of S1 and S2, it can be easily seen that it is difficult to have a combined continuous time series of S1 and S2 data, which is essential for agricultural studies due to their different sensitivities to different crop's biophysical properties [7], [11], [33].

IV. Experimental Results

In order to estimate SMC over vineyards, the proposed architecture is applied in three steps. Firstly, the training strategy of cycleGAN for image translation is introduced for recovering missing S1 or S2 images. Secondly, low-dimensional representations of the data are extracted through pre-trained autoencoders. With this low-dimensional representation data, SMC estimation is finally performed using various prediction models which are presented with their respective accuracy assessments.

A. Image translation between Sentinel-1/2

This section presents some qualitative results when performing the cycleGAN model on Sentinel-1/2 data. For the cycleGAN training process, 5×1500 train images (corresponding to each of the 5 used SM sensors locations) and 5×450 test images were used. The training is performed for 100 epochs with a batch size 10 and learning rate of 0.0002. Figure 4 depicts the results for the translation from S1 to S2 and vice-versa. We can see from Figure 4 that the model can extract some marginal S2 features from the S1 counterpart. However, when translating from S2 to S1, the performance of the model

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2021.3134127, IEEE Transactions on Geoscience and Remote Sensing

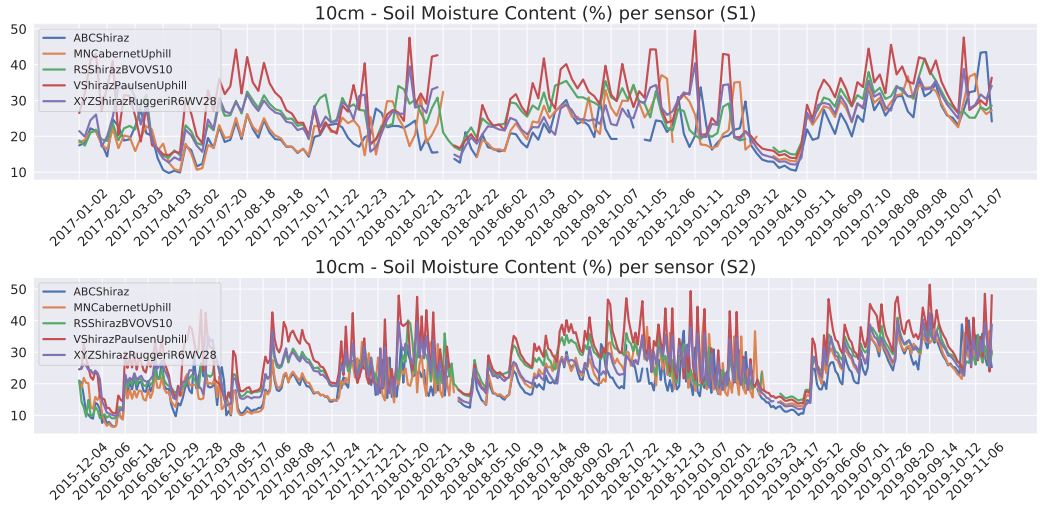

Fig. 3. Soil moisture sensor measurements acquired in five different sites corresponding to S1 and S2 from top to bottom respectively.

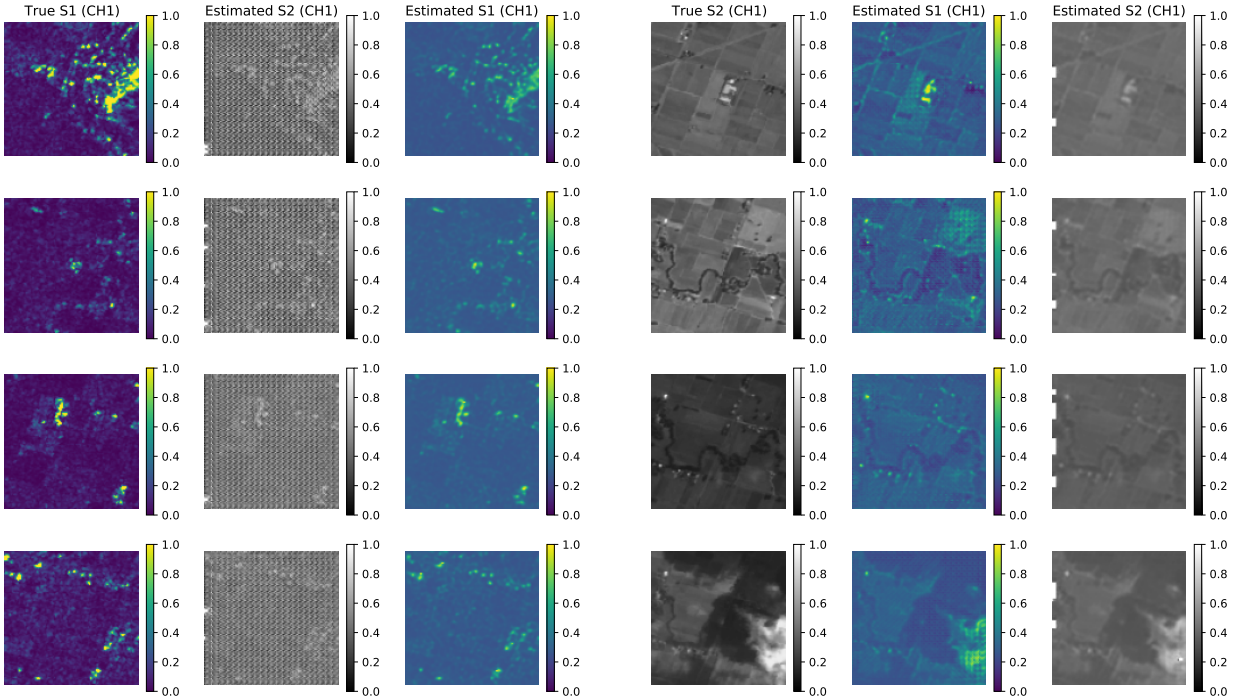

Fig. 4. We have generated Sentinel-2 true colour images from Sentinel - 1 imagery and vice versa. The learned features were used for soil moisture content prediction. (left) Translation from S1 to S2. (right) Translation from S2 to S1. For each pixel, we used the grey-scale value (CH1 on the images).

decreases (Figure 4). On the other hand, when we consider NDVI components, the model seems to generalise well when recovering S1 features as depicted in Figures 5 and 6. The key role of feature-adapted solutions based on CNN is also underlined in recent studies [33], [34], [35] when there is no available (fully cloudy condition) training data at a certain time for dynamic monitoring of agricultural fields.

B. Learning low-dimensional representations

For each domain (S1 and S2 data) we use autoencoders to extract low-dimensional representations (of dimension 784). We train the autoencoders using 7740 satellite images of size 100×100 pixels. Examples of the used images to train the autoencoder for NDVI features are depicted in Figure 7.

Figure 8 shows reconstruction results of the autoencoder on NDVI data. We can see that the model preserves information about the original data while compressing it in a low dimensional representation obtained by the encoder model (of dimension 784). Figure 9 depicts the training and testing loss of the autoencoder. For autoencoders, we used Adam optimizer and the mean squared error (MSE) loss function. During our experiments, we have tried cross-entropy loss, however it did not add any improvement to the results. The resulting training loss achieved was 0.00086614 and validation loss 0.00093009 on 150 epochs of training.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2021.3134127, IEEE Transactions on Geoscience and Remote Sensing

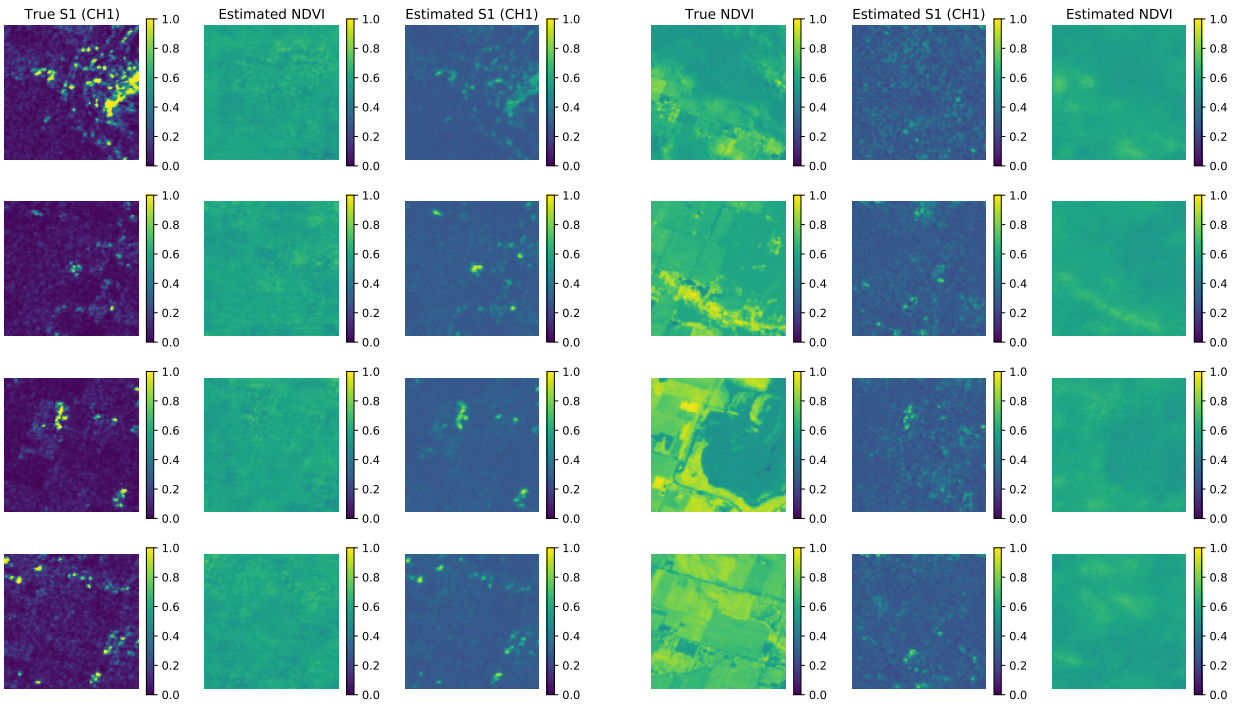

Fig. 5. (left) Translation from S1 (channel 1, VH) to NDVI images. (right) Translation from NDVI to S1 (channel 1, VH) images.

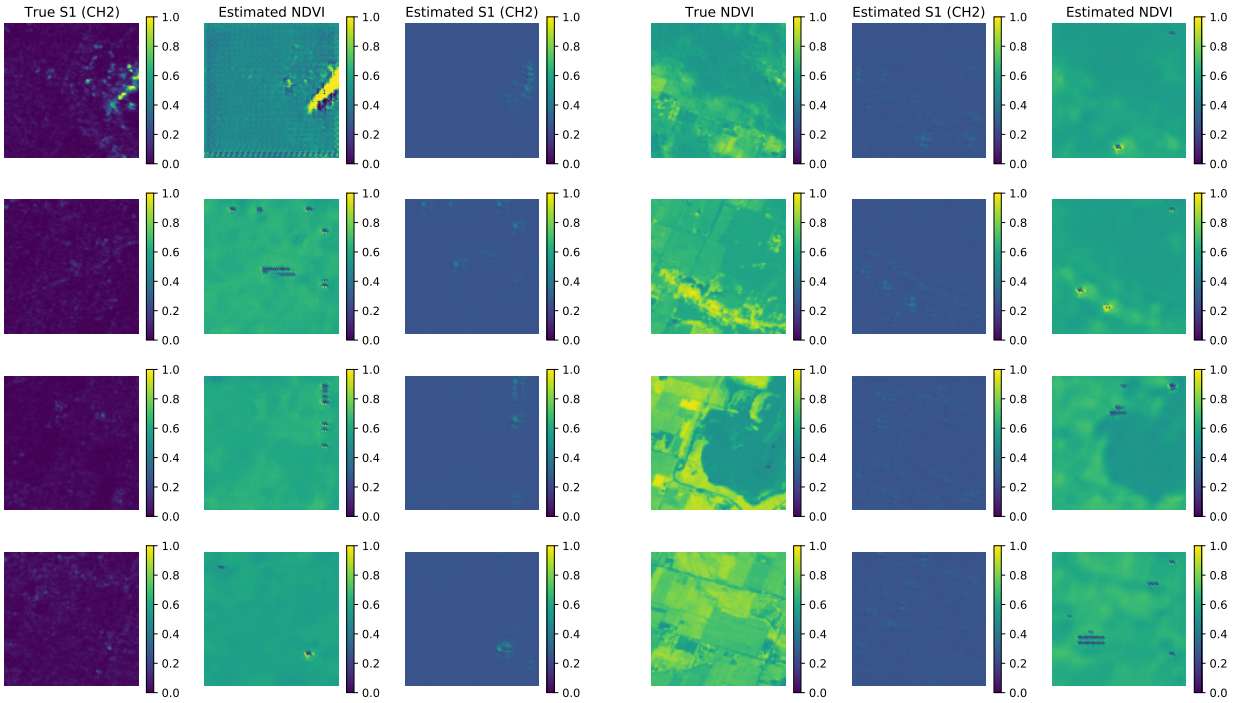

Fig. 6. (left) Translation from S1 (channel 2, VV) to NDVI images. (right) Translation from NDVI to S1 (channel 2, VV) images.

It is worth noting that the dimensionality reduction can be performed with well-known linear principal component analysis (PCA) as well, however nonlinear autoencoders can learn more powerful features for a given dimensionality. These results demonstrate only marginally better performance of autoencoders on image data [36]. However, in a different scenario (by exploiting other features than S1 and S2 images), the relationship between the raw features space and the latent space could be highly non-linear in which using autoencoders could be more beneficial. Our purpose is to propose a more general framework which does not restrict to linear spaces, therefore we use autoencoders for dimensionality reduction.

C. Soil Moisture Prediction

As a final step we compare the performance of a few popular machine learning models on this task. We use simple

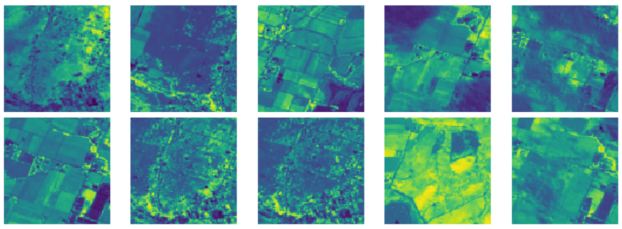

Fig. 7. Example of NDVI images used to train the autoencoder model.

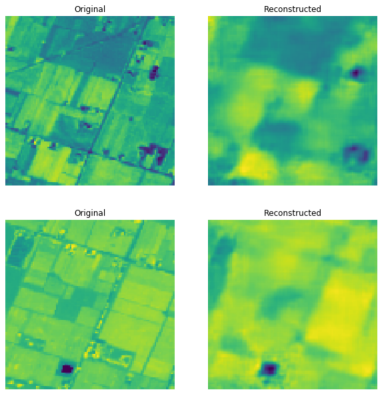

Fig. 8. Reconstruction of NDVI images with the autoencoder model.

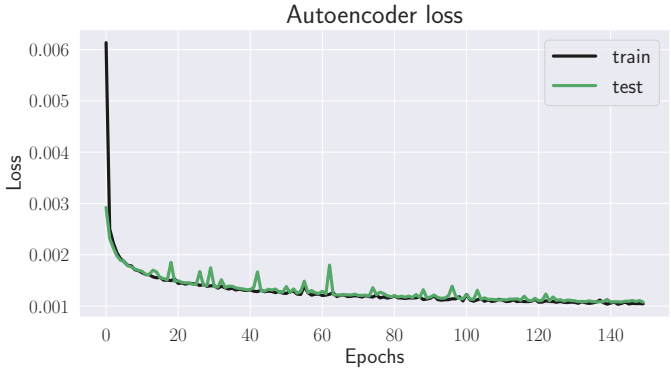

Fig. 9. Train and test loss of the auto-encoder on NDVI images.

linear regression, ridge regression, kernel ridge regression, support vector regressor and random forests to predict soil moisture from the time series that we have constructed with the cycleGAN model. A summary of the comparison among the regressors in SMC estimation (see Figure 3) is presented in Figure 10 and quantified in Table I with the MSE. Figure 10 shows the predictions of these models on the train and test data. The parameters of each regression method were optimally selected based on the five-fold cross validation.

1) Neural Networks (NN): First, we use a neural network model to predict the soil moisture on the test site. We use a simple fully connected neural network (multi-layer perceptron, MLP) with 5 layers (trained with Adam optimiser) to build a model for soil moisture prediction. Figure 3 shows the corresponding soil moisture measurements on the different sensors of the site (5 in total). Using these data and the corresponding satellite data we end up with a dataset of 578 training samples and 250 testing samples for a 30% test split. Once the encoder model is trained, the images are transformed to vectors of dimension 784 and the following model is used to predict the SM measurements.

2) Linear Regression (Linear): As a baseline model, we use a simple linear regression model. It is a linear model with coefficients $w = (w_1, ..., w_p)$ to minimise the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation. We train all the models on the same data.

3) Random Forest (RF): Additionally, we train a random forest regression model. This model is a meta estimator that fits a number of regression decision trees on various sub-samples of the same dataset and uses averaging to improve the predictive accuracy and control over-fitting.

4) SVM Regressor (SVR): Support vector machine (SVM) regression is a non-parametric machine learning model, which relies on kernel functions. SVR approach depends on some subset of the training data, called the support vectors. Radial basis kernel function was used for the SVR.

5) Ridge Regression and Kernel Ridge Regression: Finally, we train ridge regression and kernel ridge regression models. Ridge regression model differs from simple linear regression by the addition of regularisation, given by the $l2$-norm. Kernels are used to calculate the inner product of two vectors in a feature space. We add kernels to ridge regression to add non-linearity. This increases the accuracy of the model on our dataset.

The RF model performs better than the other algorithms, neural networks and kernel regression give slightly similar results and then comes Ridge regression, whereas the linear regression model demonstrated significant overfitting, while the other methods do not perform well at all.

In order to justify the usage of such a complex architecture, we have run a series of experiments with GANs and CNN, avoiding dimensionality reduction. However, one of the main drawbacks of CNNs is that they need a lot of data to perform well enough. In this case, the resulting amount of images was not enough to successfully train a CNN to predict SMC. To overcome the issue of not having enough data, we proposed to split the inference part into two different steps: dimensionality reduction followed by ML classifiers, thereby offering the possibility to work with a limited amount of data.

V. Discussion on GAN performance evaluation

Evaluating GANs is a challenging task and an active area of research that's seen a lot of progress in the last few years ([37], [38], [39], [40], [41]). On one hand, it is similar to evaluating other models by comparing its output against some metrics, which are used across models. On the other hand, there is an aspect that makes evaluating GANs challenging. In contrast to other deep learning models, the loss of a model is not descriptive of its performance. For example, in classifiers low loss on a test set indicates superior performance, whereas a low loss for the generator or discriminator suggests that learning has stopped. Therefore, there should be metrics that evaluate images on their quality. In GANs, to evaluate quality of the

TABLE I
PERFORMANCES OF THE DIFFERENT MODELS IN TERMS OF MSE.

Models	NN	Linear	RF	Ridge	Kernel	SVR
Train	$1.66 \cdot 10^{-4}$	$\mathbf{2.75 \cdot 10^{-10}}$	$7.31 \cdot 10^{-4}$	$1.32 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$3.92 \cdot 10^{-3}$
Test	$7.57 \cdot 10^{-3}$	$1.57 \cdot 10^{-1}$	$\mathbf{7.35 \cdot 10^{-3}}$	$8.34 \cdot 10^{-3}$	$7.57 \cdot 10^{-3}$	$1.08 \cdot 10^{-2}$

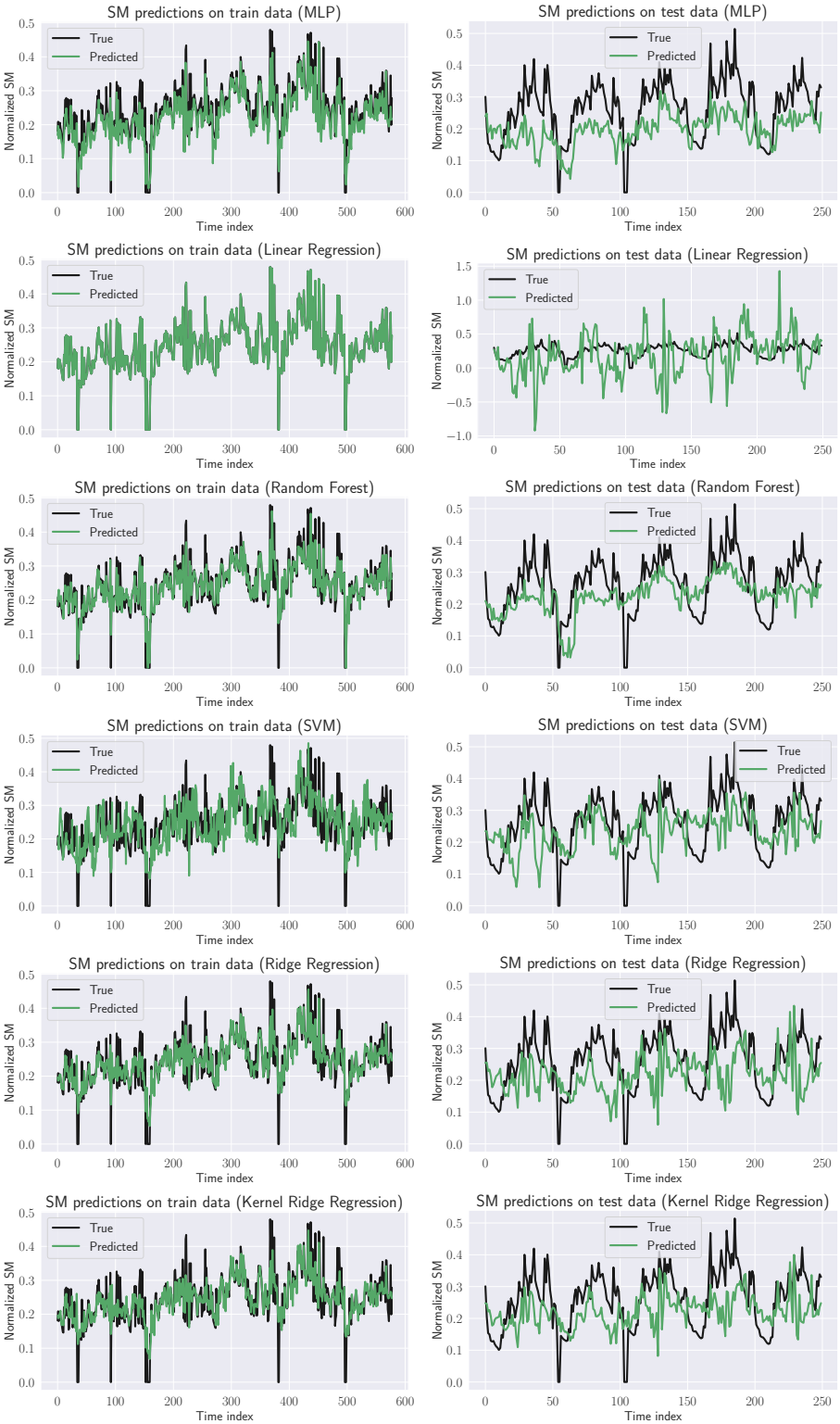

Fig. 10. Predictions of the SM measurements on the train data (left) and test data (right).

generated image, we need to access it across 2 dimensions: fidelity (how realistic do generated images look) and diversity (whether the generator is able to produce the diversity of images that's inherent in the training data set).

In generative modelling, we are given a dataset of samples x drawn from some unknown probability distribution $p_r(x)$, where r stands for 'real'. We use the samples x to derive the unknown real data distribution $p_r(x)$. A generative model G encodes a distribution over new samples, $p_g(x)$, where g stands for 'generated'. The aim is that we find a generative distribution such that $pg(x) \approx pr(x)$ according to some metric.

The most popular metric, used so far in generative models, is inception score ([38]. The Inception Score (IS) is a metric for automatically evaluating the quality of image generative models [37]. It uses Inception-V3 model[42]), pre-trained on a large dataset of general-purpose images such as ImageNet [43]. The IS calculates a statistic of the network's outputs when applied to generated images, in other words the pre-trained Inception-V3 model is used to classify generated images and the quality of images is assessed on the accuracy of the predicted class. Formally, the IS can be written as following:

$$IS(G) = exp(\mathbb{E}_{x\sim p_g} D_{KL}(p(y|X) \parallel p(y))), \tag{3}$$

where $X \sim p_g$ indicates that x is an image sampled from p_g, $D_{KL}(p \parallel q)$ is the KL-divergence between the distributions p and q, $p(y|X)$ is the conditional class distribution, $p(y) = \int_X p(y|X)p_g(X)$ is the marginal class distribution [37]. IS metric is widely applied in literature, however, it was recently found that applying the Inception Score to generative models trained on datasets other than ImageNet gives misleading results. Barratt and Sharma [38] showed that the results of IS performance even on a dataset, close to ImageNet, such as CIFAR-10, can be not good since the classes in ImageNet and CIFAR-10 do not line up identically. Therefore, this metric cannot be, since satellite imagery classification is very far from ImageNet in terms of both classes it is trained on and band resolution of images (ImageNet has 3 bands (RGB), whereas Sentinel imagery has 12). Even when we use only the True color version of the Sentinel-2 imagery, the classification for any possible class will be far away from ImageNet. The only option to use this metric is to train Inception-V3 on satellite imagery for some related classification task, but this is not possible without a large dataset for satellite imagery classification.

Another metric, used to evaluate GANs, was developed recently, called Fréchet Inception Distance (FID) [39]. It was proposed as an improvement over Inception Score. Similarly to IS, FID uses the Inception-V3 network as part of its calculation. However, instead of using the classification labels of the Inception-v3 network, it uses the output from a feature layer. Research has shown that deep convolutional neural networks trained on difficult tasks, like classifying many classes, build increasingly sophisticated representations of features going deeper into the network: the first few layers may learn to detect different kinds of edges and curves, also colour and texture, whereas the later layers respond to increasingly more complex stimuli, including parts of objects that they were trained to recognise. FID uses features from the Inception-V3 model, extracted from the last pooling layer of this model and represent the most high-level features of the model, which it is able to recognise. These features are called embedding in the case of FID. Using multivariate normal distribution, FID compares the distribution between the real and generated images and represents the results in terms of this distance (i.e. the smaller is the resulting metric, the closer the generated images are from the real ones). Formally, to calculate this distance between two normal distributions with means and standard deviations, we use the following equation [44]:

$$\begin{aligned} FID = &\parallel \mu_X - \mu_Y \parallel^2 + \\ &Tr(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y}), \end{aligned} \tag{4}$$

where μ_X is the mean of the real embeddings, μ_Y is the mean of the generated embeddings, Σ_X is the covariance matrix of the real embeddings, and Σ_Y is the covariance matrix of the generated embeddings. FID is currently the most widely used GAN evaluation metric. However, similar to IS, it used a pre-trained Inception-V3 model, which does not capture the features important for satellite imagery.

A different type of metric, proposed recently, is HYPE evaluation score [40]. HYPE displays a series of images one-by-one to crowdsource evaluators on Amazon Mechanical Turk (MTurk) [45], it asks the evaluators to assess whether each image is real or fake. We cannot use this method, since the person who evaluates the satellite images needs to have special GIS training and using MTurk is infeasible in this case.

Finally, the precision and recall metrics for GANs can be used for their evaluation. In [41], authors define the precision and recall metrics as follows. If we denote the distribution of real images with P_r and the distribution of generated images with P_g, precision is the probability that a random image from P_g falls within the support of P_r and recall is the probability that a random image from P_r falls within the support of P_g. In other words, precision is the ratio of the generated images that look real to all of the generated images and recall is how much overlap between the samples is divided by all of the real samples. The estimates are obtained by calculating pairwise Euclidean distances between all feature vectors in the set and, for each feature vector, forming a hyper-sphere with radius equal to the distance to its k_th nearest neighbour. Together, these hyper-spheres define a volume in the feature space that serves as an estimate of the true manifold. To determine whether a given sample is located within this volume, a binary function is used. Precision is quantified by querying for each generated image whether the image is within the estimated manifold of real images. Symmetrically, recall is calculated by querying for each real image whether the image is within an estimated manifold of generated images. The feature vector for a given image is computed by feeding it to a pre-trained VGG-16 classifier [41], this method accessible for the reasons described above. To perform similar classification, we needed to train our own classifier. We used a classifier network that we trained on the dataset, obtained in one of our previous studies. The model was trained on a large dataset of 7000 vineyard

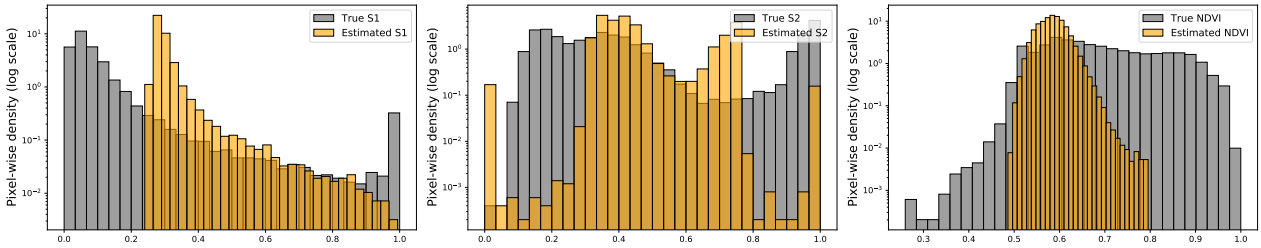

Fig. 11. Pixel-wise histograms of ground truth images and generated images with the cycleGAN model: (left) comparison of a real Sentinel-1 (S1) image with a generated one; (center) a real Sentinel-2 (S2) image with a generated one; (right) comparison of an NDVI image, calculated from a real Sentinel-2 image, with a generated NDVI image.

blocks in the Australian region. However, the results for this analysis were significantly lower than the state of the art results for GAN models, trained on large datasets. We believe that the reason for such a performance is that the features that were learned by our classifier (vineyard detection) differ from the ones that are required for soil moisture estimation.

To summarise, the best way to estimate the performance of the proposed GAN is to visually estimate the fidelity of the image. Fig.4 - Fig.6 demonstrate reconstruction results for S1 to S2 true-colour images, S1 channel 1 to NDVI and S1 channel 2 to NDVI. To demonstrate the quality of the network performance, we translated S1 images to S2 images and indices and back. Left-most and right-most images show real and reconstructed images, respectively. We can see that the translation is not perfect, but it can reproduce essential elements of the image. We do not use the right-most images from Fig.4 - Fig.6 in our simulations, they are shown to demonstrate the quality of the model.

We can analyse the differences between the images in a quantitative manner only to some extent. Given the complexity of converting images from S1 to S2, the cycleGAN model does not allow to generate fully realistic images but tends to approximate them given the input. As such, our model learns more discriminative features for the missing image domain compared to not considering such a domain. Figure11 shows pixel-wise histograms of ground truth images and generated images with the cycleGAN model. One can particularly notice that the model performs better in approaching the density of S1 images compared to the other image domains. The difference in time between image acquisition increases the discrepancies between the original and the generated images.

VI. Conclusions

In this paper, cycleGAN methodology is proposed for monitoring any biophysical parameters using S1 and S2 data. In particular, the limitations and potentials of SMC estimation using S1 and S2 data, for the first time, with the GAN-based architecture is explored. To our knowledge, this is the first time autoencoders and cycleGANs were used together to obtain features from Sentinel-1 and Sentinel-2 imagery.

This research demonstrated the feasibility of the proposed approach and allows us to fuse two different sources of information for efficient prediction of soil moisture content. Being one of the major horticulture crops consumed worldwide, grapevine is chosen to assess SMC estimations in horticulture from space. In-situ measurements conducted simultaneously with satellite acquisitions provide a discussion on the feasibility of Sentinel-1/-2 data to retrieve soil properties of vineyard, specifically SMC. The results of the experiments on the data set demonstrate that the proposed methodology is effective in regression based biophysical parameter estimation in agriculture studies, in which the possibility of having simultaneous radar and optical image acquisitions is critical due to their complementary information.

Although the proposed architecture (cycleGAN coupled with autoconders) in this paper is promising for SMC estimation in vineyards, the similar performance could be achieved with simpler ML model for different crops with different conditions (e.g., crop specific irrigation, planting patterns, the presence of dense in-situ data and etc.). Furthermore, the regression based biophysical parameter estimation using EO data is widely accepted in plantation agriculture. However, when it comes to orchard crops, the complex interaction among microwave signal, canopy morphology and soil makes the radar based studies more complicated on woody crops compared to herbaceous crops.

This study may serve as the basis for future regression based biophysical parameter estimation in agricultural studies with increasingly large amounts of freely available remote sensing data. Additionally, even though this paper examines the application of cycleGANs to the field of agriculture, significant contributions of this type of algorithms can be made in any problem that require large amounts of satellite imagery. Specifically, we envision applications of the proposed network in climate change mitigation and adaptation.

Acknowledgments

We are grateful to the editor and the reviewers for their attention to the work and valuable comments. This work was supported by the Space Research and Innovation Network for Technology (SPRINT) under project ID 1243832, and in part by the Research Fund of the Istanbul Technical University. Project Number: MGA-2021-43018.

REFERENCES

[1] S. H. Yueh, R. Shah, M. J. Chaubell, A. Hayashi, X. Xu, and A. Colliander, "A semiempirical modeling of soil moisture, vegetation, and surface roughness impact on CYGNSS reflectometry data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 5800117, pp. 1–17, 2020.

[2] K. A. Semmens, M. C. Anderson, W. P. Kustas, F. Gao, J. G. Alfieri, L. McKee, J. H. Prueger, C. R. Hain, C. Cammalleri, Y. Yang, T. Xia, L. Sanchez, M. M. Alsina, and M. Vélez, "Monitoring daily evapotranspiration over two California vineyards using Landsat-8 in a multi-sensor data fusion approach," *Remote Sensing of Environment*, vol. 185, pp. 155 – 170, 2016.

[3] V. García-Santos, J. Cuxart, M. A. Jiménez, D. Martínez-Villagrasa, G. Simó, R. Picos, and V. Caselles, "Study of temperature heterogeneities at sub-kilometric scales and influence on surface-atmosphere energy interactions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 640–654, Feb 2019.

[4] L. Suárez, P. Zarco-Tejada, V. Gonzélez-Dugo, J. Berni, R. Sagardoy, F. Morales, and E. Fereres, "Detecting water stress effects on fruit quality in orchards with time-series PRI airborne imagery," *Remote Sensing of Environment*, vol. 114, no. 2, pp. 286 – 298, 2010.

[5] Y. Liu, J. Qian, and H. Yue, "Combined Sentinel-1A with Sentinel-2A to Estimate Soil Moisture in Farmland," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1292–1310, 2021.

[6] R. Nasirzadehdizaji, Z. Cakir, F. Balik Sanli, S. Abdikan, A. Pepe, and F. Calò, "Sentinel-1 interferometric coherence and backscattering analysis for crop monitoring," *Computers and Electronics in Agriculture*, vol. 185, p. 106118, 2021.

[7] S. Vanino, P. Nino, C. D. Michele, S. F. Bolognesi, G. DÚrso, C. D. Bene, B. Pennelli, F. Vuolo, R. Farina, G. Pulighe, and R. Napoli, "Capability of Sentinel-2 data for estimating maximum evapotranspiration and irrigation requirements for tomato crop in central Italy," *Remote Sensing of Environment*, vol. 215, pp. 452 – 470, 2018.

[8] E. Erten and C. Rossi, "The worsening impacts of land reclamation assessed with Sentinel-1: The Rize (Turkey) test case," *International Journal of Applied Earth Observation and Geoinformation*, vol. 74, pp. 57 – 64, 2019.

[9] K. S. Rawat, S. K. Singh, and R. K. Pal, "Synergetic methodology for estimation of soil moisture over agricultural area using Landsat-8 and Sentinel-1 satellite data," *Remote Sensing Applications: Society and Environment*, vol. 15, p. 100250, 2019.

[10] A. Amazirh, O. Merlin, S. Er-Raki, Q. Gao, V. Rivalland, Y. Malbeteau, S. Khabba, and M. J. Escorihuela, "Retrieving surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and Landsat thermal data: A study case over bare soil," *Remote Sensing of Environment*, vol. 211, pp. 321 – 337, 2018.

[11] N. N. Baghdadi, M. El Hajj, M. Zribi, and I. Fayad, "Coupling SAR C-band and optical data for soil moisture and leaf area index retrieval over irrigated grasslands," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 3, pp. 1229–1243, March 2016.

[12] D. Palmisano, F. Mattia, A. Balenzano, G. Satalino, N. Pierdicca, and A. V. M. Guarnieri, "Sentinel-1 sensitivity to soil moisture at high incidence angle and the impact on retrieval over seasonal crops," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7308–7321, 2021.

[13] T. Lees, G. Tseng, S. Dadson, A. Hernández, C. G. Atzberger, and S. Reece, "A machine learning pipeline to predict vegetation health," *ICLR workshop on tackling climate change with ML*, 2020.

[14] V. Schmidt, M. A. Muhammed, K. Sankaran, T. Yuan, and Y. Bengio, "Modeling cloud reflectance fields using conditional generative adversarial networks," *ICLR workshop on tackling climate change with ML*, 2020.

[15] C. J. Foley, S. Vaze, M. E. A. Seddiq, A. Unagaev, and N. Efremova, "Smartcast: Predicting soil moisture interpolations into the future using earth observation data in a deep learning framework," *ICLR workshop on tackling climate change with machine learning*, 2020.

[16] A. Mestre-Quereda, J. M. Lopez-Sanchez, F. Vicente-Guijalba, A. W. Jacob, and M. E. Engdahl, "Time-series of Sentinel-1 interferometric coherence and backscatter for crop-type mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4070–4084, 2020.

[17] E. Erten, G. Taşkin, and J. M. Lopez-Sanchez, "Selection of PolSAR observables for crop biophysical variable estimation with global sensitivity analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 766–770, 2019.

[18] M. Wang, J. Wang, and L. Chen, "Mapping paddy rice using weakly supervised long short-term memory network with time series Sentinel optical and SAR images," *Agriculture*, vol. 10, no. 10, p. 483, 2020.

[19] T. Gong, X. Zheng, and X. Lu, "Cross-domain scene classification by integrating multiple incomplete sources," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 035–10 046, 2021.

[20] K. Fang, M. Pan, and C. Shen, "The value of SMAP for long-term soil moisture estimation with the help of deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2221–2233, April 2019.

[21] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, 2017.

[22] J. Song, J. H. Jeong, D. S. Park, H. H. Kim, D. C. Seo, and J. C. Ye, "Unsupervised denoising for satellite imagery using wavelet directional cycleGAN," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, pp. 6823 – 6839, 2020.

[23] A. B. L. Larsen, S. K. Sønderby, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," *CoRR*, vol. abs/1512.09300, 2015. [Online]. Available: http://arxiv.org/abs/1512.09300

[24] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 700–708. [Online]. Available: http://papers.nips.cc/paper/6672-unsupervised-image-to-image-translation-networks.pdf

[25] E. H. Sanchez, M. Serrurier, and M. Ortner, "Learning disentangled representations via mutual information estimation," 2019.

[26] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1917–1929, 2021.

[27] W. He and N. Yokoya, "Multi-temporal Sentinel-1 and -2 data fusion for optical image simulation," *ISPRS International Journal of Geo-Information*, vol. 7, no. 10, 2018. [Online]. Available: https://www.mdpi.com/2220-9964/7/10/389

[28] J. Chen, L. Wang, R. Feng, P. Liu, W. Han, and X. Chen, "CycleGAN-STF: Spatiotemporal fusion via cycleGAN-based image generation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5851 – 5865, 2020.

[29] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017. [Online]. Available: http://arxiv.org/abs/1703.10593

[30] M. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *CoRR*, vol. abs/1703.00848, 2017. [Online]. Available: http://arxiv.org/abs/1703.00848

[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[32] T. Zhou, P. Krähenbühl, M. Aubry, Q. Huang, and A. A. Efros, "Learning dense correspondence via 3d-guided cycle consistency," *CoRR*, vol. abs/1604.05383, 2016. [Online]. Available: http://arxiv.org/abs/1604.05383

[33] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, "A CNN-based fusion method for feature extraction from Sentinel data," *Remote Sensing*, vol. 10, no. 2, 2018.

[34] W. Zhao, Y. Qu, J. Chen, and Z. Yuan, "Deeply synergistic optical and SAR time series for crop dynamic monitoring," *Remote Sensing of Environment*, vol. 247, p. 111952, 2020.

[35] A. Shakya, M. Biswas, and M. Pal, "Cnn-based fusion and classification of SAR and optical data," *International Journal of Remote Sensing*, vol. 41, no. 22, pp. 8839–8861, 2020.

[36] N. Efremova and E. Erten, "Biophysical parameter estimation using Earth observation data in a multi-sensor data fusion approach: Cycle-GAN," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 5965–5968.

[37] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc.,

2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf

[38] S. T. Barratt and R. Sharma, "A note on the inception score," *NeurIPS*, vol. abs/1801.01973, 2018.

[39] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

[40] S. Zhou, M. L. Gordon, R. Krishna, A. Narcomey, D. Morina, and M. S. Bernstein, "HYPE: human eye perceptual evaluation of generative models," *NeurIPS 2019*, vol. abs/1904.01121, 2019. [Online]. Available: http://arxiv.org/abs/1904.01121

[41] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved precision and recall metric for assessing generative models," *NeurIPS 2019 conference*, 2019.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CVPR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[44] D. Dowson and B. Landau, "The fréchet distance between multivariate normal distributions," *Journal of Multivariate Analysis*, vol. 12, no. 3, pp. 450–455, 1982.

[45] K. Crowston, "Amazon mechanical turk: A research tool for organizations and information systems scholars," in *Shaping the Future of ICT Research. Methods and Approaches*, A. Bhattacherjee and B. Fitzgerald, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 210–221.

Natalia Efremova has over 20 years of experience in application of ML methods to real-world problems, both in industry and academia. She received her Ph.D. degree in Computer Science (in Neural Networks for Computer Vision) from Graduate School of Informatics, University of Kyoto, Japan, in 2012. She worked as an Associate Professor in computer science in Plekhanov University of Economics in Russia in 2012-2016. In 2016 - 2018, she obtained her MBA degree from the University of Oxford and she worked as a Teradata Research Fellow in the University of Oxford 2018-2021. Her research focuses on applications of state-of-the-art machine learning (ML) methods and artificial intelligence (AI) for pursuing sustainable goals. This includes defining the areas where ML tools would be especially advantageous within targets related to climate change, sustainable agriculture, water and food scarcity, as well as monitoring progress towards market adoption of the new AI-based technologies for agritech industry. Natalia is currently a lecturer in digital economy in Queen Mary University, London.

Mohammed El Amine Seddik received a Master of Engineering in Data Science in 2017 from Institut Mines-Telecom de Lille, with the final year completed at Telecom Paris Tech and ENS Cachan where he studied Mathematics, Computer Vision and Machine Learning. He received a PhD (entitled: Random Matrix Theory for AI: From Theory to Practice) in signal and image processing from University Paris-Saclay in 2020. He is currently a researcher at Huawei Paris Research Centre. His main research interests are around random matrix theory, random tensor theory, machine learning, deep learning and graphs.

Esra Erten 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), Oberpfaffenhofen, Germany, from April 2008 to June 2010, where she worked on information theory for multi-channel SAR images. From 2010 to 2012, she was with the Chair of Earth Observation and Remote Sensing, Institute of Environmental Engineering, ETH Zurich, Switzerland, where she worked on applied radar remote sensing for environmental parameter estimation. In 2018 she was a visiting professor at Satellite Applications Catapult, U.K. She was supported by Sabbatical Professorship at The Open University School of Engineering, U.K., during 2018-2019 academic period. Currently, she is an associated professor at the department of Geomatics Engineering, Faculty of Civil Engineering, Istanbul Technical University, Turkey. Her 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.

She was the President of the IEEE GRSS Turkish Chapter for the period 2015 - 2019 (she served as vice-president in 2013-2015).