KNOWLEDGE-GUIDED MACHINE LEARNING FOR SPATIOTEMPORAL ENVIRONMENTAL DATA ANALYSIS

A Dissertation

by

HANZI MAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Nicholas Duffield
Committee Members,	Binayak Mohanty
	Dilma Da Silva
	Xia Hu
Head of Department,	Scott Shaefer

August 2020

Major Subject: Computer Science

ABSTRACT

With the emergence of ubiquitous environmental monitoring systems in the past few decades, we are gaining unprecedented ability to collect vast amounts of spatiotemporal environmental data. However, it still remains a challenge to mine information from the spatiotemporal data for many environmental sustainability tasks. This dissertation focuses on utilizing machine learning to analyze spatiotemporal environmental data with the guidance of domain knowledge and specifically proposes several novel algorithms for the soil moisture gap-filling task and crop yield prediction task.

First, we study the problem of soil moisture gap-filling. Large spatiotemporal gaps can be incurred for daily soil moisture product that adopts the radar-radiometer fusion approach. This is normally due to the relatively low revisit schedule and the associated poor spatiotemporal coverage of radar observations. Gap-fill high resolution soil moisture in regional scale for remote sensing soil moisture product is however a great challenge. It requires models learned at neighboring regions to produce predictions at a new region with reasonable accuracy. To address this issue, we propose a novel two-layer machine learning-based algorithm that is motivated by the soil moisture radar-radiometer fusion retrieval algorithm. It predicts the brightness temperature and subsequently the soil moisture at gap areas. Compared with the traditional one-layer machine learning approach, this two-layer approach shows superior performance in extensive experiments at four study areas with distinct climate regimes.

We then focus on information mining from the multi-channel geo-spatiotemporal data. Existing approaches adopt various dimensionality reduction techniques without fully taking advantage of the data. In addition, the lack of labeled training data raises another challenge for modeling such data. We propose a novel semi-supervised self-attentive model

that learns global spatiotemporal representations for prediction tasks. Spatial and temporal variations in the geo-spatiotemporal data are extracted to produce accurate predictions. To overcome the data scarcity issue, we introduce sampled spatial and temporal context that naturally reside in the largely-available unlabeled geo-spatiotemporal data. The proposed algorithm is validated specifically on a large-scale real-world crop yield prediction task. The results show that our semi-supervised self-attentive model outperforms existing state-of-the-art yield prediction methods and its counterpart, the supervised-only self-attentive model, especially under the stress of training data scarcity.

DEDICATION

To my parents, Bintao Mao and Chengnian Xia.

ACKNOWLEDGMENTS

I would like to devote my deepest gratitude to my advisor Dr. Nicholas Duffield. Dr. Duffield took me as his PhD student at my hardest time and has been providing full support to my research work since then. He was the person who introduced me to my research area, applying machine learning in geo-spatiotemporal data analysis, which I have been passionate about and truly believe its potential impact in the future. Starting research work in this area, however, was not easy for me. Not only because I had litter experience in machine learning back in 2016, but also it required extensive work to understand the problems and challenges in this interdisciplinary area itself. Dr. Duffield offered tremendous encouragement and support during those times. He continuously fuelled me with his confidence and passion for our research work. Dr. Duffield always acknowledges and appreciates my research progress, no matter how litter they are. He also shows his best support when I meet obstacles doing research work, with his vast knowledge, critical thinking, and professionalism. I am thankful for his mentorship and support which help me grow into an independent researcher who still feels love in research. I am also grateful that I am lucky enough to find an advisor who not only guides me in research but also is a lifetime role model whom I can always look up to in my future career.

I acknowledge Dr. Binayak Mohanty, Dr. Dilma Da Silva, and Dr. Xia Hu for kindly serving on my committee and patiently commenting on my research. Special thanks go to Dr. Mohanty who guided my soil moisture gap-filling research. Dr. Mohanty provided many valuable and insightful suggestions towards my research with his great kindness and patience. I also would like to thank my other research collaborators, Dhruva Kathuria, Xi Liu, Hao Yuan, and Processor Shuiwang Ji. I appreciate their willingness to join my research work and feel fortunate to have chances to work with these kind and wise people.

I benefited a lot from countless sparkling discussions with them.

I would also like to thank the great mentors I had in my industry internships, Yevgeniy Zagayevskiy at Halliburton, Andrei Curelea, Yun Lin, and Rui Li at Pinterest. They kindly offered tremendous help and support through my internships. The vast industry experiences I learned from them will benefit me in building my future career. My managers, Jiajing Xu and Hari Venkatesan, kindly took me under their wings and helped me enjoy my summer internships at Pinterest. I was also fortunate to meet many nice coworkers at Halliburton and Pinterest. It might need to take pages to name them all and express my gratitude to their help and support, but due to the limited space, I will bear my deep, sincere appreciation to them in mind.

I am also thankful to Dr. Dilma Da Silva and Dr. Hank Walker, who guided me a lot about how to achieve academic excellence, especially in the early stages of my doctorate program. Many thanks to my lab mates, Mahmood Ettehad, Yunhong Xu, and Yixiao Feng, who made my 5-year stay in College Station enjoyable.

Last but not least, I am deeply indebted to my parents, Bintao Mao and Chengnian Xia. They offer tremendous support to my doctorate study and always give me the most freedom to explore the world on my own. Their endless love is always the source of my strength and courage.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Nicholas Duffield of the Department of Electrical and Computer Engineering and the Department of Computer Science and Engineering, Processors Dilma Da Silva and Xia Hu of the Department of Computer Science and Engineering, and Professor Binayak Mohanty of the Department of Biological and Agricultural Engineering.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a research assistantship from Texas A&M University. No outside funding sources were used in this study.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Overview

In the past few decades, vast amounts of spatiotemporal data have been generated from ubiquitous environmental monitoring systems. The emergence of remote sensing technologies coupled with local monitoring workstations enables us the unprecedented ability to monitor the environment in large scale. The spatiotemporal coverage of the geographical data has also been improved remarkably. For example, the Moderate Resolution Imaging Spectroradiometer (MODIS) product tracks changes of the earth surface globally with high temporal resolution [2]. It provides estimates of many environmental factors, such as surface temperature, land cover type, and vegetation growth. The captured environmental heterogeneity over large spatial and temporal extents benefits many regional- and global-scale applications.

Recent years have seen a proliferation of studies that mine information from the geospatiotemporal data utilizing machine learning techniques. Improvements in terms of prediction accuracy have been seen in many computational sustainability tasks, such as land cover classification [3], wildfire prediction [4], precipitation downscaling [5], and climate modelling [6]. Machine learning techniques like random forests [7], support vector machine [8], and deep learning models [9] are adopted to extract complex nonlinear relationships existing among various environmental factors.

However, there are still many challenges remaining in exploiting spatiotemporal data with machine learning techniques. This dissertation specifically contributes to this research area with two domain knowledge-guided machine learning methods. The first is a two-layer machine learning framework motivated by the soil moisture radar-radiometer fusion retrieval algorithm for the challenging gap-filling problem. The second is a semi-

supervised self-attentive model to produce predictions from the multi-channel spatiotemporal data with data scarcity issue. The semi-supervised self-attentive model is validated through a large-scale real-world crop yield prediction task. In the rest of this chapter, we first briefly introduce knowledge-guided machine learning for environmental data analysis with more details being presented in Chapter 2. We then introduce the two proposed knowledge-guided machine learning methods in turn.

1.2 Knowledge-guided Machine Learning for Environmental Data Analysis

Recognizing the success of machine learning in numerous applications like image recognition [5], natural language processing [10], and online recommender systems [11], attentions have been paid to apply machine learning on environmental monitoring data to solve environmental science problems. Domain knowledge regarding the characteristics of collected data and underlying complex earth system processes guide the design of machine learning models in various forms, such as feature engineering, model architecture design, and learning problem formalization. Traditional machine learning approaches like support vector machine [8] and random forest [7] take scalar features as input while ignoring the spatiotemporal structure of the data. Hand-coded features embedded with the domain knowledge of the monitored processes are of great importance to improve the model performances [12]. Deep learning models like convolutional neural networks and recurrent neural networks can capture spatial and temporal information contained in the data and are found to be able to produce state-of-the-art performances in tasks like crop yield prediction [13], land cover detection [3], and oceanic eddy detection [14]. In recent years, there are more and more studies that push forward the progress of this research area with advanced techniques proposed [3, 15]. These techniques are driven by the unique challenges posed by the analysis of environmental data and associated learning tasks. We present two studies in this dissertation in pursuit to contribute to this emerging research

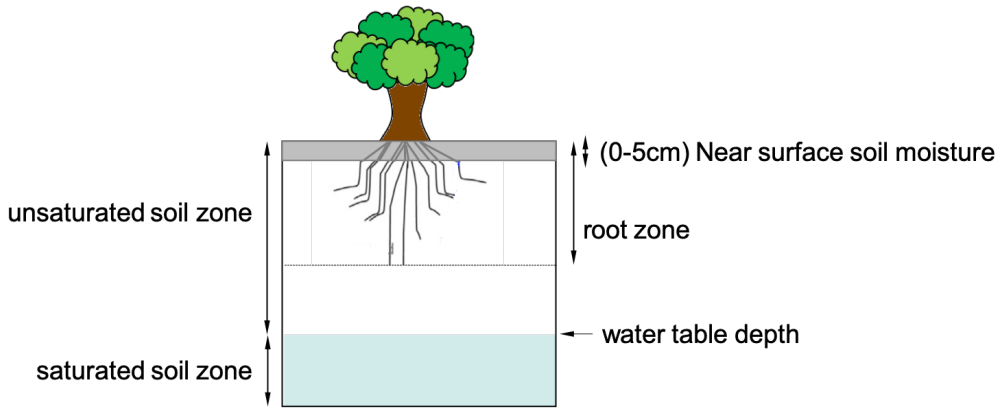

Figure 1.1: Near-surface soil moisture measured by remote sensing satellite (adapted from [1]).

area.

1.3 Two-layer Machine Learning Framework for Gap Filling of Soil Moisture

Soil moisture is one of the key environmental variables that controls the exchange of water and energy between the land surface and the atmosphere. High-resolution soil moisture retrieval with good spatiotemporal coverage benefits many fields, including agricultural and irrigation practices [16], drought and flood predictions [17], and climate modelling [18]. Traditional methods collect point-scale soil moisture estimates relying on either human labor resources or monitoring networks, which are infeasible to scale and are normally sparsely distributed. Remote sensing techniques, on the contrary, enable soil moisture retrieval at a global scale with soil moisture heterogeneity at large spatiotemporal extents captured.

Existing microwave remote sensing techniques are limited to retrieve soil moisture data in the near-surface soil layers (near-surface soil moisture as shown in Figure 1.1). Passive radiometer measures the brightness temperature, which relates to soil moisture status, in the form of land surface emissions, from which highly accurate soil moisture estimates can be generated [19]. The accuracy advantage of the passive radiometer re-

trieved soil moisture data is, however, compromised by its coarse footprint scale ($\sim$40 km). Active radar, on the contrary, measures the backscattered signal after transmitting an active radar signal. While the spatial resolution of its retrieved soil moisture estimates can be much higher than the ones from the passive radiometer, the accuracy of its soil moisture retrievals is not as satisfactory as the prior due to the high sensitivity of radar backscatter signals to surface roughness and vegetation [20].

One of the most promising ways to improve the spatial resolution of soil moisture estimates from the passive radiometer is to combine the observations from both radiometer and radar [21]. This radar-radiometer fusion approach is based on the co-variation of the radar and radiometer signals to the observed ground surface. Specifically, coarse-resolution brightness temperature is first downscaled to high-resolution with the aid of high-resolution radar backscatter estimates, from which high-resolution soil moisture can be produced. On 31 January 2015, the National Aeronautics and Space Administration (NASA) launched the Soil Moisture Active Passive (SMAP) observatory. Its mission was to provide high-resolution (3-km) soil moisture at global scale with 2- to 3-day revisit schedule combining observations from the equipped L-band radiometer and L-band radar.

Unfortunately, the L-band radar carried by the SMAP observatory unexpectedly ceased its operation after three months. To recover its capability to provide high-resolution soil moisture at global scale, C-band Sentinel-1 radars of the Eurapean Space Agency was proposed to replace the original L-band SMAP radar [22]. The SMAP/Sentinel-1 3-km soil moisture product we focus on is just the product from this work. One restriction of the SMAP/Sentinel-1 3-km soil moisture product is that it has large spatiotemporal gaps incurred by the narrow swath and low revisit schedule of the Sentinel-1 radars. This severely restrict the applications of the resulting soil moisture product for many regional- and global-scale hydro-climatological studies [23].

Existing machine learning researches fill the gap of high-resolution soil moisture by learning linear or nonlinear models that capture the relationship between coarse-resolution and fine-resolution soil moisture [24, 25, 26]. Auxiliary environmental factors, such as land surface temperature, soil properties, and precipitation, are included in the models as they have been studied as major drivers of soil moisture variations [27, 28]. One shortcoming of this approach is that additional uncertainty is incurred when introducing coarse-resolution soil moisture as the main input feature. The radar-radiometer fusion retrieval algorithm downscales the coarse-resolution brightness temperature to fine-scale first, then converts the fine-scale brightness temperature to soil moisture. Coarse-resolution soil moisture, as a by-product of this process, is converted from coarse-resolution brightness temperature and carries extra noise. Moreover, the gap filling task itself poses extra challenges to the machine learning models as it requires models learned at neighboring regions to perform well at a new region with similar hydroclimate conditions. Traditional machine learning methods sometimes cannot produce predictions with satisfactory accuracy under this setting. Additionally, existing machine learning models normally rely on fine-resolution soil moisture collected from either catchment-scale campaigns or sparsely distributed monitoring networks. This restricts the size and spatial coverage of available training data, and thus the applicability of the produced machine learning models.

To address these issues, we propose a two-layer machine learning-based framework. Different from existing approaches which directly downscales soil moisture from coarse-resolution to fine-resolution, we first downscale coarse-resolution brightness temperature to fine-resolution, then predict fine-resolution soil moisture using two consecutive machine learning models. This avoids the dependence on the coarse-scale soil moisture which carries extra uncertainties. We decide the input features of the two machine learning models with the guidance of the original radar-radiometer retrieval algorithm.

The average of historical radar measurements are introduced, at the first time, as input features and are found to be useful in terms of boosting the predictive performance of the machine learning models. To produce predictions that are consistent with available SMAP/Sentinel-1 soil moisture product, we use estimates from the SMAP/Sentinel-1 product as target variables. The expanded training data also enable us to conduct experiments at regional scale and produce machine learning models which adapt to large spatial extents.

Another intuition we have here is that the feature input designed based on the radar-radiometer fusion retrieval algorithm might be more consistent across different regions in terms of both space and time, compared with the feature set of environmental factors used in previous studies. The SMAP/Sentinel active-passive retrieval algorithm [22], which is introduced in detail in Chapter 3, is designed to fuse the radar and radiometer observations over vegetated soil for soil moisture estimates by isolating the vegetation contribution. Feature set based on the retrieval algorithm is assumed to have less spatial and temporal heterogeneity, from which machine learning models with better generalization ability can be trained.

We conduct extensive experiments to validate the performance of the proposed two-layer machine learning framework using 3.5-year data from 1 April 2015 to 30 September 2018. Multiple regions with disparate climate regimes and land surface heterogeneities across the Contiguous United States are chosen to conduct the experiments. It is shown that the proposed two-layer machine learning approach consistently outperforms traditional machine learning method with great margin. It also shows satisfactory results for the comparisons of the predictions from the two-layer framework with in situ measurements from monitoring networks, as well as with airborne soil moisture campaign data.

1.4 Semi-supervised Self-attentive Model for Crop Yield Prediction

For the previous soil moisture gap-filling task, scalar feature are taken as input to the regression model. The next topic we focus on is to extract necessary information from higher-dimensional environmental data, the multi-channel geo-spatiotemporal data, to improve the prediction performance of computational sustainability tasks. Information mining from multi-channel geo-spatiotemporal data, however, poses great challenges to many computational sustainability applications, and machine learning models themselves. Traditional machine learning models like support vector machine and random forest are not good at extracting spatial and/or temporal variations from the spatiotemporal data. On the contrary, deep learning models, such as convolutional neural networks and recurrent neural networks, have the ability to extract spatial/temporal variations from the data, but require more training samples with the increased model complexities.

We propose a novel semi-supervised self-attentive model that learns global spatiotemporal representations for prediction tasks. Spatial and temporal variations in the geo-spatiotemporal data are extracted to produce accurate predictions. To overcome the data scarcity issue, we introduce sampled spatial and temporal context that naturally reside in the largely-available unlabeled geo-spatiotemporal data. The proposed model is mainly based on the following observations that can generally be applied to many computational sustainability tasks. First, spatially explicit heterogeneity and interactions among environmental factors contain important information regarding the characteristics of environmental factors at larger spatial scales [29, 30]. Spatially averaging the input environmental factors would incur unnecessary information loss. Second, temporal variations of the environmental factors also carry valuable information that can aid the predictions [3, 30]. Third, the spatial variations and interactions among the environmental factors often show distinct but relatively consistent temporal patterns [30]. Machine learning models which

enable knowledge transfer among different time steps can aid the generation of more accurate predictions. Forth, many computational sustainability tasks suffer from the lack of ground-truth labels [15]. This is an especially crucial issue for deep learning models that deal with high-dimensional data. Fifth, the spatial and temporal context signals contained in the largely-available unlabeled geo-spatiotemporal data haven't been taken full advantage of to aid the learning of machine learning models.

To evaluate the performance of the proposed model, especially under the stress of data scarcity, we conduct extensive experiments on a large-scale real-world crop yield prediction task. Accurate large-scale crop yield estimation is of great importance to food security, grain policy making, and market management [31, 32]. With the increasing food demand due to population growth [33], accurate and timely crop yield estimation becomes ever more imperative. Traditional crop yield estimation methods that rely on manual survey or simulation-based crop models are cost-inefficient and limited to local scales [34].

In recent years, remote sensing has drawn great attention for crop yield estimation due to its large-scale spatial coverage and timely monitoring of crop progress. Specifically, multi-spectral satellite images and their derived vegetation indices are highly sensitive to aboveground biomass [35] and have been frequently used in crop yield estimation [13, 36]. In addition to remote sensing data that directly monitors cropland growth status, there are various environmental factors that indirectly impact the end-of-season crop yields. Meteorological factors such as precipitation and temperature are shown to have significant effects on crop yield, especially in its late grain-filling period [35]. Geographically localized, time-invariant factors such as soil properties and elevation also influence how different climate conditions impact crop growth [37]. More importantly, it varies by growing phases of the crop sensitivities toward these environmental variables [30].

There are however, very limited studies that integrate remote sensing and climate data

in spatiotemporal space to capture the cumulative impact of climate dynamics on crop progress. Existing crop yield prediction approaches rarely handle the fusion of heterogeneous geodata in spatiotemporal space directly. Most approaches take spatially averaged time-series variables [34, 30, 38] while ignoring the spatial distributions of crop growth and environmental events. In a more recent research [13], variables with spatial distributions are first converted to histograms, with the assumption of permutation invariance, before feeding to CNN or LSTM. Their assumption is less appropriate for the fusion of crop growth variables and environmental information as environmental factors like precipitation and temperature usually have spatially explicit and temporally varying impacts on crop growth. Unmanned Aerial Vehicles (UAVs) can acquire spatially distributed multispectral and thermal images in high resolution. However, existing researches [39, 40, 41] only focus on collecting UAV data at fixed time steps during the growing season and are limited to local scales.

The novel semi-supervised self-attentive-based CNN model we proposed *jointly* learns the crop yield prediction function and embedding function for crop status representation. Crop status representations at each time step are first extracted from heterogeneous geospatial data through convolutional neural networks. A self-attention layer is then applied to temporally aggregate crop representation at all time steps to generate the final yield predictions. To aid meaningful representation learning from the scarce but high-dimensional data, we further introduce the spatial and temporal coherence signals that naturally reside in the data. Similar to the proximity-based word embedding models in natural language, we make an assumption here that images of crop status that are spatially/temporally close should have similar representation, and vice versa.

We evaluate our approach through the challenging county-level soybean yield prediction task. The results show that our semi-supervised self-attentive model outperforms existing state-of-the-art yield prediction methods and its counterpart, the supervised-only

self-attentive model, especially under the stress of training data scarcity. Additionally, we examine the generalization ability of the proposed model across different regions. This sheds light on how the model trained using input data from one location would perform for prediction tasks of another location. Extensive feature importance analysis is also conducted for both early in-season prediction and right-before harvesting prediction.

The rest of this dissertation is organized as follows. We present in Chapter 2 a survey of machine learning for environmental data analysis. Chapter 3 presents the two-layer machine learning algorithm for the soil moisture gap-filling problem. Chapter 4 focuses on the semi-supervised self-attentive model that produces crop yield predictions from multi-channel geo-spatiotemporal data. We finish this dissertation with Chapter 5 to summarize the dissertation and discuss future work.

2. SURVEY OF MACHINE LEARNING FOR ENVIRONMENTAL DATA ANALYSIS

The past few decades see unprecedented progresses in both machine learning and data produced by environmental monitoring systems. While there are promising opportunities in environmental sustainability of applying machine learning to analyze the environmental data, unique challenges are posed by the characteristics of the environmental data and the underlying earth system processes. In recent years, a proliferation of studies have been done to address some of the challenges in pursuit of advancing both machine learning from the computer science discipline and many other disciplines such as geophysics, hydrology, and agricultural economics. We provide in this chapter a brief survey of the progress of machine learning for environmental data analysis across various disciplines, which also serves as a background for the rest of the thesis.

2.1 Sources of Environmental Data

Human beings have long been monitoring the earth and its interacting components like biosphere and atmosphere through in-situ sensors/workstations. These in-situ sensor-based observations are either available for a small local region or scattered non-uniformly and sparsely across a large spatial region. For example, there are more than 10,000 precipitation gauges scattered through the United States which provide daily precipitation measurements [42]. Observations provided by the in-situ sensors are often perceived as the most reliable and accurate estimates of many environmental factors and events. Their availability to large spatial/temporal extents are however limited by the nature of their collection processes.

Satellite-based remote sensing techniques, on the contrary, can provide regional or global estimates of an environmental factor, such as surface temperature and spectral re-

flectance, at the remote sensing footprint scale. The most common remote sensing data utilized by the machine learning field is the multispectral images of the earth surface [13, 15] from satellites like Moderate Resolution Imaging Spectroradiometer (MODIS) [2] and Landsat [43]. The visual spectral bands (red, green, and blue) of the multispectral images can reconstruct the visual view of the earth surface and benefit computer vision tasks like urban building classification [44] and economic tasks like poverty mapping from nighttime light intensity [45, 46]. There are also numerous environmental/geoscience variables like soil moisture [22], vegetation indices [47] can be derived from the multispectral images which consecutively benefits a wide range of sustainability disciplines. Besides the satellite-based remote sensing techniques, Unmanned Aerial Vehicles (UAVs) or airborne radar and radiometer can also provide observations of particular environmental factors. These techniques are often restricted to cover a local region at irregular temporal intervals [48, 40].

Another source of environmental data is the simulation data from physics-based models. Large amounts of simulation data can be generated from the models that capture the complex environmental geoscience processes based on physical laws and principles. For example, near-surface climate variables, such as precipitation, temperature, wind speed, and humidity, from the North-American Land Data Assimilation System Project Phase 2 (NLDAS-2) are used to spatially downscale coarse-resolution precipitation through the random forest model [5]. While the simulation data has the advantage over other data sources that continuous coverage at prefixed spatial and temporal resolutions can often be guaranteed, the accuracy of the simulated data are restricted by the extent to which the underlying complex geoscience process being captured by the simulation models.

In addition to the aforementioned data sources, there are also work that fuse environmental data from various sources in pursuit to improve the data quality or spatial/temporal coverage. The Parameter-elevation Relationships on Independent Slopes Model (PRISM)

dataset, for example, fuse climate observations from a wide range of monitoring sensors and radars to provide estimates of climate variables with continuous spatial coverage, such as precipitation, temperature and vapor pressure deficit, using climate-based interpolation [49]. The improved spatial coverage of the climate data benefit many tasks like crop yield prediction [30], climate modeling and change tracking [50].

2.2 Machine Learning for Environmental Data Analysis

Recognizing the success of machine learning in numerous applications like recommendation system, natural language processing, and image recognition, in the past two decades researchers in computer science and various environmental science fields have been applying machine learning to analyze environmental data to help solve environmental science challenges facing humanity [51]. Domain knowledge from the environmental science is embedded in machine learning models in the forms of feature engineering, model architecture design, and learning problem formalization. This benefits not only the specific disciplines which the learning tasks fall in but also the machine learning field itself.

While this is still an emerging research area with new techniques keeping coming out, we categorize here the work of machine learning for environmental data analysis into three kinds, traditional machine learning approaches relying on hand-coded features, deep learning models that handle the spatiotemporal structure of the data and automatically extract relevant features from the data, and more advanced techniques that attempt to address unique challenges posed by the environmental data. The rest of this chapter will present the three kinds of work in turn.

2.2.1 Traditional Machine Learning Approaches

Traditional machine learning models like linear models [52], support vector machine [8], and random forest [7] take scalar features as input, and extract either linear or non-

linear relationship between the input features and the target variable. Hand-coded feature engineering is often heavily relied here for two main reasons. First, traditional machine learning models are not able to automatically extract information regarding the spatial and temporal variations contained in the raw spatiotemporal environmental data [12]. Hand-coded features embedded with the domain knowledge of the underlying complex earth system processes are of great importance to improve model performances. Second, certain extent of explainability or interpretability are often expected from these traditional machine learning models. Utilizing hand-coded features, while might not be flexible enough, helps communications with domain experts and benefits decision making process [53].

Among numerous traditional machine learning models that capture complex nonlinear relationships between input features and output variable, random forest model is one of the most widely used and successful models applied to analyze environmental data. Random forest model gains its reputation for being less prone to the overfitting problem and have better generalization ability compared with other traditional machine learning approaches [54, 55]. For example, researchers in [56] utilize random forest model to classify land cover types for a mountainous area in Colorado. Multispectral data from the Landsat satellite and topographic features like elevation, slope, and aspect are taken as input considering that the experiment area is mountainous. Additional experiments are done to examine the feature importance of predicting each land cover type.

2.2.2 Deep Learning

The spatiotemporal structure of the environmental data brings both challenges and opportunities. On one hand, traditional machine learning approaches are not able to handle the spatiotemporal data in its raw form. Dimensionality reduction techniques like spatial averaging have to be adopted for higher-resolution spatial images when they are taken

as input for a unit-wise prediction task. Time-series scalar features are fed into machine learning models like random forest without differentiating the temporal order among the features. Since almost all environmental science phenomena happens in both spatial and temporal dimensions [12], ignoring such variations in the data would hamper the performance of the learning models. On the other hand, the spatial and temporal variations in the environmental data contain much useful information, which can be utilized to aid model learning [30]. There are also natural spatiotemporal correlation/context signals existing in the data [15]. For example, a location covered by one type of land cover is often surrounded by locations with similar land cover types [3]. These domain knowledge-guided observations can be utilized to boost the performance of machine learning models.

We present here three classic kinds of deep learning models which can handle the spatial and/or temporal structure of the environmental data, the deep residual neural networks [5] for learning of spatial images, the Long Short Term Memory-Attention (LSTM-Attention) networks [57] for learning of time-series data, and more advanced deep learning models that handle spatiotemporal data, such as convolutional LSTM network [58] and 3d convolutional neural networks (C3D) [59, 60, 61].

2.2.2.1 CNNs and Deep Residual Neural Networks

Convolutional neural networks have demonstrated its superior ability to extract spatial information from the data and have been applied successfully to many vision-oriented tasks [62]. Its achievement is established mainly due to the convolutional layer which applies the convolution operator across the first two dimensions (height and width) of the input images. The shared weights of the convolutional filter greatly reduces the number of model parameters. Deep convolutional neural networks are constituted of stacked layers of convolutional neural networks and can integrate low/middle/high- level features and produce powerful classifiers/regressors.

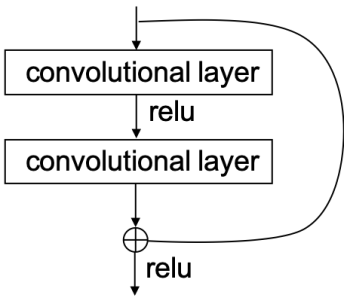

Figure 2.1: Building block for residual neural network.

Training deep convolutional neural networks is however not easy. The degradation problem is noticed by researchers in [5] that the prediction accuracy gets saturated, and then degrades rapidly when increasing the network depth. It is further argued that the degradation problem happens as learning models might have difficulties in approximating identity mappings when many nonlinear layers get stacked. Instead of approximate complicated functions, e.g. $f(\mathbf{x})$, using neural networks, it is hypothesized that approximate the residual function, e.g. $f(\mathbf{x}) - \mathbf{x}$, can ease the learning. Reformulation of the problem is used here to help precondition the problem. Figure 2.1 shows the basic residual block. The input of the whole building block gets added to the output of two convolutional layers. The intuition behind this is that if the underlying function to be learned is closer to an identity mapping than to a zero mapping, it should be easier for the model equipped with such residual building block to learn a function closer the the identity mapping, than to learn a new function from the beginning. It is then proposed in [5] different versions of residual neural networks with different depths. The ResNet-18, for example, has eight residual building blocks stacked together and eighteen convolutional layers in total.

Deep residual convolutional neural networks can be applied to extract spatial patterns contained in the environmental data. In [15], ResNet-18 is applied to static multispectral images from the USDA's National Agriculture Imagery Program (NAIP) and Landsat 8 satellite to produce predictions for tasks such as land cover classification, poverty predic-

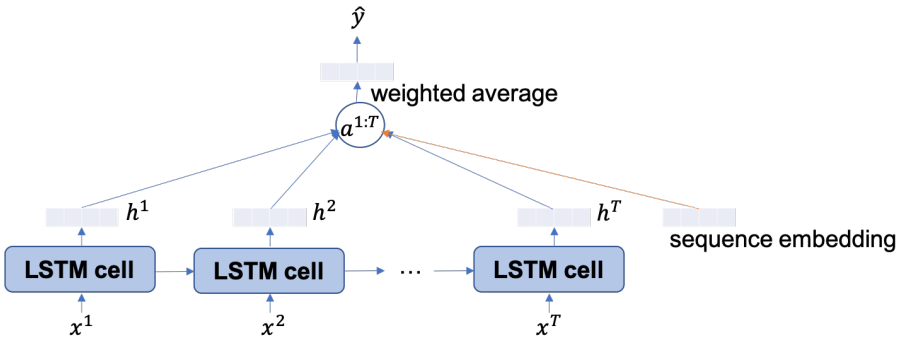

Figure 2.2: The architecture of LSTM-Attention networks.

tion, and country health index prediction.

2.2.2.2 LSTM-Attention Networks

Instead of handling the spatially distributed data, the Long-Short Term Memory (LSTM) networks have been utilized to mine temporal patterns from time-series environmental data. In [63], LSTM networks are trained to expand satellite-provided soil moisture observations to gap areas in pursuit of seamless coverage of soil moisture in large scale. Model-simulated moisture, atmospheric forcings like precipitation, temperature, humdity, and static physiographic attributes are taken as inputs. LSTM is found to be able to learn temporal patterns of soil moisture dynamics where physics-based simulation models fail.

A domain knowledge-guided observation is made for the land cover detection task in [57] that there exists a discrimination period in the sequence of input data. Successful detection of such discrimination period is required to produce satisfactory results for the land cover classification task. For example, croplands can be more easily identified/classified during the growing season compared to the early planting stage or late harvesting stage. The attention mechanism [64] is added to LSTM networks to aid the discovery of the discriminative period as shown in Figure 2.2. Specifically, a sequence embedding is learned

separately to represent the entire sequence data. After the generation of hidden representations h^t at each time step t, relevance scores of each time step are calculated based on the similarity between a hidden representation and the sequence embedding. Finally, all hidden representations can be aggregated into one unifying embedding through weighted average based on the relevance scores, from which the classification prediction can be produced.

2.2.2.3 Convolutional LSTM and 3D CNNs

While there are many deep learning models that attract the attention of environmental scientists after achieving their success in computer science discipline, convolutional LSTM is one of the few standard deep learning models whose design was driven by the unique characteristics of environmental data and the precipitation nowcasting task [58]. The precipitation nowcasting task is formulated as as a sequence prediction problem, of which both the input and output are spatiotemporal sequences. Different from traditional LSTM-based encoder-decoder framework [65], the convolutional LSTM networks proposed in [58] keep the convolutional structures for both the input-to-state and state-to-state transitions. All the inputs, hidden states, gates, and cell outputs of the new LSTM cell (noted as ConvLSTM) are in the forms of 3D tensors instead of 2D ones. The convolutional LSTM is assumed to be able to better preserve the spatial information, in addition to the temporal information, of the input data. It is further validated in [58] that the convolutional LSTM networks outperform the state-of-the-art operational algorithm for precipitation nowcasting.

3-dimensional convolutional networks (3D CNNs) proposed in [59] instead designs a novel 3D architecture for the convolutional neural networks in order to handle spatiotemporal data. Observing that the traditional 2D convolutional neural networks can only apply 2D convolution operator across two dimensions, i.e. height and width of the input,

the 3D convolution is designed to preserve the temporal information of the input signal. Researchers in [66] apply the 3D CNNs to conduct crop yield prediction from satellite images and find that the additional temporal information captured by the 3D CNNs help the model achieve better performance than traditional 2D CNNs.

2.2.3 Advanced Techniques Addressing Unique Challenges

In addition to the aforementioned traditional machine learning approaches and classic deep learning models, in recent years, there are more and more interesting work that adopt advanced techniques to address unique challenges posed by the environmental data and associated learning tasks. Observations based on domain-knowledge are infused into the model architecture design and problem formulation, which greatly improve the produced prediction accuracy than standard deep learning models.

For example, it is observed in [3] that both long- and short-term temporal variation patterns exist in the vegetation growth features for the land cover prediction task. Traditional LSTM fails to capture both long-term seasonal changes and short-term recent environmental changes, which are vital for accurate land cover forecasting. A dual-memory LSTM is thus designed where two sets of cell states with different time windows are kept. This customized design based on domain knowledge achieves better performance than standard LSTM model.

While proximity-based word embedding models have been widely used in natural language processing area [67, 68], it is until recently that a similar spatial context idea is applied to spatially distributed data to learn unsupervised representations [15]. Image tiles here are treated as atomic units with the first law of geography that "everything is related to everything else, but near things are more related than distant things" applied. A triplet loss function is designed to minimize the distance, which is measured as the Euclidean distance of representation embedding vectors, between image tiles that are close in space,

while maximize the distance between image tiles that are far away. It is shown in [15] that the learned unsupervised representations of multispectral images can improve the predictions of many downstream tasks, especially when the labeled data are scarce.

2.3 Conclusion

Geo-spatiotemporal data have their unique characteristics due to the nature of the underlying complex geoscience processes and the collection process of the data, which poses challenges to machine learning techniques that are applied to analyze such data. Addressing these challenges will not only benefit the specific discipline the tasks fall in but also the machine learning field itself. As stated in the beginning of this chapter, applying domain knowledge-guided machine learning to analyze environmental data is still an emerging research topic. Joint efforts of computer scientists and environmental scientists are needed to push forward progress in this area.

3. GAP FILLING OF HIGH-RESOLUTION SOIL MOISTURE FOR SMAP/SENTINEL-1: A TWO-LAYER MACHINE LEARNING-BASED FRAMEWORK*

3.1 Introduction

Soil moisture is an essential variable controlling the exchange of water and energy fluxes between the land surface and the atmosphere. Retrieving high-resolution soil moisture at local, regional and global scales is critical for best agricultural and irrigation practices [16], weather and climate predictions [18, 69], biogeochemical process characterizations [70] and drought and flood predictions [17, 71]. To complement the sparsely distributed in situ soil moisture measurements [72, 73, 74], various microwave remote sensing approaches have been developed over the last few decades to retrieve soil moisture estimates at a global scale such as the Advanced Scatterometer [75], the Advanced Microwave Scanning Radiometer-EOS [76], the Soil Moisture and Ocean Salinity [77], the Soil Moisture Active Passive (SMAP) mission [78], and the Sentinel-1 mission [79]. Global-scale microwave sensing soil moisture products have the advantage of capturing soil moisture heterogeneity over large spatial extents. This further benefits regional- and global-scale applications that are impacted by land-atmosphere interactions and/or soil moisture variability [80, 81, 82].

Existing microwave remote sensing techniques measure the near-surface soil moisture (≤ 5 cm) indirectly through its relationship with water content and soil dielectric constant. They can be classified into two categories according to the source of their signals, 1) the active radar measuring the backscattered signal after transmitting an active pulse and 2)

*Reprinted with permission from "Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework" by H. Mao, D. Kathuria, N. Duffield, and B. P. Mohanty, 2019, Water Resources Research, vol. 55, no. 8, pp. 6986-7009, doi:10.1029/2019WR024902. Copyright 2019 by John Wiley and Sons, Inc.

the passive radiometer measuring the land surface emissions. Active radars typically retrieve soil moisture at a spatial resolution varying from several meters to $\sim$3 km. In addition to soil moisture, they are also highly sensitive to surface roughness and vegetation, which reduces the accuracy of the retrieved soil moisture estimates [20, 83]. On the contrary, the relatively mature passive radiometers have the ability to provide better soil moisture estimates due to their high sensitivity to soil moisture [19]. The high accuracy of the soil moisture estimates retrieved by the passive remote sensing is, however, compromised by its coarse resolution, $\sim$40 km for both the Soil Moisture and Ocean Salinity and SMAP passive soil moisture products [77, 84], which restricts the application of the resulting soil moisture products for regional hydro-climatological studies [23, 78, 85, 86].

Among the various methods proposed to downscale coarse-scale soil moisture [87, 88], the fusion of radar and radiometer microwave observations is one of the most promising and robust techniques to provide high-resolution (1-10 km) soil moisture estimates at a global scale. The SMAP mission, specifically, adopts the active-passive retrieval algorithm proposed by [21]. In this two-layer retrieval algorithm, the radar backscatter measurements are first used to downscale the coarse-scale passive brightness temperature measurements from which the high-resolution soil moisture is then retrieved through the single channel algorithm (SCA) [89]. The original SMAP observatory was launched on 31 January 2015. Equipped with an L-band radiometer and an L-band radar as a single observation system, its mission was to combine 36-km brightness temperature and 3-km radar backscatter to provide 9-km soil moisture at a global scale [78, 90].

Unfortunately, the SMAP L-band radar unexpectedly ceased its operation on 7 July 2015. To recover the capability of providing high-resolution soil moisture, one possible solution was to replace the SMAP L-band radar measurements with the C-band radar measurements from the European Space Agency's Sentinel-1 radars, Sentinel-1A and Sentinel-1B [91, 92]. Many studies have investigated the performance of integrating the

C-band active and the L-band passive observations to retrieve high resolution soil moisture with promising results [91, 93, 94, 95].

As the most recent 3-km soil moisture product from the SMAP mission, the SMAP/Sentinel-1 product L2_SM_SP [96] combines the 33-km brightness temperature from the SMAP-enhanced passive soil moisture product L3_SM_P_E with the 3-km aggregated copolarization (co-pol) and cross-polarization (cross-pol) backscatter measurements retrieved by the Sentinel-1 radars. They adopt the SMAP/Sentinel active-passive retrieval algorithm [22] similar to the original SMAP active-passive algorithm where high-resolution brightness temperature and soil moisture are retrieved consecutively. One restriction of this product, however, is that the Sentinel-1 radar observations have much narrower swath ($\sim$250 km) compared with the relatively wide swath (1,000 km) of the SMAP radiometer. As the SMAP/Sentinel active-passive retrieval algorithm requires concurrent radar and radiometer observations, the narrow swath of the Sentinel-1 radar significantly reduces the spatial coverage of the resulting daily soil moisture compared with the SMAP-enhanced passive 33-km soil moisture product (Figure 3.1). The temporal resolution also drops from a 3-day revisit schedule to a 6-day one when using the collocated Sentinel-1A and Sentinel-1B radar backscatter observations [78, 91, 97]. Additionally, considering that the Sentinel-1A and Sentinel-1B sensors provide backscatter measurements with different swath patterns and interleaved revisit schedules, the temporal resolution further drops to 12 days or more for some locations. The narrow swath and low revisit schedule of the SMAP/Sentinel-1 3-km soil moisture cannot achieve the required temporal resolution for many soil moisture-related applications [88, 98].

This chapter thus focuses on downscaling/gap-filling soil moisture at areas where the SMAP/Sentinel active-passive retrieval algorithm cannot be applied due to missing radar backscatter measurements. [87] and [88] outline the various techniques to downscale passive remote sensing soil moisture. The techniques that do not rely on concurrent radar

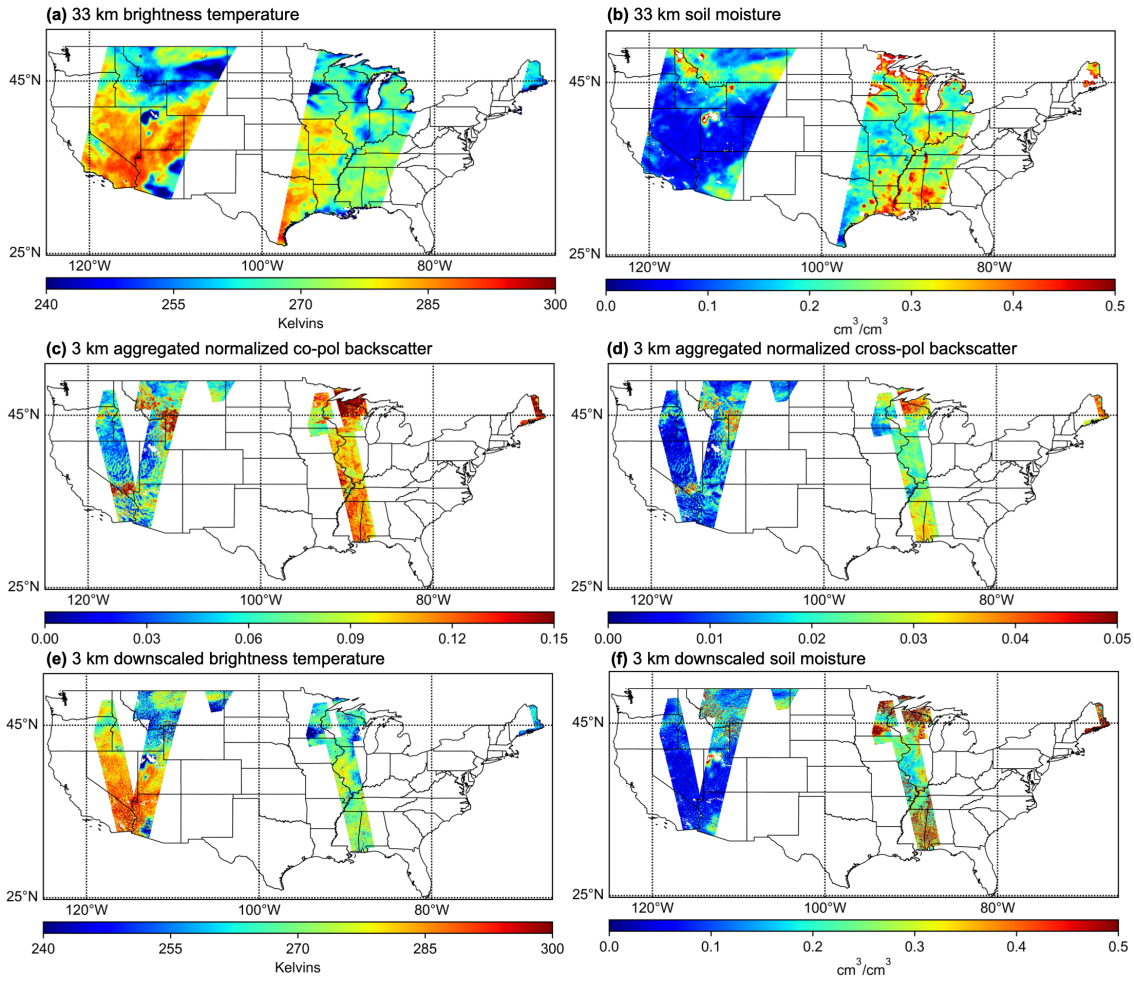

Figure 3.1: Spatial availability of the SMAP/Sentinel-1 active-passive 3-km soil moisture product (L2_SM_SP) and the SMAP-enhanced passive 33-km soil moisture product (L3_SM_P_E) on 16 June 2018 for the Continental United States. (a) The 33-km brightness temperature retrieved by the SMAP radiometer. (b) The 33-km soil moisture converted from (a) using the single channel algorithm. (c) The 3-km aggregated normalized copolarization backscatter retrieved by the Sentinel-1A/Sentinel-1B radars. (d) The 3-km aggregated normalized cross-polarization backscatter retrieved by the Sentinel-1A/Sentinel-1B radars. (e) The 3-km downscaled brightness temperature from the SMAP/Sentinel active-passive algorithm. (f) The 3-km downscaled soil moisture converted from (e) using the single channel algorithm. SMAP = Soil Moisture Active Passive.

observations can be applied to gap fill the SMAP/Sentinel-1 3-km soil moisture product, such as fusion of optical/thermal and microwave methods [99, 100, 101], data assimila-

tion methods [102, 103], and machine learning methods [26, 24, 104, 105, 25]. Among these approaches, machine learning techniques have shown great potential to downscale soil moisture with its ability to handle noisy data and learning complex nonlinear processes [106, 5]. Existing machine learning methods capture the relationship between coarse-scale and fine-scale soil moisture directly with the aid of auxiliary environmental variables, such as soil properties, topography, land cover, and meteorological forcings. The fine-scale soil moisture estimates used to train/calibrate the machine learning models are restricted to ground data obtained from either catchment-scale campaigns [24, 25] or sparsely distributed soil moisture networks [26].

In this chapter, estimates from the SMAP/Sentinel-1 active-passive 3-km soil moisture product serve as target variables to train the machine learning models and learn complex relationships existing between coarse-scale and fine-scale soil moisture [107]. Directly using remote sensing soil moisture estimates for training a machine learning model also has the advantage of maintaining high consistency, or "high fidelity" [63], between the downscaled predictions and the SMAP/Sentinel-1 3-km soil moisture product. We propose a two-layer machine learning-based framework motivated by the SMAP/Sentinel active-passive retrieval algorithm to improve the consistency between the downscaled soil moisture and the retrieved 3-km soil moisture from the SMAP/Sentinel-1 soil moisture product. The downscaled brightness temperature and soil moisture are predicted consecutively using machine learning models in the two layers of the framework. Instead of directly downscaling the 33-km soil moisture from the SMAP-enhanced passive soil moisture product to 3 km, the 33 km brightness temperature (from the same product) is first downscaled to 3 km through a machine learning model. A separate machine learning model is then applied to convert the 3-km downscaled brightness temperature to 3-km soil moisture. This machine learning approach, different from the SMAP/Sentinel active-passive retrieval algorithm, removes the dependence on concurrent radar and ra-

diometer observations and thus can be applied to gap fill soil moisture at areas where the SMAP/Sentinel active-passive retrieval algorithm cannot be used. Compared with the conventional one-layer machine learning framework that directly downscales soil moisture from 33 km to 3 km, the two-layer framework avoids the dependence on the 33-km soil moisture that has additional uncertainties incurred by the indirect conversion from the 33-km brightness temperature. Moreover, it allows us to naturally introduce the average of historical (past 30 days) radar backscatter measurements as a key input to downscale the 33-km brightness temperature. This is critical as the vegetation and surface heterogeneity captured by the average of historical radar backscatter measurements can help produce more accurate downscaled brightness temperature, from which high resolution soil moisture with higher accuracy can be subsequently derived.

Using the proposed framework, downscaled soil moisture that is highly consistent with the SMAP/Sentinel-1 3-km soil moisture product can be reconstructed at gap areas. The framework is a general one though and can embed any machine learning model. We implement both a general one-layer framework that directly downscales 33-km soil moisture to 3 km and our proposed two-layer framework with the Random Forest [7] serving as the predictive machine learning model. The Random Forest model is specifically chosen here for its ability to handle complex nonlinear relationships with reduced overfitting. It has also shown promising results in many soil moisture downscaling studies [26, 104, 108, 109].

In this chapter, relationships between the SMAP/Sentinel-1 3-km soil moisture/brightness temperature and the input features (ancillary data) are learned through machine learning models at regions where the 3-km soil moisture/brightness temperature is available. The learned models are then applied to predict 3-km soil moisture/brightness temperature at spatially/temporally neighboring gap regions where the Sentinel radar observations and thus the 3-km soil moisture and brightness temperature values are missing.

We validate the performance of the models in disparate regions across Contiguous United States from 1 April 2015 to 30 September 2018 by predicting holdout SMAP/Sentinel-1 3-km soil moisture data. The efficacy of the proposed framework is also demonstrated by comparing it with in situ measurements from soil moisture networks as well as with airborne soil moisture data retrieved during the SMAPVEx 2016 campaign in Iowa.

The contribution of this chapter is threefold. First, it is the first study using machine learning to address the poor spatial and temporal coverage for downscaled soil moisture estimates retrieved by the radar-radiometer combination approach. Second, it designs a novel two-layer machine learning framework motivated by the original SMAP/Sentinel active-passive retrieval algorithm to improve 3-km soil moisture predictions. Third, the 30-day average of historical radar backscatter measurements is introduced for the first time as a key input to aid in capturing subgrid soil moisture heterogeneity. The rest of the chapter is organized as follows. Section 3.2 describes the data sets and study area. Section 3.3 gives a brief introduction of the SMAP/Sentinel active-passive retrieval algorithm. The specifics of the methodology are presented in section 3.4. The results and validations are discussed in section 3.5. The conclusions and discussion for future work are then provided in section 3.6.

3.2 Data Sets and Study Area

3.2.1 Soil Moisture and Brightness Temperature

The most recent SMAP/Sentinel-1 active-passive soil moisture product (L2_SM_SP) provides 3-km soil moisture and 3-km brightness temperature that can serve as the learning targets of our machine learning models. As a Level 2 product, for any specific day, there are hundreds of data files each containing a subpatch of the globally covered data. We generate daily composites from this Level 2 product with the averaged value applied to overlapping estimates. Besides the 3-km brightness temperature and soil moisture,

3-km radar backscatter cross-section measurements and other physically based parameters are available in L2_SM_SP from which averages of historical (past 30 days) values can be generated for both covered and gap regions. We take the 33-km soil moisture and 33-km brightness temperature (values are posted on 9-km resolution grids) from the SMAP-enhanced passive soil moisture product (L3_SM_P_E). For both the aforementioned SMAP soil moisture products, only the descending overpass, or local 6 A.M., data are used in this chapter considering the thermal equilibrium effect [110]. These products are gridded using the Equal-Area Scalable Earth-2 (EASE2) projection [111] where the 3-km subgrids are nested consistently inside the 9-km grid. In situ soil moisture observations from two soil moisture networks, Soil Climate Analysis Network (SCAN) [112] and U.S. Climate Reference Network (USCRN) [113] were used for additional validation at selected study areas from April 2015 to September 2018. The spatial resolution performance of the gap-filled soil moisture was further assessed by the airborne soil moisture data from the SMAP Validation Experiment 2016 (SMAPVEX16) at South Fork, Iowa (August 2016) [114].

3.2.2 Ancillary Data for Proposed Two-Layer Framework

Ancillary data used in the SMAP/Sentinel active-passive retrieval algorithm are also used in the proposed two-layer framework to capture the fine-scale heterogeneity in soil moisture and brightness temperature. Soil properties represented by the percent sand, percent clay, and bulk density were obtained from Soil Geographic (STATSGO) Data Base [115] with 1-km resolution. Land cover class was extracted from the MCD12Q1 [116] with 500-m resolution. Both the soil properties and land cover class are linearly averaged to 3 km. Effective soil temperature, scattering albedo, vegetation water content, vegetation opacity, and roughness coefficient were obtained from the 33-km SMAP-enhanced passive soil moisture product and are rescaled to 3 km using bilinear interpolation.

3.2.3 Ancillary Data for Traditional One-Layer Framework

Different from our proposed two-layer framework, existing machine learning methods normally predict soil moisture by capturing its relationship with other environmental variables, such as soil properties, topography, land cover, and meteorological forcings. We also implement this one-layer framework to compare with the proposed two-layer approach. Again, we take the soil properties from the STATSGO and landcover class from MCD12Q1. The elevation data were obtained from the NASA Shuttle Radar Topography Mission Global 3 arc second product [117]. Additional land surface controls represented by leaf area index and land surface temperature were extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS) products, MCD15A3H [118] with 500-m resolution and MOD11A1 [119] with 1-km resolution, respectively. Precipitation forcing data were obtained from the Global Precipitation Measurement [120] 3IMERGDF product with $\sim$11-km resolution. All these inputs were resampled to the 3-km EASE2-Grid projection through linear averaging (for inputs with resolutions higher than 3 km) or using bilinear interpolation (for inputs with resolutions lower than 3 km) to match the grid size of the learning targets. Slope was then computed from the resampled elevation data sets. Land surface temperature and precipitation data have a daily temporal resolution, while the leaf area index has a 4-day temporal resolution. We assume that the leaf area index stays the same for each 4-day period.

Additional post-processing was applied to land surface temperature and precipitation data. As retrieval for land surface temperature depends on the satellite thermal infrared channels, invalid surface temperature values are generated for surfaces that are obscured by clouds [121]. For these grids with invalid values, we search for historical values within a 2-week finite window and select the most recent valid historical values as substitutes. For precipitation, to incorporate its lagged effect on vegetation, surface characteristics,

and soil moisture, we use precipitation values from recent 3 days separately as inputs.

3.2.4 Study Area

We apply the proposed framework separately at four U.S. states with distinct climate regimes and land surface heterogeneities: Arizona, Oklahoma, Iowa, and Arkansas (Figure 3.2). Arizona has primarily an "arid" climate (BSk and BWh classification), Oklahoma is "warm-temperate" (Cfa classification), Iowa is primarily "snow" with fully humid hot summer (Dfa classification), while Arkansas is "warm temperate" (Csc classification) according to the Köppen-Gieger classification [122]. While Arizona, Oklahoma, and Iowa are predominantly composed of open shrublands (70.8%), grasslands (75.5%), and croplands (85.1%) respectively, Arkansas is covered with forest (28.7%), croplands (25.5%) and woody savannah (19%) [116]. Arizona has a high mean elevation (above mean sea level) of 1283.1 m, while Oklahoma, Iowa, and Arkansas have mean elevations of 398.3, 323.5, and 166.2 m, respectively [117]. In Figure 3.2, we show daily examples of the 33- and 3-km soil moisture. Arizona with the "arid-steppe-hot" climate generally has dry conditions compared with the other three states (Figure 3.2). Even with low soil moisture values, high levels of heterogeneity can still be seen at the 3-km resolution, especially for areas with latitude lower than 35°N. Oklahoma, Iowa, and Arkansas have soil moisture values ranging from 0.0 to 0.5 cm^3/cm^3. While the high-resolution soil moisture has more variability at Arkansas (Figure 3.2(g) and 3.2(h)), Oklahoma and Iowa are spatially homogeneous at 3 km (Figure 3.2(c) and 3.2(d) and Figure 3.2(e) and 3.2(f)).

3.3 The SMAP/Sentinel Active-Passive Retrieval Algorithm

We provide here an overview of the retrieval algorithm adopted by the SMAP/Sentinel-1 3-km soil moisture product [21, 123, 22] that motivates the design of our proposed two-layer machine learning-based framework. This retrieval algorithm adopts a two-layer strategy where coarse-resolution SMAP radiometer brightness tem-

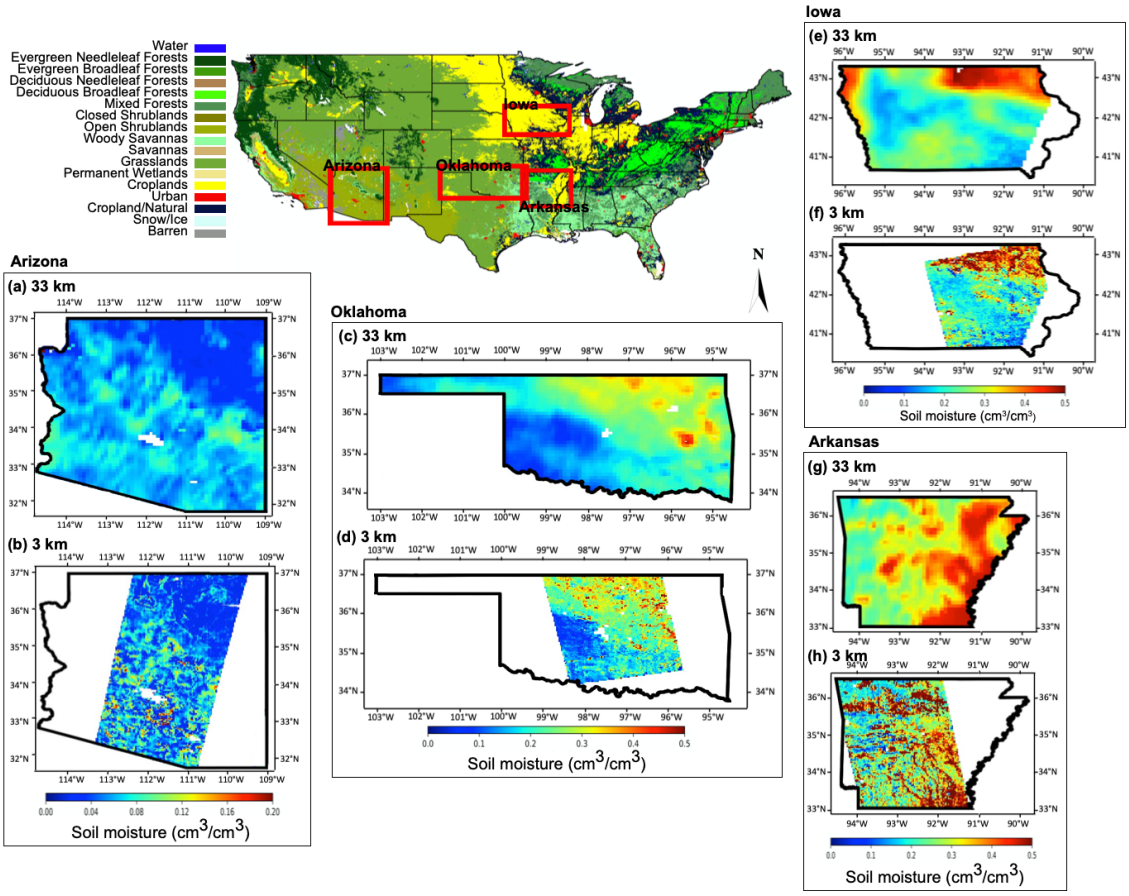

Figure 3.2: Locations of the four study areas. The background shows the MODIS International Geosphere Biosphere Programme (IGBP) land cover class. (a) The 33-km soil moisture and (b) 3-km soil moisture at Arizona on 13 June 2018. (c) The 33-km soil moisture and (d) 3-km soil moisture at Oklahoma on 1 June 2018. (e) The 33-km soil moisture and (f) 3-km soil moisture at Iowa on 9 June 2018. (g) The 33-km soil moisture and (h) 3-km soil moisture at Arkansas on 3 June 2018. The 33-km soil moisture is converted from 33-km brightness temperature through the single channel algorithm, while 3-km soil moisture is the downscaled product from the Soil Moisture Active Passive-Sentinel active-passive algorithm.

perature is first downscaled using overlapping high-resolution Sentinel radar backscatter data followed by soil moisture retrieval through the SCA (Figure 3.3(a)).

The algorithm first disaggregates the coarse-scale brightness temperature measurements (T_{B}, at polarization v or h) with the aid of fine-scale radar backscatter measure-

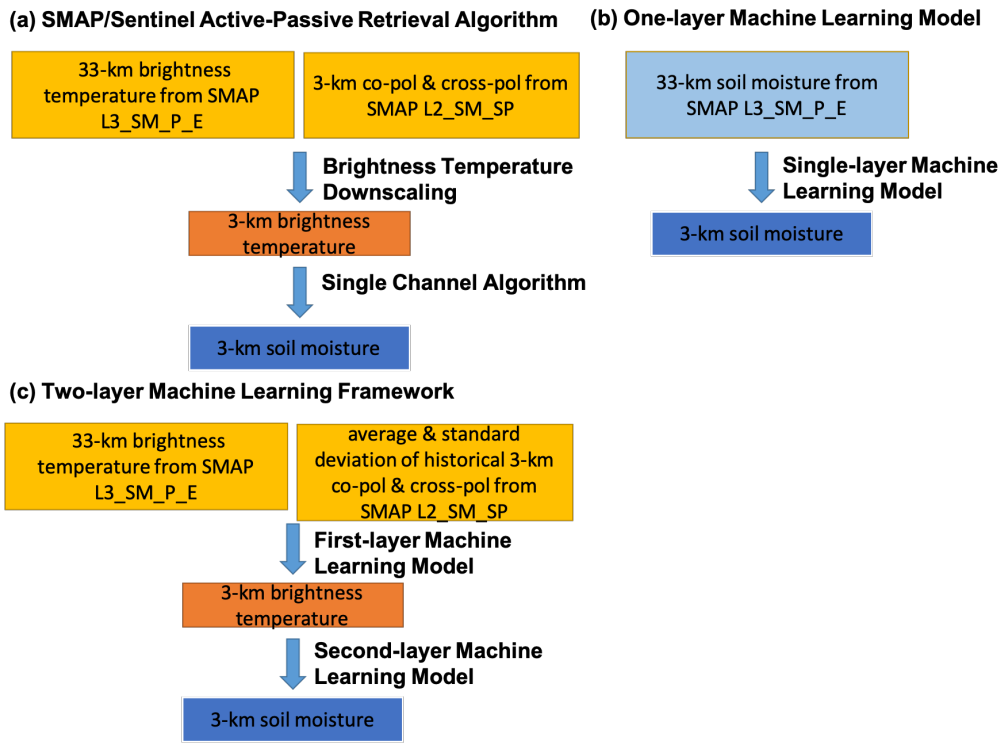

Figure 3.3: Architecture comparison of (a) the SMAP/Sentinel active-passive retrieval algorithm, (b) the one-layer machine learning model, and (c) the two-layer machine learning framework. Only key inputs are listed here. SMAP = Soil Moisture Active Passive.

ments. Both co-pol (σ_{vv} or σ_{hh}) and cross-pol (σ_{vh}) radar backscatter cross sections (in normalized/linear scale) are used in the algorithm. Cross-pol measurements are included in the algorithm to capture the subgrid variations of brightness temperature within a coarse-scale footprint due to the variations in vegetation characteristics and surface roughness. The vegetation characteristics and surface roughness are assumed to be invariant within a short period of time. Suppose that we have brightness temperature measurements at a coarse support "*C*" and radar backscatter measurements at a fine support "*M*." Within a single pixel of the grid at support *C*, the disaggregated brightness temper-

ature of a subpixel at support *M* can be obtained using

$$T_{\mathrm{B}}(M) = [\frac{T_{\mathrm{B}}(C)}{T_{\mathrm{S}}} + \beta(C) \cdot \{[\sigma_{vv}(M) - \sigma_{vv}(C)] + \Gamma \cdot [\sigma_{vh}(C) - \sigma_{vh}(M)]\}] \cdot T_{\mathrm{S}}. \quad (3.1)$$

T_{S} is the effective surface temperature of the near-surface soil (top $\sim$5 cm). Parameter β is derived from retrieved brightness temperature, radar backscatter, and other physically based parameters, such as scattering albedo and vegetation opacity, to estimate the covariation of the emission and backscatter observations. The sensitivity parameter, $\Gamma \equiv [\delta\sigma_{vv}(M)/\delta\sigma_{vh}(M)]$, is used to convert the subgrid heterogeneity captured by the cross-pol deviations $[\sigma_{vh}(C) - \sigma_{vh}(M)]$ to co-pol deviations. Once the downscaled brightness temperature is obtained, the SCA [89] is applied to retrieve the downscaled soil moisture.

Table 3.1 lists the inputs required for downscaling brightness temperature and SCA separately. For some of the derived inputs like $\sigma_{vv}(M)$, we also list their corresponding data sources. Physically based parameters like scattering albedo and vegetation opacity depend on vegetation structures, polarization, microwave frequency, and so forth. They are currently obtained through a land cover-driven lookup table that contains parameter values under different land cover classes [110]. It can be seen from the table that the SMAP/Sentinel active-passive retrieval algorithm, especially the brightness temperature downscaling, requires concurrent radar and radiometer observations on the same day. The narrow imaging swath and low revisit schedule of the Sentinel radars thus severely restricts the spatial and temporal coverage of the resulting downscaled soil moisture estimates. In the next section, we propose a machine learning framework to account for this shortcoming.

Table 3.1: Inputs of the SMAP/Sentinel Active-Passive Retrieval Algorithm.

Algorithm	Inputs
Brightness temperature downscaling	$T_{\mathrm{B}}(C)$, T_{S}, $\sigma_{vv}(M), \sigma_{vv}(C), \sigma_{vh}(M), \sigma_{vh}(C)$, $\beta(C) \leftarrow (T_{\mathrm{B}}(C), T_{\mathrm{S}}, \sigma_{vv}(M), \sigma_{vh}(M)$, scattering albedo, vegetation opacity, incidence angle), $\Gamma(C) \leftarrow (\sigma_{vv}(M), \sigma_{vh}(M))$
Single channel algorithm	T_{B}, T_{S}, vegetation water content, vegetation opacity, sand and clay percentages, bulk density, scattering albedo, roughness coefficient, land cover class

3.4 Methodology

3.4.1 Two-layer Machine Learning Framework

To downscale soil moisture that has a high consistency with the SMAP/Sentinel-1 3-km soil moisture estimates at gap areas, we hypothesize that it would be beneficial to learn a two-layer machine learning framework whose structure and inputs are guided by the SMAP/Sentinel active-passive retrieval algorithm.

First, we formulate the relationship learned by the first-layer machine learning model to disaggregate the 33-km brightness temperature to 3 km. Following the SMAP/Sentinel active-passive algorithm, we rewrite equation (3.1) as

$$\frac{T_{\mathrm{B}}(3\mathrm{km}) - T_{\mathrm{B}}(33\mathrm{km})}{T_{\mathrm{S}}} = \beta \cdot \{[\sigma_{vv}(3\mathrm{km}) - \sigma_{vv}(33\mathrm{km})] + \Gamma \cdot [\sigma_{vh}(33\mathrm{km}) - \sigma_{vh}(3\mathrm{km})]\}. \tag{3.2}$$

The left side of equation (3.2) can serve as the learning target of our first-layer machine learning model. It is available at areas where the radar and radiometer observations are available and missing at the gap areas where the radar observations are absent.

The parameters on the right side of equation (3.2) are restricted by the availability

of radar observations and therefore incur gap areas for the SMAP/Sentinel-1 3-km soil moisture product. To remove this dependence, we define the function f_1 that will be learned by our first-layer machine learning model as

$$y = \frac{T_{\mathrm{B}}(3\mathrm{km}) - T_{\mathrm{B}}(33\mathrm{km})}{T_{\mathrm{S}}} = f_1(rb^{\mathrm{hist}}, \beta^{\mathrm{hist}}, \Gamma^{\mathrm{hist}}, \omega, \tau, lat, lon). \tag{3.3}$$

ω and τ refer to the scattering albedo and vegetation opacity, respectively, and can be obtained from the SMAP-enhanced passive soil moisture product. They are thus available in both covered and gap areas. Latitude and longitude were added to this framework, and all following frameworks, to aid the machine learning models to learn more localized patterns. Instead of directly using the radar observations $\sigma_{vv}(3\mathrm{km})$ and $\sigma_{vh}(3\mathrm{km})$ that are not available at gap areas, we use the average of historical radar backscatter measurements rb^{hist} within a 30-day window as substitutes to capture the subgrid heterogeneity in vegetation and surface physical characteristics. This substitution is based on the knowledge that radar backscatter measurements can capture the subgrid land-surface heterogeneity and can be assumed to have less variability, compared with the dynamics of soil moisture, within a finite temporal window [21, 123]. To the best of our knowledge, this is the first study utilizing the average of historical radar backscatter measurements for downscaling/gap-filling soil moisture. In addition to the average, we also include the standard deviation of the historical 3-km radar backscatter measurements as an input to reflect the extent to which they change during the past 30 days. Thirty-day historical averages of β (β^{hist}) and Γ (Γ^{hist}) are also introduced to improve the efficacy of the models.

Using these ancillary data, the first-layer machine learning model will be trained at areas where the learning target and inputs are available. The learned model can then be used to obtain the predicted $\hat{y}$ at areas where $T_{\mathrm{B}}(3\mathrm{km})$ is missing. We use the "hat" symbol over a variable to denote that it is a predicted value. The 3-km brightness temperature at

these gap areas can, therefore, be reconstructed through

$$\hat{T}_B(\text{3km}) = T_{\text{B}}(\text{33km}) + \hat{y} \cdot T_{\text{S}}. \tag{3.4}$$

After downscaling the 33-km brightness temperature, the downscaled 3-km soil moisture is obtained through the second-layer machine learning model f_2 given by

$$\widehat{SM}(\text{3km}) = f_2(\hat{T}_B(\text{3km}), T_{\text{S}}, VWC, S, \omega, \tau, h, L, lat, lon). \tag{3.5}$$

VWC, S, h, and L refer to vegetation water content, soil properties including sand and clay percentages and bulk density, roughness coefficient, and land cover class, respectively. These features, along with the predicted 3-km brightness temperature, effective surface temperature, scattering albedo, and vegetation opacity, also form inputs to the original SCA algorithm. We call this machine learning framework as "two-layer."

3.4.2 Traditional One-Layer Machine Learning Framework

In addition to the two-layer machine learning-based framework, we also implement a one-layer framework which directly downscales soil moisture from 33 to 3 km using ancillary data in section 3.2.3. It also has the ability to gap fill soil moisture at missing areas as long as the ancillary data are available. The function f_3 learned by this framework can be written as:

$$\widehat{SM}(\text{3km}) = f_3(SM(\text{33km}), \mathcal{H}, lat, lon), \tag{3.6}$$

where $\mathcal{H}$ denotes the feature set including leaf area index, land surface temperature, soil properties, elevation, slope, land cover, and precipitation which have been shown to be the major drivers of soil-moisture variability in past studies [27, 124, 125, 28]. In this chapter, we call this framework as "one-layer."

Figure 3.3(b) and 3.3(c) show the structures of the one-layer framework and the pro-

posed two-layer framework, respectively. Comparing Figure 3.3(a) and Figure 3.3(c), it can be seen that the two-layer machine learning-based framework follows a similar structure as the original SMAP/Sentinel active-passive retrieval algorithm while using a machine learning paradigm. Its performance over the one-layer framework in terms of downscaling soil moisture at gap areas will be demonstrated through extensive machine learning-based experiments as discussed in section 3.5.

3.4.3 Additional Machine Learning Frameworks for Comparison

For a comprehensive evaluation, we also compare the proposed two-layer with the following machine learning frameworks:

- Two-layer+: Two-layer framework but with ancillary data from both sections 3.2.2 and 3.2.3 as inputs.
- One-layer+ (SM): One-layer framework but with ancillary data from both sections 3.2.2 and 3.2.3 as inputs.
- One-layer+ (BT): One-layer framework which converts 33-km brightness temperature (instead of 33-km soil moisture) directly into 3-km soil moisture and with ancillary data from both sections 3.2.2 and 3.2.3 as inputs.

The "two-layer+" and "one-layer+ (SM)" allows us to evaluate the effect of adding "extra" covariate information in the machine learning frameworks. The "one-layer+ (BT)" is also evaluated to demonstrate the potential advantage of using a two-layer framework as opposed to directly downscaling brightness temperature to 3-km soil moisture.

3.4.4 Random Forest

In this chapter, for all the machine learning frameworks, the Random Forest [7] will be used as the "machine learning model" for downscaling soil moisture/brightness temperature (i.e. the model used in f_1, f_2 and f_3 in equations (3.3), (3.5), (3.6), respectively).

Proposed in 2001, the Random Forest model quickly gained its reputation for being able to learn complex nonlinear relationships between the target variable and input features while resisting to the overfitting problem. Its superior performance is achieved through an ensemble of individually learned decision trees combined with the bagging method and random selections of features. During the training process of the Random Forest model, multiple hierarchical decision trees are learned independently with the goal of maximizing the variance reduction for each split. The individually grown decision trees can, however, easily suffer from the overfitting problem where significantly disparate performances can be observed on the training and test sets. To tackle this issue, the Random Forest model averages the predictions generated by all trees in the forest with additional strategies to reduce the correlation between them. The correlation reduction is done through bagging and the random selections of features. During the training process, each decision tree is exposed to different data subsets that are randomly sampled with replacement from the same training set. Additionally, randomly selected subset of the features are considered for each split. The Random Forest model generalizes well on test set when strong decision trees are constructed with controlled correlations between them.

3.4.5 Spatiotemporal Regional Learning

For a machine learning framework, we test the following three hypotheses in this chapter: High-resolution soil moisture/brightness temperature at gap areas can be reconstructed from the models learned at (1) spatial neighboring regions, (2) temporal neighboring regions, and (3) both spatial and temporal neighboring regions. We define these three settings as the spatial regional learning, the temporal regional learning, and the spatiotemporal regional learning, respectively. For the temporal regional learning and the spatiotemporal regional learning, a 30-day window is adopted to search for the most recent temporal neighboring regions for the gap regions. As shown in Figure 3.4,

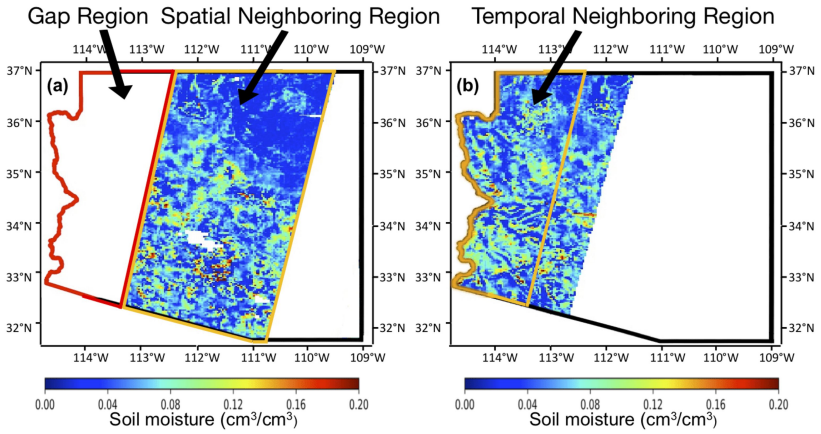

Figure 3.4: 3-km soil moisture swath at Arizona on (a) 13 June 2018 and (b) 5 June 2018 from Soil Moisture Active Passive L2_SM_SP.

machine learning models are built using data available at covered spatial and/or temporal neighboring regions (inside the orange frames) and are subsequently used to predict high-resolution soil moisture at the gap region (inside the red frame) where the original SMAP/Sentinel-1 active-passive soil moisture product is missing. In this chapter, all machine learning frameworks are trained for individual regions at a daily scale. The success of these machine learning frameworks depends on how effectively information learned at spatial/temporal neighboring regions can be extrapolated to the gap region.

The three regional learning settings enable expanded training data from spatiotemporally neighboring data regions. The expanded training data enabled by the regional learning settings allow the models to learn complex relationships between coarse-scale and fine-scale soil moisture and achieve higher predictive capabilities. Note that we focus on gap filling the "neighboring" spatiotemporal regions in this chapter which fall into the same or, at least, similar hydroclimates as the regions where the models are learned. Building models that can be applied to "almost everywhere" or discussing whether a model learned under one hydroclimate is applicable to another is beyond the scope of this chapter and will be explored in the future.

3.5 Results and Discussions

We validate the downscaled soil moisture at four selected study areas, Arizona, Oklahoma, Iowa, and Arkansas, for the 3.5-year period between 1 April 2015 and 30 September 2018. Besides the one-layer framework (equation (3.6)), we also compare the performance of our proposed two-layer framework with the 33-km soil moisture from the SMAP-enhanced passive soil moisture product. As the SMAP-enhanced passive soil moisture product is posted on 9-km EASE2 grids, we resample the 33-km soil moisture to 3 km by simply assuming that all the 3-km EASE2 subgrids lying inside one 9-km EASE2 grid have the same value. We denote this product as the "resampled 33-km soil moisture." The resampled 33-km soil moisture estimates serve as a reference for comparison with the one-layer and the proposed two-layer frameworks. In terms of metrics, the correlation coefficient (R) and unbiased root mean square error (ubRMSE) are used to quantify and compare the overall agreement between the predicted 3-km soil moisture and the retrieved ones. Note that, for every study region (Arkansas, Arizona, Iowa, and Oklahoma), each machine learning framework is trained individually for each day, and the results in the subsequent sections are summarized using boxplots. Average metrics from five machine learning runs are reported for all the experiments.

3.5.1 Thirty-Three Kilometers versus One-Layer versus Two-Layer

To demonstrate the performance of our proposed two-layer framework, we split the covered regions (where the 3-km soil moisture estimates are available) into training and test regions. Machine learning models learned at the training region for each day are used to predict the downscaled soil moisture/brightness temperature at the test region. At each of the selected study areas, Arizona, Oklahoma, Iowa, and Arkansas, there are two to three different major swath patterns with $\sim$12-day interleaved revisit schedules. As shown in Figure 3.5, we select two swath patterns for each of the study areas, the main

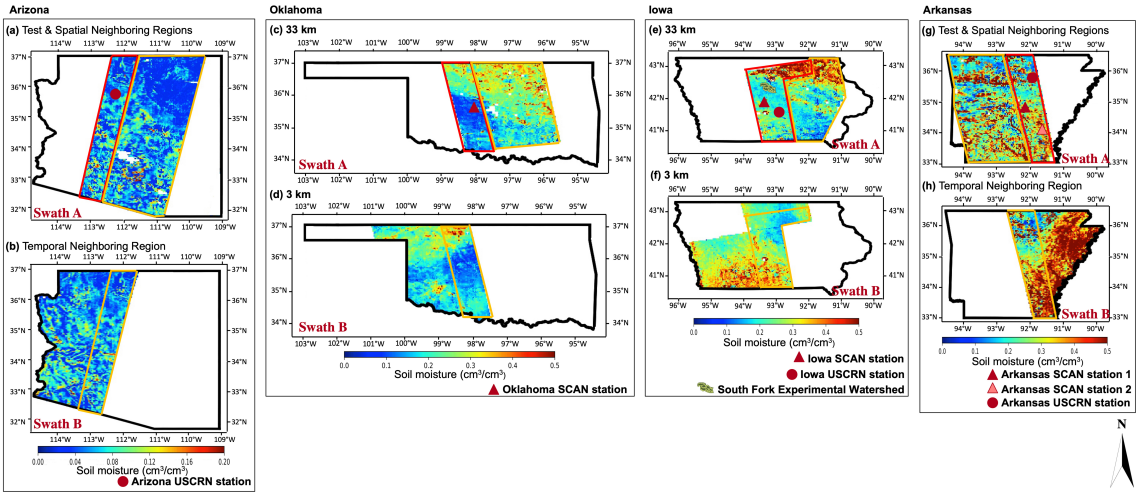

Figure 3.5: Train and test splits at Arizona, Oklahoma, Iowa, and Arkansas. (a) Selected swath A at Arizona on 13 June 2018. (b) Selected swath B at Arizona on 5 June 2018. (c) Selected swath A at Oklahoma on 1 June 2018. (d) Selected swath B at Oklahoma on 24 May 2018. (e) Selected swath A at Iowa on 9 June 2018. (f) Selected swath B at Iowa on 3 June 2018. (g) Selected swath A at Arkansas on 8 July 2018. (h) Selected swath B at Arkansas on 21 June 2018. SCAN = Soil Climate Analysis Network; USCRN = U.S. Climate Reference Network.

swath A and the shadow swath B. The overlap of the two swaths will be used as a filter to obtain the holdout test region. For any specific day which has the main swath A available, the filter will be applied to obtain the test region (as shown in the red frames of Figures 3.5(a), 3.5(c), 3.5(e) and 3.5(g)), while the rest will be used as spatial neighboring regions for the training process (as shown in the orange frames in Figures 3.5(a), 3.5(c), 3.5(e) and 3.5(g)). For the temporal regional learning task, a 30-day sliding window is adopted to search for the most recent shadow swath B. Once the most recent shadow swath B is obtained, the filter of the test region is applied to obtain the temporal neighboring region (as shown in the orange frames of Figures 3.5(b), 3.5(d), 3.5(f) and 3.5(h)). This training and test splitting strategy mimics how the regional learning is applied to fill/downscale soil moisture at real gap areas where the 3-km soil moisture estimates are indeed missing. It also allows us to compare the performance of the three regional learning settings,

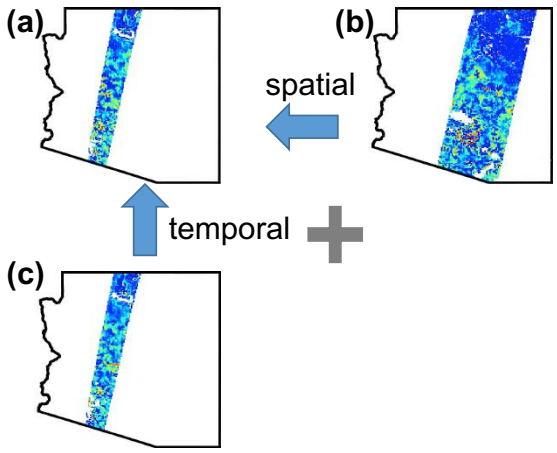

Figure 3.6: Spatial/temporal regional learning for (a) Test area at Arizona on 13 June 2018. (b) Spatial neighboring region on the same day. (c) Search for temporal neighboring regions within 30-day sliding window is applied to obtain the most recent temporal neighboring region on 5 June 2018.

Table 3.2: Number of Days Covered for Each Quarter of a Year (1 April 2015 to 30 September 2018) and Average Sizes of Spatial Training Regions and Test Regions. rT = Ratio.

Year	2015			2016				2017				2018					
Quarter	2	3	4	1	2	3	4	1	2	3	4	1	2	3	Train	Test	rT
Arizona	-	-	-	-	-	-	-	2	4	6	7	6	7	3	11,463	3,676	3.1
Oklahoma	-	2	-	-	4	4	5	1	5	2	3	4	6	8	4,306	2,330	1.8
Iowa	-	-	-	-	-	6	-	4	5	7	6	2	7	7	7,773	2,512	3.1
Arkansas	-	-	-	-	-	-	3	7	5	7	5	7	7	7	4,782	2,215	2.1

the spatial regional learning, temporal regional learning, and the spatiotemporal regional learning. Figure 3.6 shows, as an example, the test region at Arizona on 13 June and its corresponding spatial neighboring region and temporal neighboring region for the training. Table 3.2 lists the number of days covered for each quarter of a year at the selected study areas where the main swath A is available for test and spatial neighboring regions and the shadow swath B is available for a temporal neighboring regions search. It can be seen from the table that Arizona, Iowa, and Arkansas have no coverage in 2015 and the

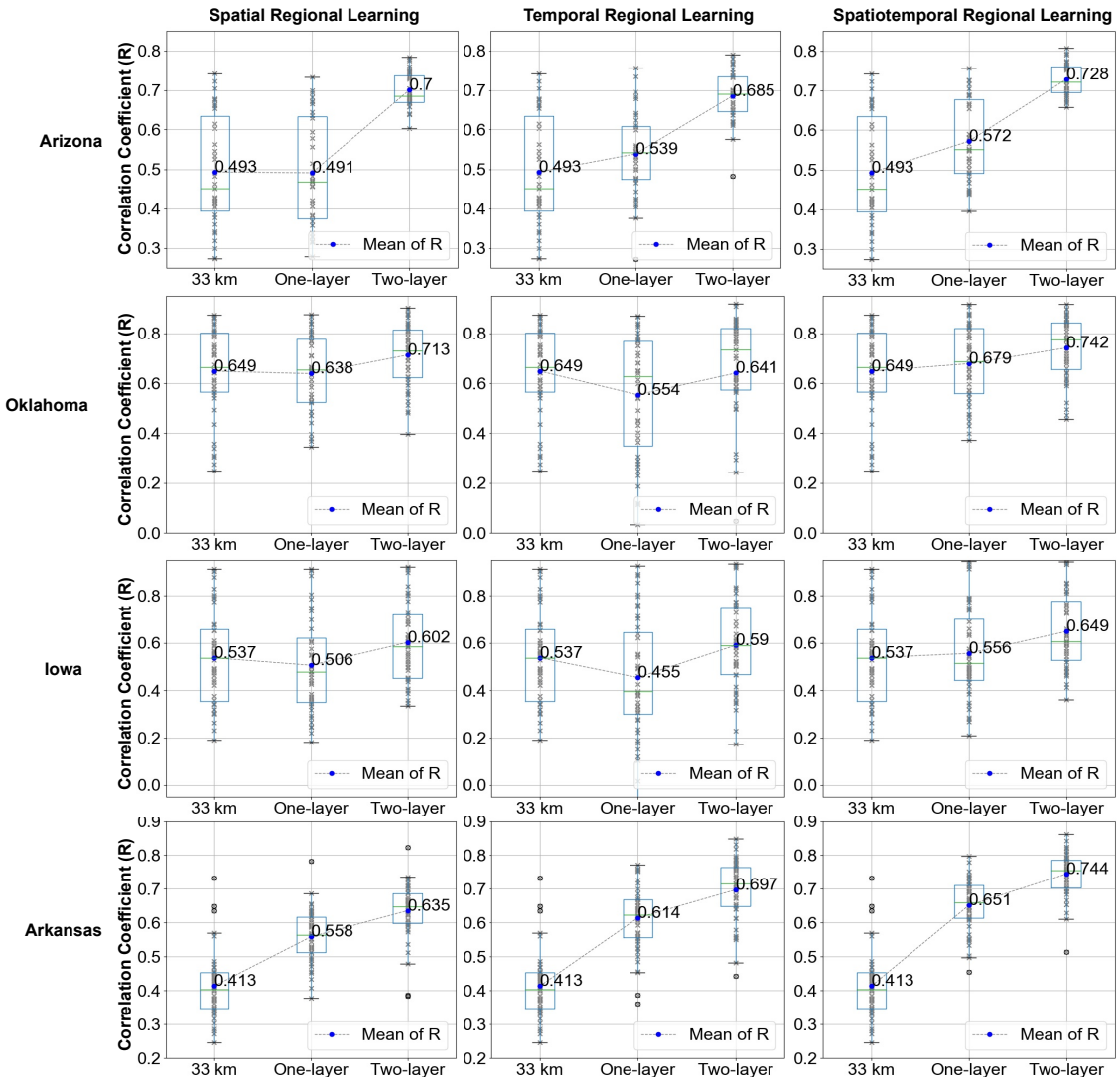

Figure 3.7: Correlation coefficient (R) comparisons of the resampled 33 km, one-layer model and our proposed two-layer framework for the selected study areas under different regional learning settings.

first three quarters of 2016 which is due to the limited/no coverage of the SMAP/Sentinel-1 3-km soil moisture product during that period. The average sizes (number of pixels available) of spatial training and test regions and their corresponding ratio for each area are also listed in the table. The sizes of temporal training regions are the same as the sizes of the test sets.

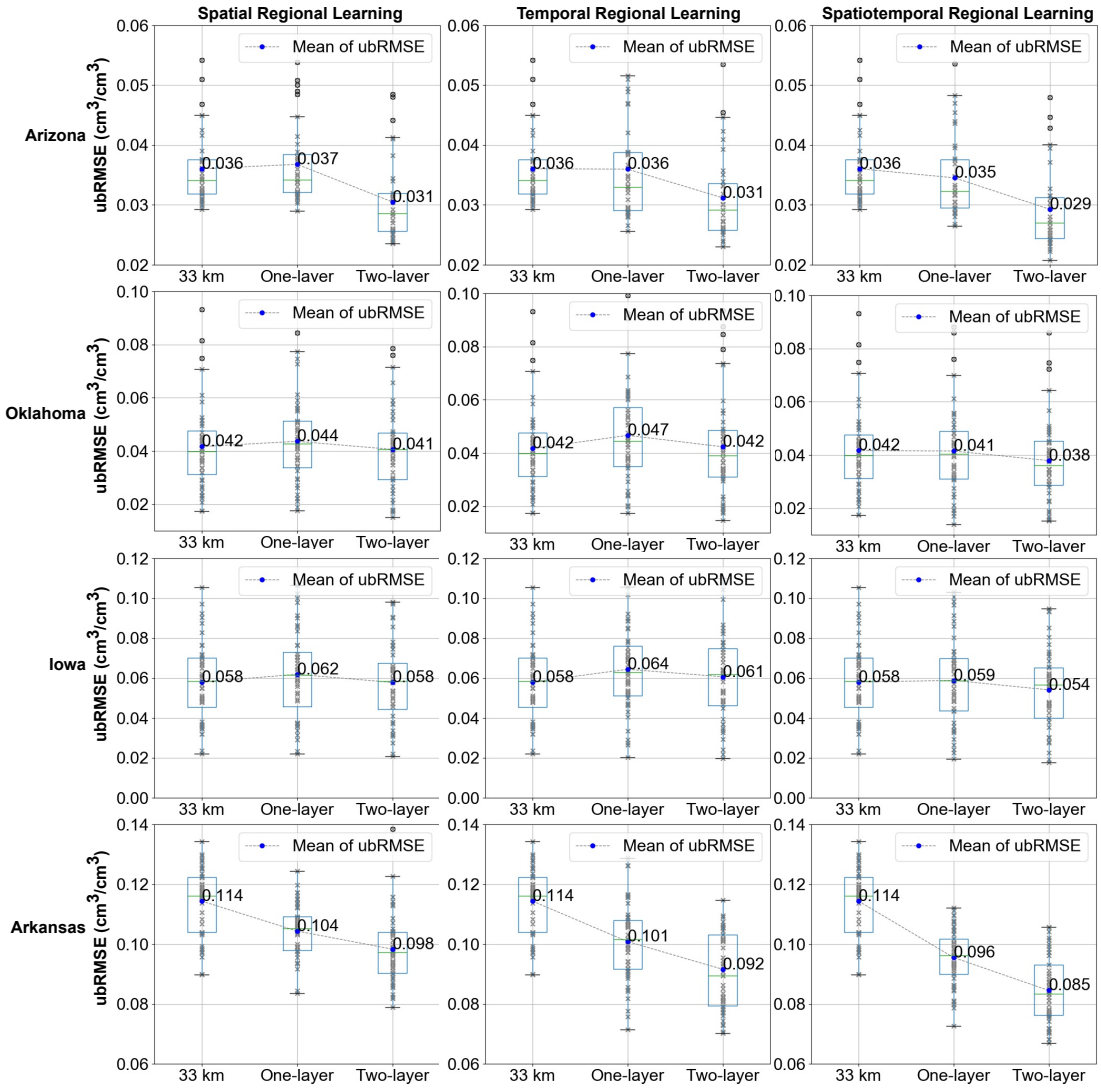

Figure 3.8: ubRMSE comparisons of the resampled 33km, one-layer model and our proposed two-layer framework for the selected study areas under different regional learning settings. ubRMSE = unbiased root mean square error.

Figures 3.7 and 3.8 summarize the performance (in terms of R and ubRMSE, respectively) of the resampled 33-km soil moisture, one-layer framework and our proposed two-layer framework under different regional learning settings at available test regions for the entire duration of the analysis. At each selected study area, performance metrics are summarized by boxplots to show a general pattern of how different frameworks per-

form. It can be seen that the proposed two-layer framework generally outperforms both the resampled 33-km soil moisture product and the one-layer framework in terms of R and ubRMSE under all three regional learning settings. The improvement in R is significant, especially at Arizona and Arkansas. Taking the spatiotemporal regional learning as an example, the mean R at Arizona is increased to 0.728 for the two-layer framework compared to 0.572 (27% improvement) for the one-layer framework and 0.493 (47% improvement) for the 33-km soil moisture, while the mean R at Arkansas is increased to 0.744 for the two-layer framework compared to 0.651 (14% improvement) for the one-layer framework and 0.413 (80% improvement) for the 33-km soil moisture. Exceptions of improvement in performance are noticed at Oklahoma and Iowa for the spatial regional learning and temporal regional learning where relatively insignificant changes in the chosen metrics are found when comparing the resampled 33-km soil moisture product, the one-layer framework, and the two-layer framework. This is, however, expected as the retrieved 3-km soil moisture has much lesser spatial heterogeneity at Oklahoma and Iowa as discussed in section 3.2.4.

To visually compare the performance of these frameworks, Figure 3.9, as an example, focuses on a daily test region on 13 June 2018 at Arizona to show the disaggregated soil moisture produced by the resampled 33-km soil moisture product, the one-layer framework, and the proposed two-layer framework. It can be seen that although the one-layer framework under the spatiotemporal regional learning has the ability to reconstruct fine-scale soil moisture to some extent (Figure 3.9(c)), its downscaling quality is still inferior to the proposed two-layer framework, even when compared to the spatial regional learning only approach (Figure 3.9(d)). With the proposed two-layer framework, more fine-scale soil moisture heterogeneity can be predicted/downscaled with improvement in both R and ubRMSE (Figure 3.9(f)). Figure 3.10 shows the pixel-by-pixel comparison of the correlation of the resampled 33 km, one-layer framework, and the proposed two-layer framework

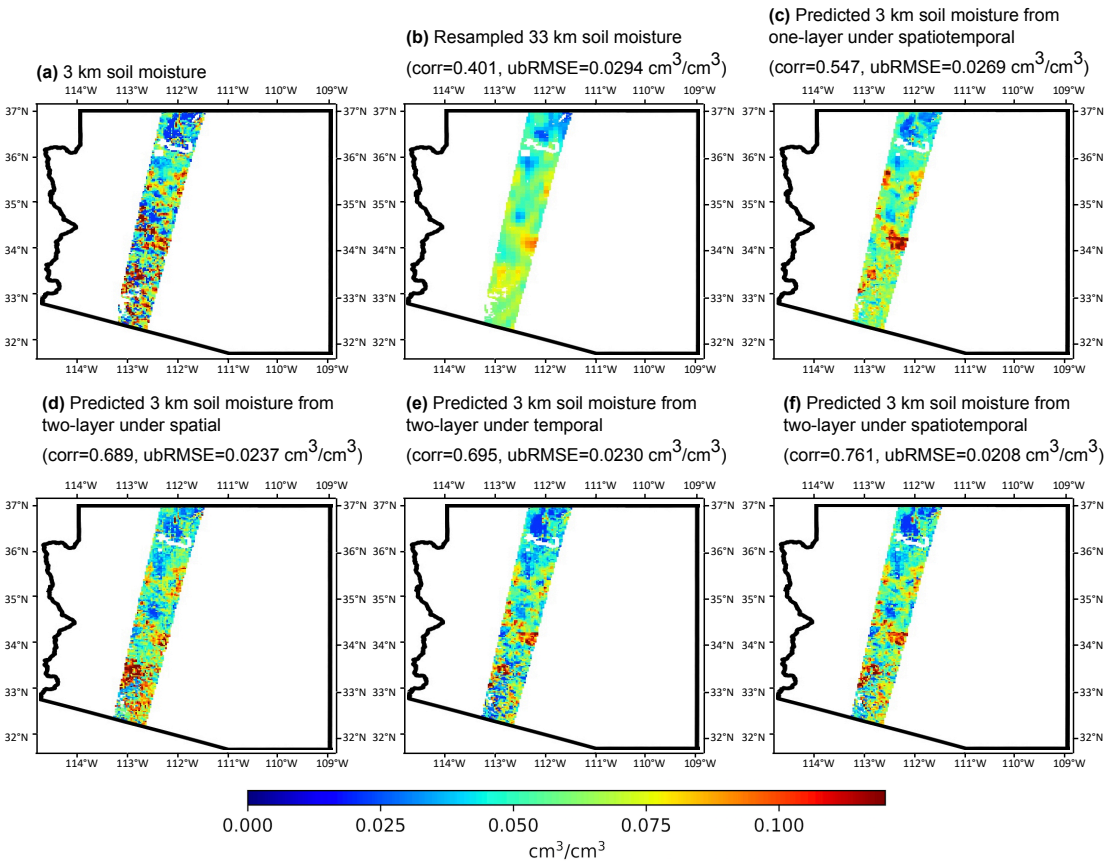

Figure 3.9: Performance comparisons for test region at Arizona on 13 June 2018. (a) The 3-km soil moisture from SMAP/Sentinel-1 active-passive product. (b) Resampled 33-km soil moisture from SMAP-enhanced passive product. Predicted 3-km soil moisture from (c) the one-layer model under spatiotemporal regional learning, (d) the two-layer framework under spatial regional learning, (e) the two-layer framework under temporal regional learning, and (f) the two-layer framework under spatiotemporal regional learning. SMAP = Soil Moisture Active Passive; ubRMSE = unbiased root mean square error.

under the spatiotemporal regional learning with the SMAP/Sentinel-1 observations. The correlation values are calculated for all available values inside each 33 km $\times$ 33 km pixel of the test region. As the SMAP 33-km-enhanced passive soil moisture product is actually posted on a 9-km resolution grid, it has nonzero correlation values inside the 33 km $\times$ 33 km pixels. Figure 3.10, again, shows that our two-layer framework increases the correlation between the gap-filled predicted soil moisture and the SMAP/Sentinel-1

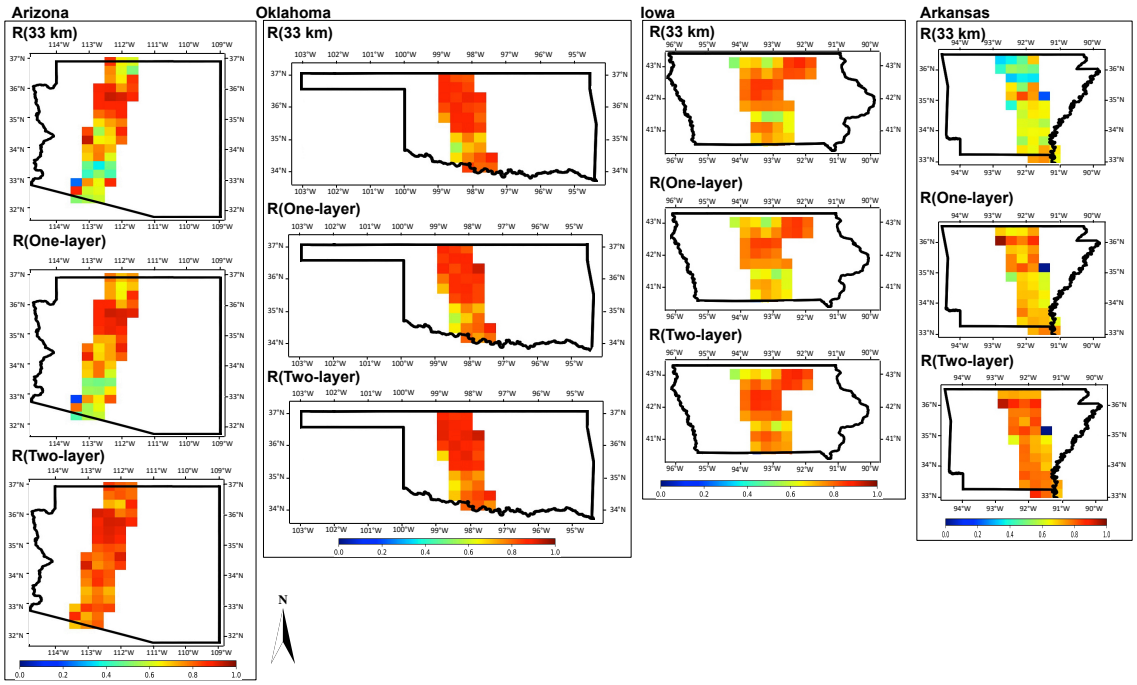

Figure 3.10: Correlation of the resampled 33 km, one-layer model, and our proposed two-layer model with the SMAP/Sentinel-1 observations under the spatiotemporal regional learning. Each pixel in this figure is a patch of 11 $\times$ 11 3-km Equal-Area Scalable Earth-2 grids. SMAP = Soil Moisture Active Passive.

3-km observations compared with both the resampled 33-km soil moisture and the one-layer predictions, though as mentioned earlier, the improvement is moderate in Iowa and Oklahoma.

Since the above discussion summarizes the performance of the proposed framework for the entire duration of the data, it is also worthwhile to look at how the framework performs seasonally. To achieve this, we group the results from all years into four quarters (January-March, April-June, July-September, and October-September) and compare R and ubRMSE across these time periods. (Figure 3.11). It can be seen that the two-layer framework consistently outperforms the 33-km soil moisture and the one-layer framework across seasons. It can also be seen that for the majority of the seasons, the scatter of the boxplot for the two-layer is also less, making the predictions consistently accurate across

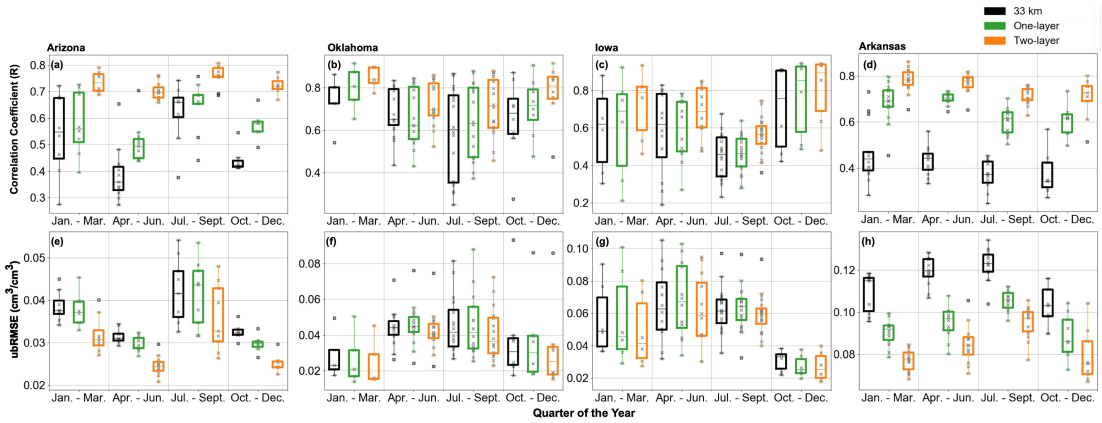

Figure 3.11: Seasonal performance metrics of the two-layer framework, one-layer framework, and 33-km soil moisture. The two-layer framework outperforms the other two consistently across seasons.

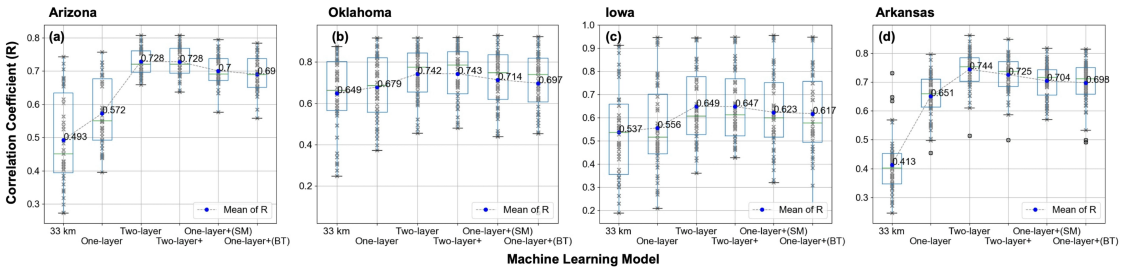

Figure 3.12: Performance comparisons of the proposed two-layer machine learning models with three other machine learning models with different feature designs. Two-layer+ adds auxiliary environmental variables to the two-layer model. One-layer+ (SM) is a one-layer model which uses all the features that are used in the two-layer models but directly downscale soil moisture from 33 to 3 km. One-layer+ (BT) is a one-layer model which directly downscales 33-km brightness temperature to 3-km soil moisture.

all days in a season.

3.5.2 Comparison With Additional Frameworks

Since in the previous section, we found that the spatiotemporal regional learning outperforms the spatial-only and temporal-only approaches, we now compare the two-layer framework with machine learning frameworks defined in section 3.4.3 under the spa-

tiotemporal regional learning using the same test-train split criterion as section 3.5.1. Figure 3.12 compares the various frameworks in terms of the mean R between the predictions and observation in the test region for all four study areas. We find that for all the study areas, there is insignificant improvement between the two-layer and two-layer+ frameworks denoting that adding the extra features from section 3.2.3 has no effect in improving soil moisture predictions. We can also see that the two-layer outperforms both the one-layer+ (SM) and one-layer+ (BT) across all four study regions which underscores the importance of using a two-layer machine learning framework for soil moisture downscaling rather than directly downscaling 33-km soil moisture or 33-km brightness temperature to high-resolution soil moisture. Note that, as mentioned earlier, the Random Forest forms the "machine learning model" for all the analyzed frameworks.

3.5.3 Feature Importance of Input Data in Two-Layer Framework

Another advantage of using the Random Forest model in the two-layer framework is that it has the ability to assess the relative importance of the covariates used as inputs to the model. Figure 3.13 shows the relative feature importance (normalized by the sum) of the input data (Table 3.1) in the first layer of the framework. It can be seen that the 30-day historical radar-backscatter data (co-pol hist and cross-pol hist in Figure 3.13) are the most important input of the first layer of the framework followed by a moderate effect of the rest of the parameters. This implies that incorporating the historical radar backscatter data from both spatial and temporal regions can significantly improve the prediction capabilities of the machine learning model to predict downscaled brightness temperature. For the second layer of the model, as expected, the predicted 3-km brightness temperature from Layer 1 is the most important input followed by surface temperature for Arizona and land cover for Arkansas, Iowa, and Oklahoma (Figure 3.14). For Arkansas, the effect of land cover is especially significant because, as mentioned earlier in section 3.2.4,

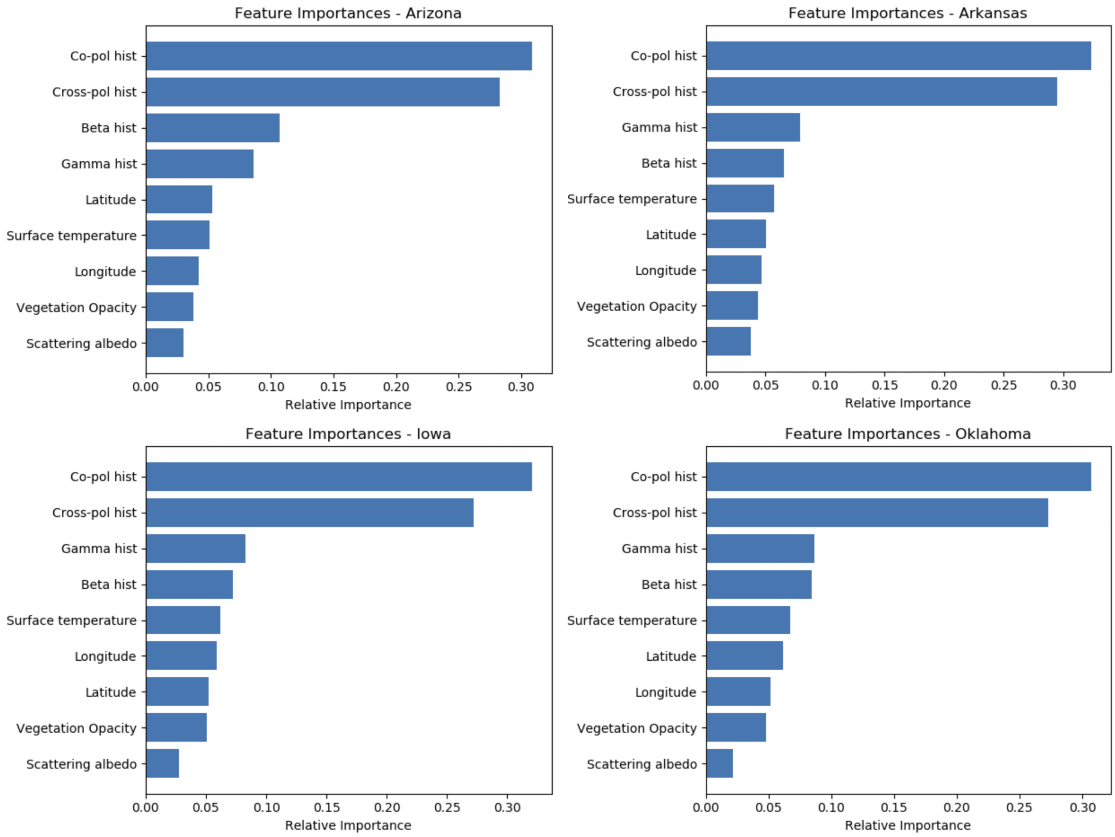

Figure 3.13: The 30-day average historical radar backscatter data (co-pol hist and cross-pol hist) are the most important feature for Layer 1 of the machine learning framework with moderate importance of the rest of the features. Co-pol = copolarization; cross-pol = cross-polarization.

Arkansas consists of diverse land cover classes compared to the other study regions which comprise predominantly (more than 70%) of one land cover only. It is interesting to note that, for the second layer, the rest of the inputs are not significant for any study region.

3.5.4 Comparison With Airborne and In Situ Soil Moisture Data

We also compare the proposed machine learning framework (under spatiotemporal learning) using publicly available high-resolution (airborne) soil moisture data set retrieved during the SMAPVEx 2016 campaign at South Fork, Iowa (Figure 3.5(e)). Due to the poor temporal coverage of the SMAP/Sentinel-1 product, only one day (6 August

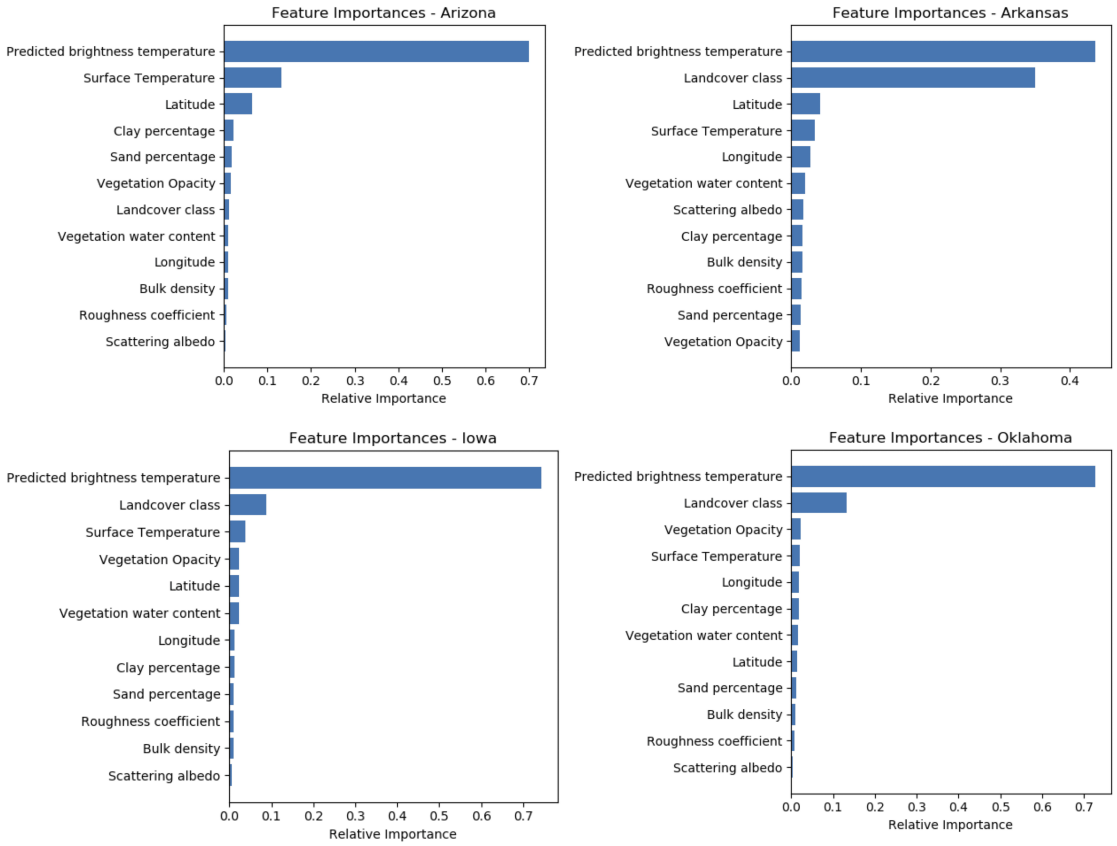

Figure 3.14: For Layer 2 of the machine learning framework, the predicted 3-km brightness temperature from Layer 1 is the most important input followed by surface temperature for Arizona and land cover for Arkansas, Iowa, and Oklahoma. For Arkansas, the effect of land cover is more significant due to the presence of multiple land classes in the state.

2016) of the airborne campaign coincided with the SMAP/Sentinel-1 overpass. Figure 3.15(a) shows the Passive Active L-band System airborne soil moisture data upscaled to the 3-km EASE2 grid. Figure 3.15(b) shows the retrieved 3-km soil moisture from the SMAP/Sentinel-1 product, and Figure 3.15(c) shows the soil moisture predictions from the two-layer framework. The two-layer predictions show good consistence with the airborne soil moisture with correlation value being 0.23 and ubRMSE being 0.032 cm^3/cm^3. These values are similar to the ones between the actual SMAP/Sentinel-1 product and the

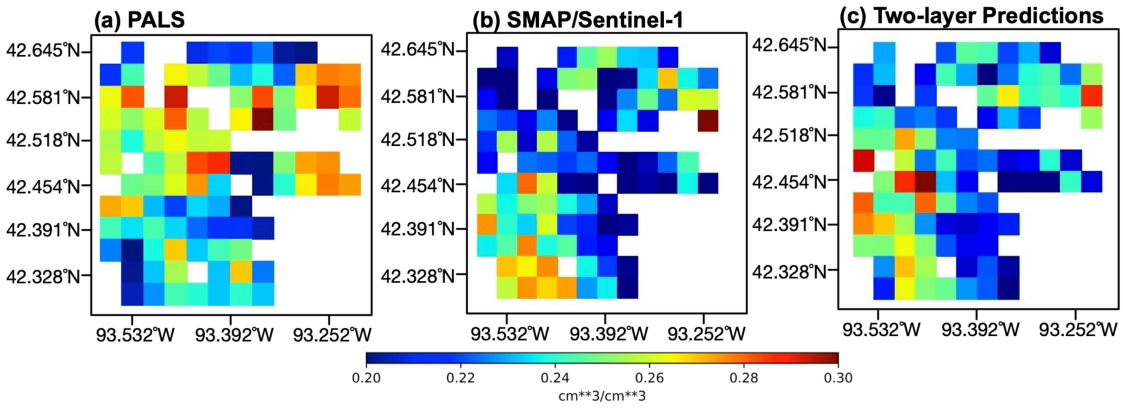

Figure 3.15: Soil moisture comparisons at South Fork, Iowa, on 6 August 2016. (a) Airborne data upscaled to 3-km Equal-Area Scalable Earth-2 grid. (b) The 3-km soil moisture from SMAP/Sentinel-1 active-passive product. (c) Predictions from two-layer model under spatiotemporal regional learning. PALS = Passive Active L-band System; SMAP = Soil Moisture Active Passive.

airborne soil moisture, which has a correlation of 0.09 and ubRMSE of 0.038 cm^3/cm^3. This serves as a good validation of the proposed framework.

To explore the temporal stability of the proposed framework, we further validate the soil-moisture predictions with in situ data from SCAN and USCRN. Figure 3.5 lists the available SCAN and USCRN stations falling into the test regions at the four selected areas (one each for Arizona and Oklahoma, two for Iowa, and three for Arkansas). Figure 3.16 shows the time series of the in situ data, our soil moisture predictions, and the actual SMAP/Sentinel-1 soil moisture values. Note that, for predictions, the entire test region, following Figure 3.5, was held out during the training process. It can be seen that the predictions from our two-layer machine learning framework mimic the majority of the wetting and drying phases of the observed SMAP/Sentinel-1 product and in situ data. The performance of the predictions follows a similar trend as that of the actual SMAP/Sentinel-1 product (Table 3.3).

For instance, compared with USCRN data in Arizona, the correlation of the pre-

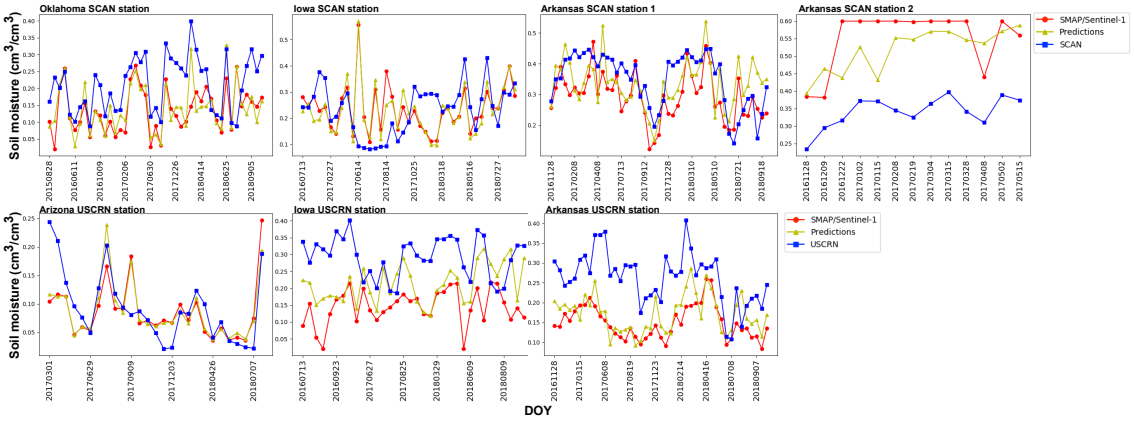

Figure 3.16: Time series matchup of the in situ soil moisture collected by two sparse soil moisture networks, SCAN and USCRN, with the SMAP/Sentinel-1 retrievals and predictions from two-layer machine learning framework under the spatiotemporal regional learning. SMAP = Soil Moisture Active Passive; SCAN = Soil Climate Analysis Network; USCRN = U.S. Climate Reference Network; DOY = day of year.

Table 3.3: Comparison of the SMAP/Sentinel-1 Observations and the Two-Layer Predictions with in-Situ Soil Moisture. SMAP = Soil Moisture Active Passive; SCAN = Soil Climate Analysis Network; USCRN = U.S. Climate Reference Network.

R/ubRMSE(cm^3/cm^3)		SMAP/Sentinel-1 v.s. in situ	Two-layer predictions v.s. in situ	Two-layer predictions v.s. SMAP/Sentinel-1
SCAN	Oklahoma	0.55/0.071	0.61/0.068	0.81/0.042
	Iowa	0.04/0.125	0.10/0.126	0.92/0.037
	Arkansas-1	0.65/0.069	0.47/0.085	0.76/0.053
	Arkansas-2	0.79/0.056	0.68/0.045	0.49/0.075
USCRN	Arizona	0.65/0.047	0.71/0.043	0.93/0.018
	Iowa	0.06/0.077	0.08/0.079	0.38/0.060
	Arkansas	0.52/0.056	0.42/0.062	0.76/0.032

dictions is 0.71 (ubRMSE = 0.047 cm^3/cm^3) when the correlation of the actual SMAP/Sentinel-1 product is 0.65 (ubRMSE = 0.043 cm^3/cm^3), while for Iowa SCAN data, the correlation of predictions drops down to 0.10 (ubRMSE = 0.126 cm^3/cm^3) when the SMAP/Sentinel-1 soil moisture correlation is 0.04 (ubRMSE = 0.125 cm^3/cm^3). This

is a good validation of the proposed framework as our aim is to extend the established SMAP/Sentinel-1 product to unobserved areas (and not improve the SMAP/Seninel-1 retrieval). The efficacy of the predictions for a given surface heterogeneity and hydroclimate is therefore limited by the accuracy of SMAP/Sentinel-1 retrievals in those conditions and future improvements in the retrieval algorithm will enhance the predictive capability of the machine learning algorithm as well.

3.5.5 Regional Learning Limitations

3.5.5.1 Spatial Regional Learning

To compare the downscaling performances of different regional learning settings, fixed-size test regions were adopted for the above experiments at each study area. It is worthwhile to explore whether our proposed two-layer framework can achieve satisfactory performance when the size of the test regions expands. To project how our models will perform when we fill real gaps, we split the training and test regions by shrinking the regions covered by the 3-km soil moisture. Taking the available swath at Arkansas on 14 June 2018 as an example, as shown in Figure 3.17, we split the region covered by the SMAP/Sentinel-1 3-km soil moisture product (Figure 3.17(a)) into training and test regions with various size ratio (Figure 3.17(b)-3.17(e)). Here four different size ratios, 1:4, 1:2, 1:1 and 2:1, are chosen. A higher ratio indicates that less training samples are used during the training process and more samples are used to validate predictions. Models learned using data from the training region are used to downscale soil moisture at test region. Figure 3.18 shows the corresponding results at the three selected study areas. While generally deteriorating performances are observed with the growing size ratio, at Arizona and Arkansas, our proposed two-layer framework outperforms the resampled 33 km soil moisture and the one-layer framework. Especially at Arkansas, the R values are approximately 0.7 when the test region is twice the size of the training region compared

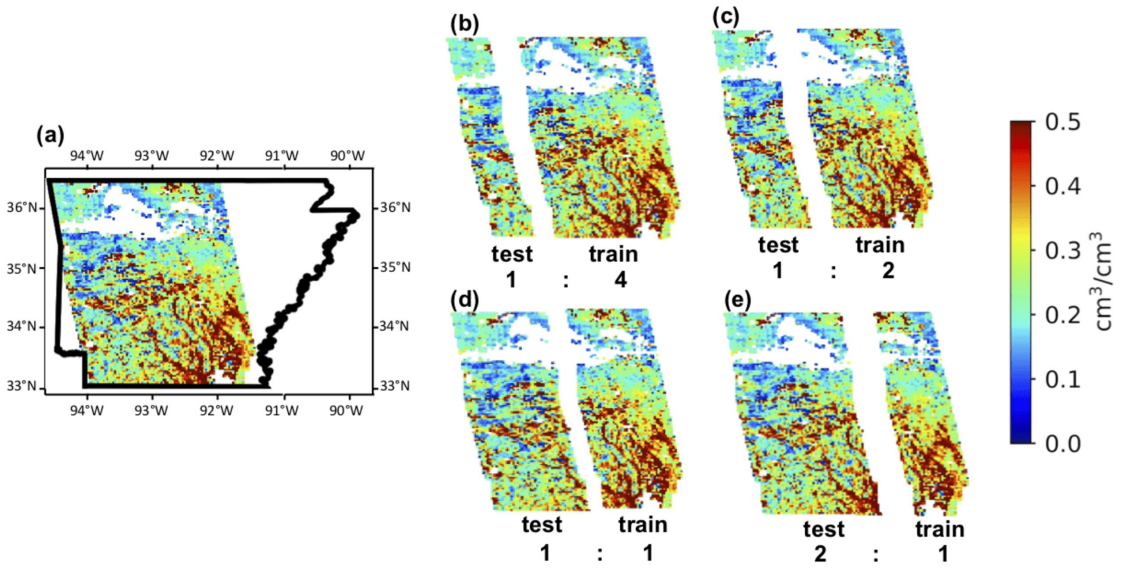

Figure 3.17: Vertical split of the whole swath at Arkansas on 14 June 2018 (a) with various size ratios of test region versus training region. (b), (c), (d), (e) show size ratios of 1:4, 1:2, 1:1, and 2:1, respectively.

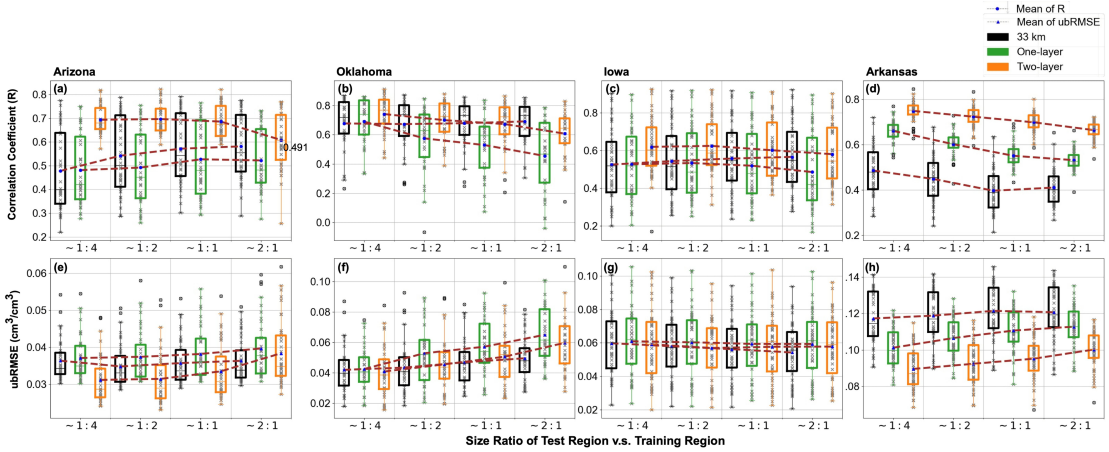

Figure 3.18: Performance comparisons of the resampled 33-km soil moisture, one-layer model, and our proposed two-layer framework under different size ratios of the test region versus the training region. ubRMSE = unbiased root mean square error.

with $\sim$0.4 R from the 33-km soil moisture and $\sim$0.5 R from the one-layer framework.

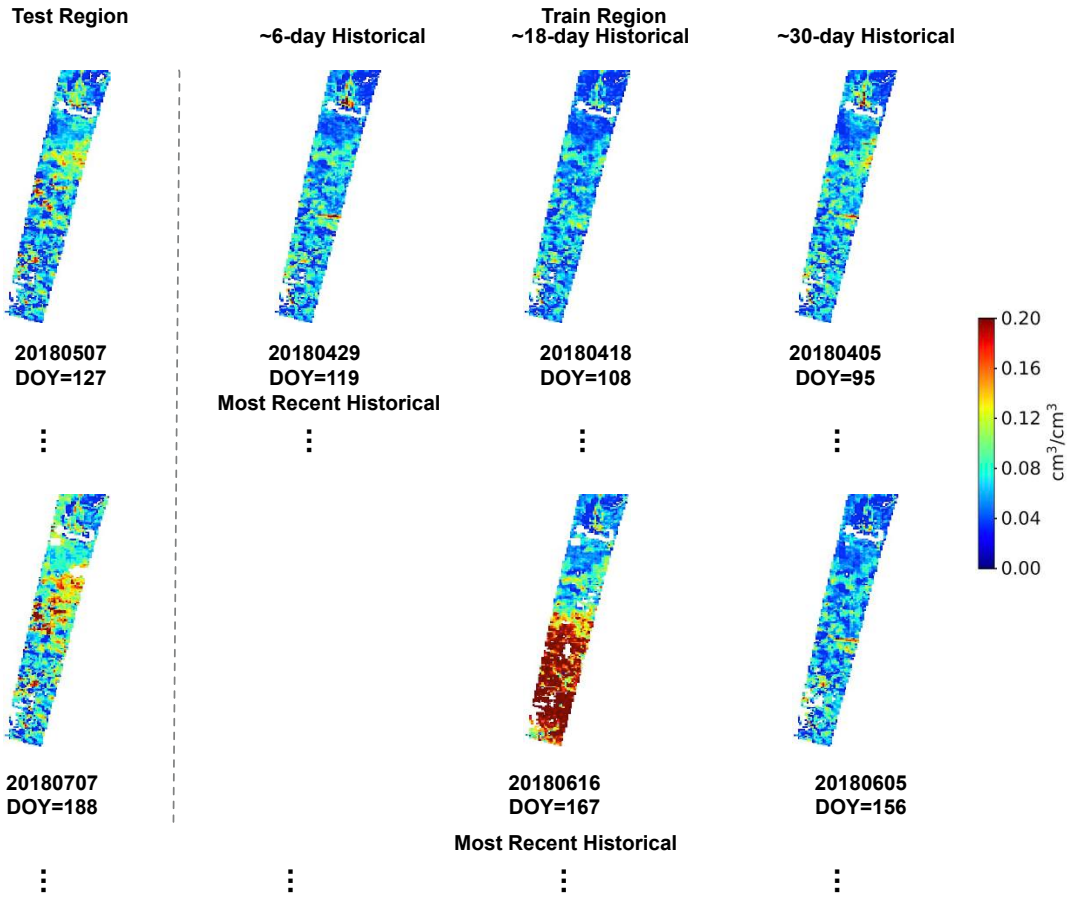

Figure 3.19: Search for temporal neighboring regions within a 30-day window for test regions at Arizona on 7 May 2018 and 7 July 2018. DOY = day of year.

3.5.5.2 Temporal Regional Learning

To obtain the temporal neighboring region for the temporal regional learning and spatiotemporal regional learning, we search back from the current day to find the most recent historical temporal neighboring region from another swath pattern within a 30-day window. In this section, we explore how the timespan between the current day test region and the obtained temporal neighboring region influences the predictive performance. For the Sentinel-1 radars, the temporal resolution of one swath pattern is $\sim$12 days, while the timespan between two different but consecutive swath patterns is $\sim$6 days. Consid-

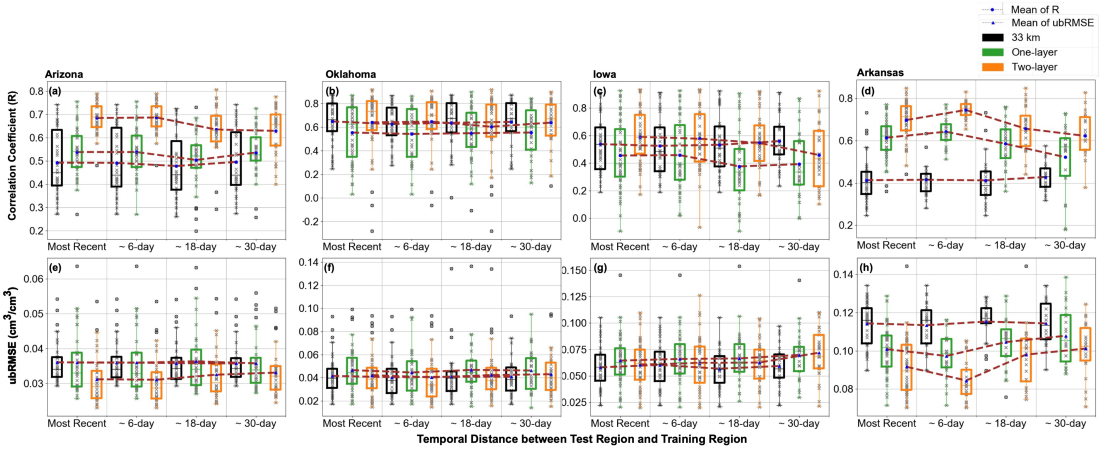

Figure 3.20: Performance comparisons of the resampled 33-km soil moisture, one-layer model, and our proposed two-layer framework under different temporal distances between the test region and the temporal neighboring region. ubRMSE = unbiased root mean square error.

ering that the swaths may be missing due to various reasons, the timespan between the current-day test region and the obtained temporal neighboring region can be $\sim$6 days, $\sim$18 days or $\sim$30 days. To explore how these different timespans influence the downscaling performances of the one-layer framework and our two-layer framework, as shown in Figure 3.19, we continue searching within the 30-day window to find all available temporal neighboring regions and compare the results aggregated by different timespans. Figure 3.20 shows the comparisons at the three study areas. All the three study areas observe deteriorating performances, to different extents, as the timespan increases between the test region and the temporal neighboring region. Again, our two-layer framework returns better downscaling performances, especially at Arizona and Arkansas, compared with the one-layer framework and the resampled 33-km soil moisture.

3.5.6 Prediction at Gap Areas

In this subsection, we predict and compare the downscaled soil moisture produced by the models at real gap areas where the 3-km soil moisture estimates are indeed missing for

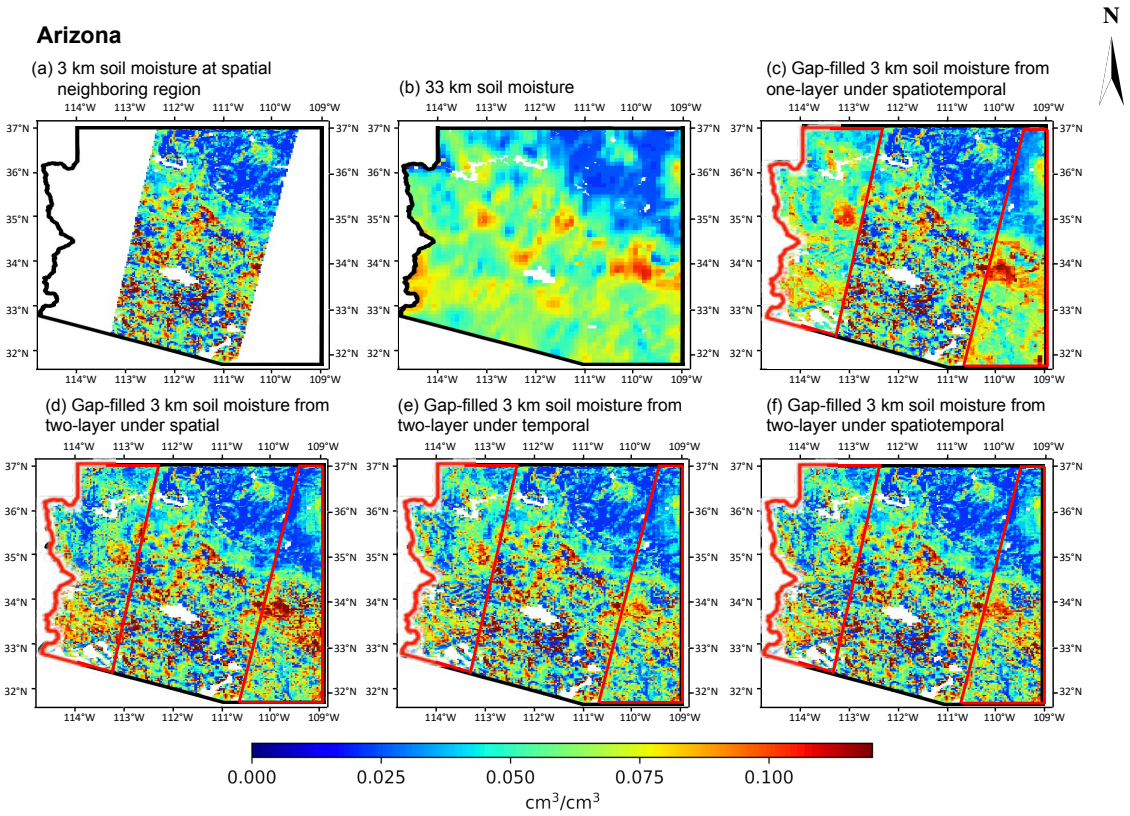

Figure 3.21: Real gap filling at Arizona on 13 June 2018. (a) The 3-km soil moisture from Soil Moisture Active Passive/Sentinel Product for spatial neighboring region on the same day. (b) Resampled 33-km soil moisture from Soil Moisture Active Passive L3_SM_P_E. Gap-filled 3-km soil moisture from (c) the one-layer model using spatiotemporal regional learning, (d) the hierarchical learning models using spatial regional learning, (e) the hierarchical learning models using temporal regional learning, and (f) the hierarchical learning models using spatiotemporal regional learning.

the SMAP/Sentinel-1 product. As we have discussed in section 3.4.5, we first train models using data from spatiotemporal neighboring regions of the gap regions. The learned models are then fitted to predict the 3-km brightness temperature and soil moisture at real gap areas. Figure 3.21 shows the predicted high-resolution soil moisture at Arizona on 13 June 2018, as an example, from the resampled 33-km soil moisture, the one-layer framework, and the two-layer framework under different regional learning settings. Again, our two-layer framework reconstructs more subgrid soil moisture heterogeneities compared

with the one-layer framework.

Gaps of SMAP/Sentinel-1 active-passive product for Contiguous United States can be filled adopting the same methodology. The gap-filled soil moisture has the same 3-km resolution as the SMAP/Sentinel-1 active-passive product and the same spatial and temporal coverage as the SMAP-enhanced passive product. Figures 3.22(a) and 3.22(b) show the daily 3-km soil moisture product from the SMAP/Sentinel-1 active passive product and daily resampled 33-km soil moisture product from the SMAP-enhanced passive product, on 16 June 2018, respectively. Large areas that are not covered by the fine-scale soil moisture can be observed. Again, this is due to missing high-resolution radar backscatter measurements. Figures 3.22(c) and 3.22(d) show the comparison of the gap-filled soil moisture produced by the one-layer framework and our proposed two-layer framework. We were able to gap fill most of the missing areas except those that have missing inputs, mainly the historical radar scatter measurements. Additionally, while the one-layer framework is able to reconstruct the 3-km soil moisture (under the spatiotemporal regional learning), it produces much less fine-scale soil moisture heterogeneity than proposed two-layer framework.

3.6 Conclusion

Recognizing the importance of global-scale soil moisture estimates with fine spatial and temporal resolution, this chapter introduces the regional learning strategy to enable regional gap-filled high-resolution soil moisture that maintains high consistency with the SMAP/Sentinel-1 3-km soil moisture product. It also presents a novel two-layer machine learning-based framework motivated by the SMAP active-passive retrieval algorithm. The two-layer framework can downscale/gap-fill soil moisture with more heterogeneity captured at areas where the SMAP/Sentinel-1 active-passive 3-km soil moisture product is missing. The average of historical radar backscatter measurements is introduced as one of

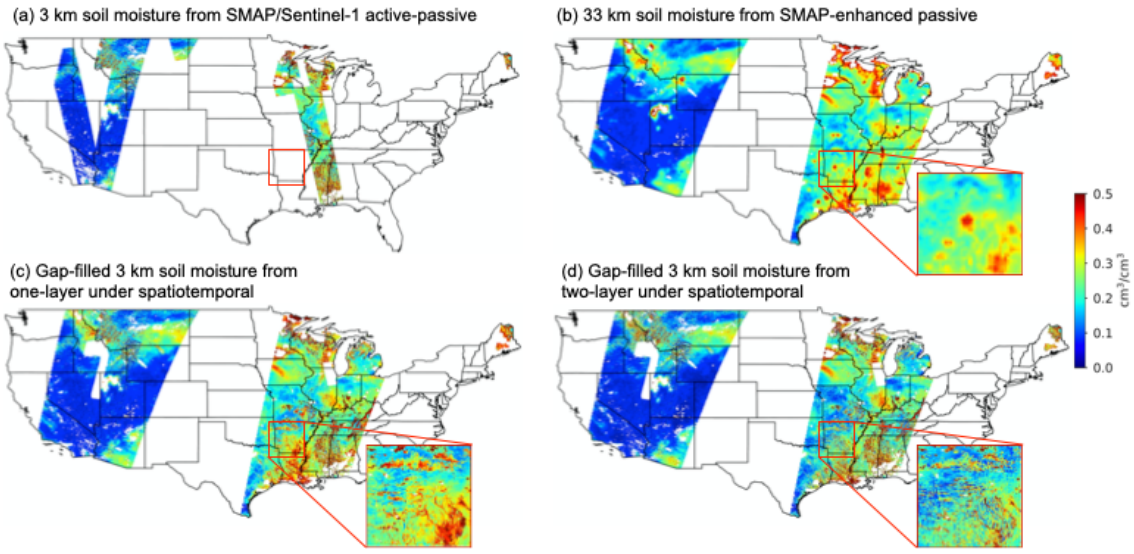

Figure 3.22: Real gap filling of the Continental United States on 16 June 2018. (a) Observed 3-km soil moisture from SMAP L2_SM_SP. (b) Resampled 33-km soil moisture from SMAP L3_SM_P_E. (c) Gap-filled 3-km soil moisture from the one-layer model under the spatiotemporal regional learning setting. (d) Gap-filled 3-km soil moisture from the proposed theory-guided hierarchical framework under the spatiotemporal regional learning setting. SMAP = Soil Moisture Active Passive.

the key inputs for soil moisture downscaling. We compare the two-layer framework with a conventional one-layer framework and the resampled 33-km soil moisture product under three different regional learning settings against holdout SMAP/Sentinel-1 active-passive soil moisture estimates. We show that our two-layer framework outperforms the one-layer model and the resampled 33-km soil moisture, in terms of R and ubRMSE, under all scenarios. The most significant performance improvements of the proposed two-layer framework were seen on the R under the spatiotemporal regional learning where it is increased to 0.728 at Arizona compared to 0.493 for the resampled 33 km and 0.572 for the one-layer, 0.742 at Oklahoma compared to 0.649 for the resampled 33 km and 0.679 for the one-layer, 0.649 at Iowa compared to 0.537 for the resampled 33 km and 0.556 for the one-layer, and 0.744 at Arkansas compared to 0.413 for the resampled 33 km and 0.651 for the one-layer.

One limitation of our gap-filled soil moisture product is that its availability is still restricted by the availability and accuracy of the inputs, mainly the 33-km soil moisture and 33-km brightness temperature from the SMAP-enhanced passive product. The best temporal resolution produced is therefore $\sim$2-3 days. This issue will hopefully be addressed by the generation of seamlessly covered coarse-scale soil moisture and/or brightness temperature data in the future. This can be achieved by developing gap-filling techniques for the 33-km SMAP soil moisture or by fusing multiple soil moisture products with SMAP.

4. CONTEXT-AWARE ATTENTIVE REPRESENTATION LEARNING FOR GEO-SPATIOTEMPORAL ANALYSIS

4.1 Introduction

Environmental monitoring is critical to risk control, policymaking, and resource protection and management [126, 127], in order to understand the effects both of systematic factors such as increasing population [33] and occurrences of dynamic or sporadic phenomena such as extreme weather events [128]. With the emergence of new remote sensing technologies deployed in ubiquitous monitoring platforms, we are gaining the ability to collect vast amounts of geo-spatiotemporal environmental data [129].

Recent years have seen a proliferation of studies concerning computational approaches that exploit geo-spatiotemporal data for applications to environmental sustainability. These applications include crop yield prediction [13], soil moisture downscaling [130], land cover classification [15], wildfire prediction [4], and climate modeling [131]. The input data for these applications are either multi-channel data from a single source (e.g. multispectral remote sensing images) or conflated geodata fused from various sources [132]. Combining multi-feature spatial data in time series results in a high-dimensional dataset. It comprises the data features associated with each represented point in the four-dimensional product space of temporal and spatial coordinates. As an example, Figure 4.1 shows the available input of a midwestern county in the U.S. for the county-level crop-yield prediction task, which consists of monthly plant growth estimates from remote sensing satellites and climate data from various sources collected during the growing season.

Extracting information from the high-dimensional data remains a challenge. Most previous studies have avoided dealing with the full spatiotemporal data directly through

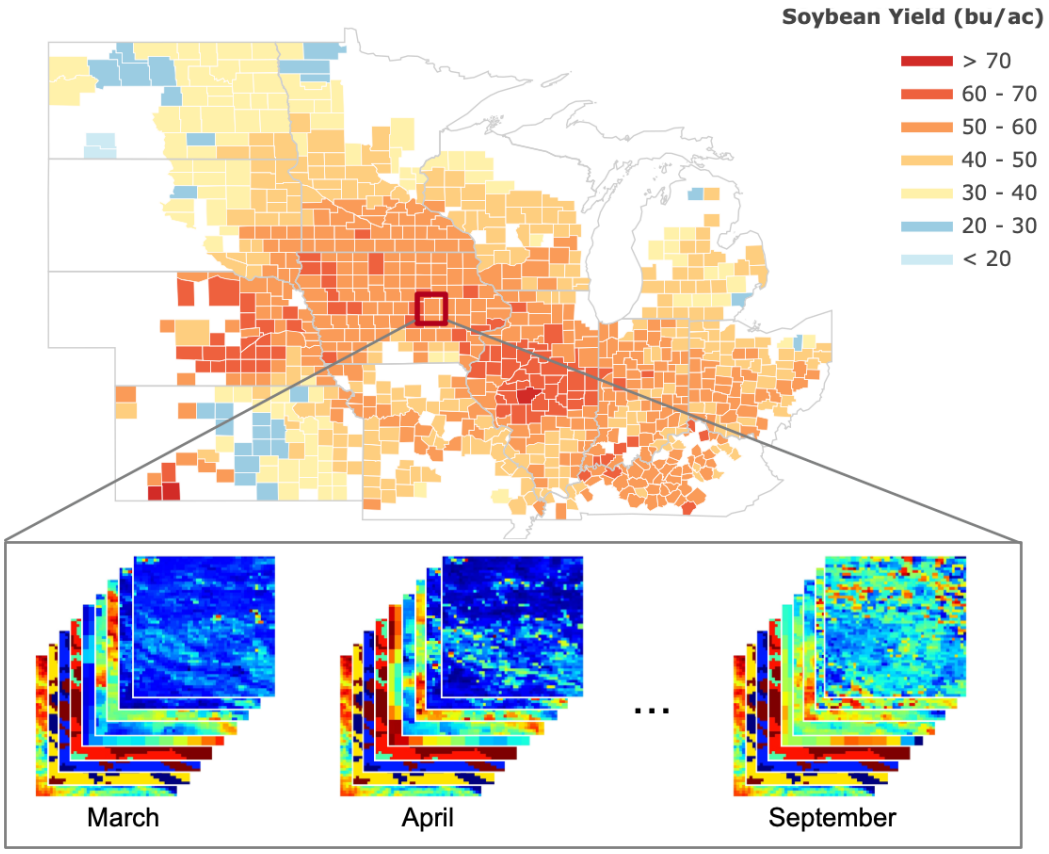

Figure 4.1: Up: County-level soybean yield of mid-western U.S. in 2017. Down: monthly plant growth estimates (including vegetation indices) and climate data (including land surface temperature, precipitation, soil properties and elevation) at Tama, Iowa.

various dimensionality reduction techniques. The most common approach is to spatially average the input resulting in time series scalar features [34]. In [13], spatial images are converted to histograms with the assumption that spatial distributions of the features do not contribute to the accuracy of the predictions. These approaches ignore the spatially explicit heterogeneity and interactions existing among the fused environmental features. The ignored subgrid heterogeneous dynamics, however, play an important role in many sustainability applications, such as crop yield prediction [30] and soil moisture downscaling [130]. In addition, there are studies that generate predictions from a one-time snapshot of the geo-spatiotemporal data [39, 41] without taking advantage of the temporal infor-

mation.

Another challenge that commonly exists in the applications of spatiotemporal geographical data is the lack of labels [15]. Many computational sustainability tasks rely on label collection procedures that are either expensive or cumbersome [133]. The scarcity of labeled training data becomes an even severe issue for machine learning models that deal with high-dimensional data [134].

In this chapter, we propose a novel semi-supervised self-attentive-based model that *jointly* learns a prediction function and an embedding representation function. At each time step, spatial information contained in the multi-channel three-dimensional images is first extracted to a latent space as representations through convolutional neural networks. A self-attentive model is then applied to temporally aggregate the spatial representations at all time steps to generate a global spatiotemporal representation, from which final predictions are produced. To overcome the scarcity of training data, we further introduce the spatial and temporal coherence signals that naturally reside in the largely-available unlabeled geo-spatiotemporal data through a novel sampling procedure. Similar to the proximity-based word embedding models in natural language, we make an assumption that images that are spatially or temporally close should have similar representation, and conversely. Under our semi-supervised training framework, this newly introduced spatiotemporal context can aid meaningful representation learning that adapts to the supervised prediction task at the same time. We evaluate our approach through a large-scale real-world crop yield prediction task. Experimental results show that our semi-supervised self-attentive model outperforms existing state-of-the-art crop yield prediction methods and its counterpart, the supervised-only self-attentive model, especially when there are less labeled training data. The contributions of this work are summarized as follows:

- We propose a novel semi-supervised self-attentive model which learns global spatiotemporal representations for prediction tasks. This model takes full advantage of

the multi-channel geo-spatiotemporal data with both spatial and temporal variations captured.

- We aid the representation learning with spatial and temporal context sampled from largely-available unlabeled geo-spatiotemporal data. Under the semi-supervised training framework, context-aware representations that adapt to the supervised prediction task are learned.

- We conduct extensive experiments on a challenging large-scale real-world crop yield prediction task. Experimental results demonstrate the effectiveness of our semi-supervised self-attentive model over existing state-of-the-art yield prediction methods and its counterpart, supervised-only model, especially under the stress of scarce labeled data.

4.2 Related Work

With the emergence of environmental monitoring in the past decade, great progress has been made in many computational sustainability tasks through the use of geo-spatiotemporal data [135]. The most common approach to prediction in this area, either as regression or classification, is to incorporate multiple features from various sources with the consideration of domain knowledge. Specifically, much research [3, 130, 34] take time series scalar features as input to generate predictions for ground-truth labels within a predefined spatial unit. Input features with higher resolutions than the labels are averaged without considering the subgrid heterogeneous dynamics. However, it has been studied and well acknowledged in many environmental sustainability domains that the spatially explicit dynamics and interactions among environmental factors play a great role determining the characteristics of environmental factors at larger spatial scales [30, 29]. Ignoring these subgrid heterogeneity results in unnecessary information loss. In [13], multispectral remote sensing images with spatial distributions are converted to histograms

with the assumption of permutation invariance. Their assumption is less appropriate when conflated geodata with spatially explicit dynamics and interactions are used as inputs. In addition, several studies only incorporates static spatially distributed data without considering the temporal patterns in the geo-spatiotemporal data, such as land cover prediction [136], poverty mapping [15]. There is however valuable information contained in the temporal variations that can aid the predictions [3, 30].

The temporal self-attentive representation learning model proposed in this chapter, instead, adopts a hierarchical framework where spatial representations at each time step are learned first, followed by a temporal aggregation through a self-attentive model. It extracts both spatial and temporal variations in the data with a global spatiotemporal representation learned for the prediction task. Recurrent convolutional networks [137] shares a similar architecture with our model, but processes latent output at each time step in a temporally explicit order. While the self-attentive model has demonstrated its superiority in capturing global information and handling sequential data at longer lengths in many fields [138, 139], there is no research, to our best knowledge, to demonstrate its ability to mine spatiotemporal patterns from geographical data. Geo-spatiotemporal data normally show distinct but relatively consistent temporal patterns [30]. The knowledge transfer among different time steps enabled by the self-attentive model can aid the generation of a more meaningful global representation. The 3d convolutional neural networks [59] also aims at spatiotemporal feature learning. They apply 3d convolutions across both the spatial and temporal dimensions.

Another remaining challenge for learning spatiotemporal environmental monitoring data is the lack of labels. In addition to data augmentation [136, 40] and transfer learning [45] that have been proposed to alleviate this issue, there have been attempts to introduce spatial context to aid model learning [140, 15]. In [140], weighted representations of all spatial neighbors within a predefined region are considered. However, their settings

are not practical when dealing with high-dimensional spatiotemporal data. An existing study [15] introduces the spatial context through sampling from the neighboring region to aid their unsupervised representation learning, and learn representations independently from downstream tasks. In our research, we sample spatial and temporal context from the largely-available unlabeled data. Information from the sampled context is learned together with a supervised prediction task under the semi-supervised training framework. Context-aware representations that adapt to the prediction task at the same time can thus be learned.

4.3 Method

4.3.1 Problem Definition

In this section, we present the formulation for the geo-spatiotemporal unit-wise prediction task. Specifically, given the multi-channel geo-spatiotemporal data as input, the goal is to predict labels for each geographical unit. This approach can be used in many real-world applications, such as land cover classification [3] and wildfire prediction [4] in field-scale, crop yield prediction in county-level [13], and health index prediction for countries [15]. Given a unit i, there are T different multi-channel spatial images centered at the unit. Different images represent the geo-spatiotemporal data for different time steps. We denote the time series spatial images as $\mathbf{A}_i \in \mathbb{R}^{T \times h \times w \times d}$. Note that h and w denote the height and width of the image, and d is the number of features incorporated from either a single dataset or a conflated dataset fused from various sources. To utilize the spatial and temporal context information, for each unit i, we obtain three context-aware images, known as the spatial neighbor $\mathbf{SN}_i$, spatial distant $\mathbf{SD}_i$, and temporal neighbor $\mathbf{TN}_i$ through a novel sampling procedure, which is introduced in Section 4.3.3. These context-aware images share the same dimensions $T \times h \times w \times d$ as $\mathbf{A}_i$.

Our objective is to learn a predictive model to predict the label for unit i from the time

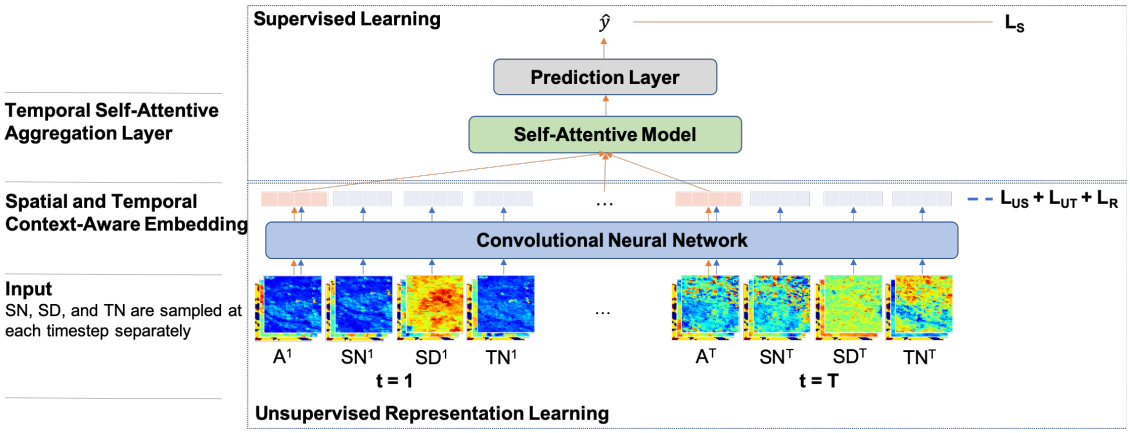

Figure 4.2: Architecture overview of the semi-supervised context-aware attentive representation learning model. Arrows in orange and blue denote the data flows for the supervised and unsupervised learning, respectively.

series spatial image quadruplets. Formally, it can be written as

$$y_i = \mathcal{F}\big(\{\mathbf{A}_i, \mathbf{SN}_i, \mathbf{SD}_i, \mathbf{TN}_i\}^1, \cdots, \{\mathbf{A}_i, \mathbf{SN}_i, \mathbf{SD}_i, \mathbf{TN}_i\}^T\big), \tag{4.1}$$

where y_i denotes the prediction for the geographical unit i.

4.3.2 Framework Overview

In this chapter we propose a novel semi-supervised model that jointly learns a prediction function and an embedding representation function. The architecture of the proposed framework is presented in Figure 4.2. The framework consists of three parts. First, for each time step, we employ convolutional neural networks (CNNs) to extract its spatial representations from the image quadruplets $\{\mathbf{A}, \mathbf{SN}, \mathbf{SD}, \mathbf{TN}\}$. Since CNNs have demonstrated its superior ability to extract spatial information in computer vision field [62], we adopt CNNs to extract spatially explicit dynamics and interactions existing among the fused environmental features. Specifically, we use a modified ResNet-18 architecture [5] where the final classification layer is removed. Formally, for each im-

age in the quadruplets, the CNNs map it to an m-dimensional embedding, denoted as $g(\cdot) \in \mathbb{R}^m$. Then we denote the resulted time series context-aware spatial representations as $\{g(\mathbf{A}), g(\mathbf{SN}), g(\mathbf{SD}), g(\mathbf{TN})\}^t$. Note that for all different time steps and images, the same CNNs are shared. Second, a self-attentive model is applied to temporally aggregate the spatial representations $g(\mathbf{A})^t$ at the labeled regions to produce a global spatiotemporal representation. Finally, a prediction layer composed of a fully-connected neural network is deployed to generate predictions from the extracted global information. The objective function of our model is composed of four parts:

$$L := L_S + \alpha(L_{US} + \beta L_{UT} + \gamma L_R), \tag{4.2}$$

where L_S denotes the supervised loss for the prediction task. L_{US} and L_{UT} denote unsupervised losses for spatial and temporal context respectively. L_R is added here to regularize the learned representations to a meaningful hypersphere. α, β, γ are trade-off weights. We first explain how we produce spatial and temporal context-aware representations through unsupervised learning and a novel sampling procedure in Section 4.3.3. Descriptions of the temporal self-attentive model for the generation of a global spatiotemporal representation are presented in Section 4.3.4.

4.3.3 Context-aware Representation Learning

The widely explored proximity-based word embedding models [67] assume that "a word is characterized by the company it keeps"; thus words that appear in similar contexts should have similar representations. Extending this idea to our spatiotemporal distributed geographical data, we make an assumption that spatial images that are close spatially or temporally should have similar representations than those that are far apart.

To constrain the closeness in spatiotemporal space, we introduce the concepts of *spatial neighborhood* and *temporal neighborhood*. First, we denote the regions where la-

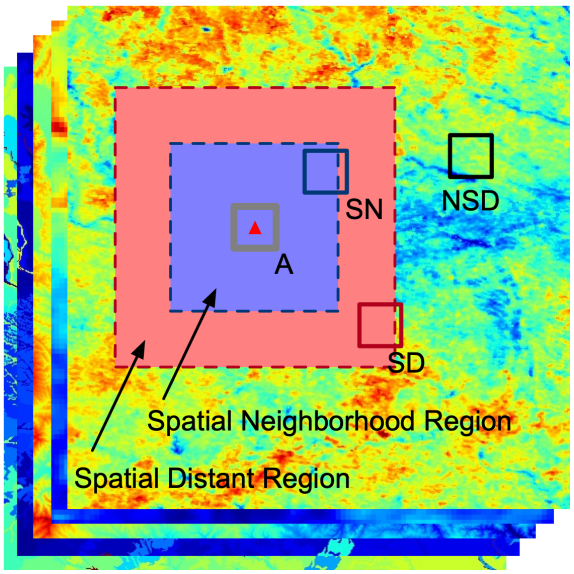

Figure 4.3: An example of spatial neighborhood region and spatial distant region corresponding to an anchor image $\mathbf{A}$. Spatial neighbor $\mathbf{SN}$ is sampled from the spatial neighborhood region. Spatial distant $\mathbf{SD}$ is sampled from the spatial distant region. A conflated geodata that fuse various environmental features, such as vegetation indices, temperature, soil properties, etc., are plotted in background.

bels are obtained and predictions ought to be produced as anchor regions. The spatial neighborhood is then defined as a larger spatial region that is within a predefined spatial distance of an anchor region and appears at the same time step. An example of a spatial neighborhood region is shown in Figure 4.3. The temporal neighborhood, instead, is defined as the same spatial region of the anchor region but appears at a time step within a predefined temporal distance. Spatial neighbor image $\mathbf{SN}$, spatial distant image $\mathbf{SD}$, and temporal neighbor image $\mathbf{TN}$ can be obtained in the context of the spatial neighborhood and temporal neighborhood through a sampling procedure, which will be introduced in 4.3.3.2.

4.3.3.1 Learning Objective

Now suppose that we have time series spatial representations $\{g(\mathbf{A}), g(\mathbf{SN}), g(\mathbf{SD}), g(\mathbf{TN})\}^t$ generated from the CNNs with $\{\mathbf{A}, \mathbf{SN}, \mathbf{SD}, \mathbf{TN}\}^t$

as input. At each time step $t \in (1 \cdots T)$, we seek to minimize the Euclidean distance between the representation vectors of the anchor image and spatial/temporal neighbor image, while maximize the distance between the representation vectors of the anchor image and spatial distant image. The unsupervised loss for the spatial and temporal context can be calculated by

$$L_{US}^{t} = max(0, ||g_\theta(A^t) - g_\theta(SN^t)||^2 - ||g_\theta(A^t) - g_\theta(SD^t)||^2 + p) \tag{4.3}$$

and

$$L_{UT}^{t} = ||g_\theta(A^t) - g_\theta(TN^t)||^2, \tag{4.4}$$

respectively. Following existing work [15], a rectifier with margin p is introduced here to control the extent how the representations of the spatial distant image are pushed away compared to the representation of the spatial neighbor image.

To constrain the learned embeddings within a hypersphere where better representations with meaningful relative distance can be learned, we further introduce a L2 regularization with loss L_R^t:

$$L_{R}^{t} = ||g_\theta(A^t)||^2 + ||g_\theta(SN^t)||^2 + ||g_\theta(SD^t)||^2 + ||g_\theta(TN^t)||^2. \tag{4.5}$$

Finally, for given a dataset of N geographical regions with T time steps, the unsupervised loss is given as

$$\min_{\theta} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} (L_{US(i)}^{t} + \beta L_{UT(i)}^{t} + \gamma \frac{L_{R(i)}^{t}}{\sqrt{m}}) \tag{4.6}$$

where m is the dimension of the embedding vectors and θ is the parameters of the CNNs. This unsupervised loss is jointly trained with the supervised prediction task to generate context-aware representations that are suitable for the prediction task [141].

4.3.3.2 Quadruplet Sampling

We adopt the following procedure to generate times-series image quadruplets with spatial and temporal context. First, anchor images $\mathbf{A}$ are collected as images that cover the area of interest with fixed image size. Second, spatial neighbor $\mathbf{SN}$ and spatial distant $\mathbf{SD}$ images are sampled with respect to the anchor images based on spatial distance. We adopt a similar sampling procedure as in [15] with the spatial neighborhood introduced to constrain the sampling of spatial neighbor images. Specifically, the center of the spatial neighbor images must be within a predefined number of pixels of the anchor image center both vertically and horizontally. Different from the pure-unsupervised learning in [15] where the size of spatial neighborhood is relatively more flexible, we find in our setting that the size of spatial neighborhood should adapt to the supervised task. In practice, we find that choosing the spatial neighborhood similar to the size of the anchor images produces better supervised predictions.

As computational sustainability tasks normally conduct experiments in large-scale, e.g. the county-level crop yield prediction across the United States [13], we further constrain the distant region from which the spatial distant images can be sampled. As shown in Figure 4.3, in addition to the spatial neighborhood region colored in blue, there is a spatial distant region colored in red around the anchor image $\mathbf{A}$. The image $\mathbf{NSD}$ in the black rectangle, while is also far away from $\mathbf{A}$ as the sampled distant image $\mathbf{SD}$, will not be considered as spatial distant in this setting. This is similar to the "hard negative" idea introduced to the temporal embedding learning in [142]. We find in practice that this newly added constraint help us learn better representations with the prediction accuracy improved. Finally, the temporal neighbor images are sampled from a fixed temporal window as temporal context for each anchor image. The sampling procedure is applied at all time steps separately to enrich the spatial and temporal proximity signals the model sees.

4.3.4 Temporal Self-Attentive Model

We present here a temporal self-attentive model used for the generation of a global spatiotemporal representation. As output from the CNNs, we have the time series context-aware spatial representations $(g(\mathbf{A})^1, g(\mathbf{A})^2, \cdots, g(\mathbf{A})^T)$ for the labeled regions. As these representations are generated independently at each time step, there is no order information learned to aid networks in later stage understanding the relative distance of these embeddings. To incorporate order of sequence information, we add "positional embedding" [143] to the spatial representations through position-wise summation. Specifically, we use the positional encoding in a sinusoidal form. The encoding added to a spatial representation at time step t is given as:

$$\begin{aligned} PE(t, 2i) &= sin(t/10000^{2i/m}) \\ PE(t, 2i+1) &= cos(t/10000^{2i/m}) \end{aligned} \tag{4.7}$$

where m is the length of the representation embedding and $i \in [0, m/2]$. The added parameter-free position-dependent embeddings can help followed networks incorporate the temporal order of the spatial representations without incurring extra computation burden.

After the adding of the positional encoding, we adopt the multi-head self-attentive model [143] for temporal aggregation. First, we reformulate the time series position-dependent spatial representations in a matrix form $\mathbf{Z} \in \mathbb{R}^{T \times m}$. This representation matrix is then linearly projected to three matrices, queries ($\mathbf{Q}$), keys ($\mathbf{K}$), and values ($\mathbf{V}$), independently. To empower learning in different representation subspaces, the three matrices are further split into h parts and attended by h parallel heads separately. For each head

$i \in h$, we have its self-attention score calculated through the scaled dot-product attention:

$$\mathbf{S}_i = \text{softmax}(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{m/h}}), \tag{4.8}$$

where $\mathbf{S}_i \in \mathbb{R}^{T \times T}$ is the self-attention score matrix which denotes how information contained in representation subspace $\mathbf{V}_i$ can transfer to each other. The scale factor $1/\sqrt{m/h}$ is added to prevent the gradient vanishing problem [144]. The updated spatial representations of i-th head is then obtained by multiplying the self-attention score matrix with the values matrix $\mathbf{V}_i$:

$$\mathbf{Z}_i^u = \mathbf{S}_i \mathbf{V}_i \tag{4.9}$$

where $\mathbf{Z}_i^u \in \mathbb{R}^{T \times m/h}$ is the refined subspace representations. With this update, spatial representation $(\mathbf{z}_i^t)^u \in \mathbb{R}^{m/h}$ at a time step t can incorporate information contained in other times steps resulting in meaningful knowledge transferring.

To generate a unifying representation from all subspaces, the updated matrices from h parallel heads are concatenated and once again projected to have:

$$\mathbf{Z}^u = concat(\mathbf{Z}_1, \cdots, \mathbf{Z}_h)\mathbf{W} \tag{4.10}$$

where $\mathbf{Z}^u \in \mathbb{R}^{T \times m}$ and $\mathbf{W} \in \mathbb{R}^{m \times m}$ is a projection matrix for the concatenated unifying matrix. Following [143], we use a residual connection [5] to add the original spatial representations to the refined representations to enable the propagation of useful features learned at low-level to deeper levels:

$$\mathbf{Z}^u = \mathbf{Z}^u + \mathbf{Z}. \tag{4.11}$$

To further increase the length of the representations, and thus improve the model's

expressive capability [145], we add position-wise feed-forward network as in [143] to process the updated representations position-wisely. Finally, the refined spatial representations at all time steps are temporally averaged through a pooling layer to generate a global spatiotemporal representation, after which a fully-connected neural network is deployed to generate the final prediction. For a regression task, as an example, the supervised loss for the predictions can be calculated through the mean squared error:

$$L_S = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.12}$$

where $\hat{y}_i$ denotes the prediction from the model for region i. We find in practice that this temporal self-attentive model is easy to train and less prone to overfitting compared with previously studied models for geo-spatiotemporal data, especially when the training data is scarce. It becomes even more powerful when accompanied with the context-aware unsupervised representation learning.

4.4 Experiments

To validate the effectiveness of our proposed approaches, we conduct experiments over a large-scale real-world dataset for the crop yield prediction task. The dataset consist of conflated geospatiotemporal data from various sources. In this section, we first introduce the dataset, including the crop yield dataset and the input datasets. Then we describe the evaluation metrics and baseline methods. Results of model performance are presented last. Codes for the models and data pre-processing will be released to public for reproducibility.

4.4.1 Dataset

County-level soybean yield prediction has been an important task and actively researched in previous studies [146, 13]. The ground-truth of the task is average county-

level soybean yields harvested in every October and November. We collect the data from the USDA National Agricultural Statistics Service (NASS) Quick Stats Database [147] for years between 2003 to 2018. 12 states of the midwestern United States are selected which account for over 80% national soybean production. There are around 850 data points per year.

We fuse plant growth estimates from remote sensing and environmental factors from various sources as inputs. Specifically, a pair of monthly vegetation indices, the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI), are collected from the MODIS satellite product MOD13A3 [148] at 1 km resolution. These two vegetation indices complement each other and have been widely used to monitor plant growth in previous studies [34]. For environmental factors, we consider precipitation and surface temperature to reflect the water and heat stress, to which the crop growing processes were observed to be highly sensitive [149]. Monthly precipitation data at 4 km resolution are from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) dataset where precipitation data are derived from nearly 13,000 stations using climate-based interpolation [49]. Daily surface temperature data during the day and night are collected from the 1 km MODIS satellite product MOD11A1 [119]. They are aggregated to a monthly time step through taking averaged. We also introduce geographically localized and time-invariant factors such as elevation and soil properties, e.g., soil sand, silt, and clay fractions. They are also important environmental factors to be considered, as they determine how stresses like water and heat influence crop growth. The elevation data is obtained from the NASA Shuttle Radar Topography Mission Global 30 m product [150] while the soil properties data are collected from 1 km Soil Geographic (STATSGO) Data Base [115]. We map all the inputs from various sources to the MODIS product grid at 1 km resolution through averaging or nearest-neighbor search.

Once the data preprocessing steps are finished, anchor images are cropped as 50×50

pixels centering at the counties which have soybean yields, considering that the average size of counties in the U.S. is around 2500 km^2. MODIS landcover product MCD12Q1 is further introduced in this step to decide the center of croplands at each county. Fixing the image size as 50 $\times$ 50 for all the counties might not be optimal but is practical as we are conducting the prediction task in large scale. Spatial and temporal contextual images are sampled corresponding to the anchor image with the same spatial size. We collect before-harvest input data from March to September, resulting in a feature space $\mathbf{X} \in \{\mathbf{A}, \mathbf{SN}, \mathbf{SD}, \mathbf{TN}\}$with four dimensions, $\mathbf{X} \in \mathbb{R}^{7\times50\times50\times9}$, for each county where the number of time steps is 7 and the number of features is 9.

Note that we do not apply masks to differentiate the land cover types or crop types inside the geographical regions. Instead, we rely on the model itself to extract necessary information from the sequences of raw images to make accurate crop yield prediction. Another caveat is that the fused heterogeneous geographic data originally have different spatial/temporal resolutions, noise sources (e.g., cloud, urban agglomerations, etc.) and data acquisition errors. This also poses challenges to our model to extract meaningful spatiotemporal signals for the prediction task.

4.4.2 Baseline Methods

The baseline methods that are compared with can be classified into three categories based on the dimensions of the input data. The first category includes ridge regression (LR) [34], random forest (RF) [37], and multilayer perceptron (MLP) [41]. These are conventional machine learning models that have been widely used in crop yield prediction tasks. The inputs they take are scalar features in time series which can be obtained by spatially averaging the features at each time step.

The second category includes four deep learning models introduced in [13], LSTM, LSTM+GP, CNN, and CNN+GP. Instead of taking average of features in spatial space,

they convert the spatial image to histogram to keep the frequency distribution which results in a three-dimensional input space. The input is then fed into a LSTM or CNN for information extraction. Additionally, they adopt a Gaussian process after the prediction of deep models to alleviate the spatial correlations in the prediction errors.

The third category includes two advanced deep learning models, 3d convolutional neural networks (C3D) [60, 66] and recurrent convolutional networks (CNN-LSTM) [137]. We adopt a similar architecture for the 3d convolutional neural networks as in [66], with minor modifications to adapt to the height and width of our input. CNN-LSTM deploys the same convolutional networks as our proposed models, but uses the LSTM model, instead of self-attentive model, for temporal feature learning. Both CNN-LSTM and C3D take the four-dimensional features as input.

Additional feature standardization processes, including subtraction of mean and division of standard deviation, are applied to the input data of LR and MLP. As for the deep models, LSTM, LSTM+GP, CNN, and CNN+GP, raw features are fed to generate histograms, after which subtraction of mean is applied as in [13]. The same standardization process as our proposed models is applied to the input taken by CNN-LSTM and C3D, which will be introduced in 4.4.3.

4.4.3 Our Approaches

We evaluate both the semi-supervised self-attentive model (Semi-Attention) and its counterpart, supervised-only self-attentive model (S-Attention). The supervised-only model shares a similar architecture as the semi-supervised one with an exception that no sampled spatial neighbor/distant and temporal neighbor data are provided to constrain the unsupervised representation learning. Both Semi-Attention and S-Attention take four-dimensional data as input.

We standardize the input by subtracting from it per-channel mean and dividing it by

per-channel standard deviation. The standardization process is applied to each month separately as we observe that there are monthly variations in the mean and standard deviation of all the features. Monthly per-channel mean and per-channel standard deviation are obtained through using input data from the year 2003 to 2013.

4.4.4 Evaluation Approach and Metrics

To evaluate the generality of the baseline methods and our proposed approaches to unseen data in future years, we adopt a temporal nested validation approach. We conduct prediction experiments for 5 years between 2014 and 2018 independently. When a year y is selected to collect the test data, data from year $y-1$ are used for validation, while data collected from year $y-N_y-1$ to $y-2$ are used for training. N_y here can be used to control the size of training data and test the performance of a model under the stress of data scarcity.

We report Root Mean Square Error (RMSE) and R^2 as the evaluation metrics. Both RMSE and R^2 have been widely used to evaluate the crop yield prediction performances in previous studies [13, 34, 30]. RMSE measures the consistency of prediction results and ground-truth values. R^2 measures the fraction of variance in ground-truth values that can be explained by predictions. R^2 is not as scale-dependent as RMSE.

4.4.5 Hyperparameter Tuning

We tune the hyperparameters of the baseline methods and the proposed approaches based on the performance of the validation dataset. For all the deep models, including our self-attentive models, 50 epochs are run with the best model saved based on validation performance. Grid search from reasonable hyper-parameter combinations is adopted for LR, RF, and MLP.

Generally, we find our self-attentive models are easy to tune and less prone to overfitting. Weights for the unsupervised loss, temporal unsupervised loss, and regularization

(α, β, and γ as in Equation (4.2)) are set as 0.2, 0.001, 0.2, respectively. An exception is made for 2016 where we find through validation that it is more sensitive to unsupervised training part than other years. The unsupervised loss weight α is changed from 0.2 to 0.1. Radius to sample spatial neighbor and spatial distant are set as 25 and 100, respectively. As the time step granularity of our feature is relatively coarse, i.e., monthly, we fixed the temporal neighborhood as the region that appears at one-time step early. It can be easily adjusted for other tasks that have more sensitive time granularity though. We set both the dimensions of the spatial representation and the global spatiotemporal representation as 256. The inner-layer of the position-wise feed-forward networks has dimensionality as 512. A dropout with value 0.2 is applied to the temporal self-attentive layer as in [143]. For all the results of our proposed approaches reported in this chapter, the aforementioned hyperparameters are adopted.

4.4.6 Results of Model Performance

We first set N_y to 10 in comparing all the methods, which means 10 years of past data are used for training. This size of training data has been adopted widely in previous studies [13, 37]. Table 4.1 and Table 4.2 show the empirical results for the comparison with baselines in terms of RMSE and R^2 respectively. It can be seen that our approaches consistently outperform all the baseline methods with significant margins. A 13.5% improvement in terms of RMSE and 16.5% improvement in terms of R^2 can be seen when comparing the semi-supervised model with the best-performing baseline C3D. In comparison with the supervised-only model, a 2.9% and 2.1% improvements in terms of RMSE and R^2 are observed. It is worth noting that S-Attention outperforms CNN-LSTM with 14.5% improvement in RMSE and 23.8% improvement in R^2. This demonstrates the advantage of self-attentive model over LSTM in capturing temporal patterns on this crop yield prediction task.

Table 4.1: RMSE comparison of various methods when 10-year data are used for training.

Method	Year					Avg
	2014	2015	2016	2017	2018	
LR	6.465	7.754	7.589	6.839	8.163	7.362
RF	5.332	6.69	8.134	6.352	7.692	6.840
MLP	5.236	6.076	6.752	6.025	8.242	6.466
LSTM	5.595	6.450	8.389	6.890	5.655	6.596
LSTM+GP	5.013	5.553	6.761	5.134	5.522	5.597
CNN	4.833	5.625	8.157	5.302	6.318	6.047
CNN+GP	4.824	5.540	8.136	5.706	6.235	6.088
C3D	4.969	5.891	5.462	5.736	5.558	5.523
CNN-LSTM	4.948	5.698	6.688	5.596	5.83	5.752
S-Attention	4.762	5.336	**4.308**	4.741	5.452	4.920
Semi-Attention	**4.579**	**5.174**	4.548	**4.464**	**5.121**	**4.777**

One thing that deserves to be mentioned is that all the baseline methods perform poorly in the year 2016, i.e., the RMSE values are significantly higher and R^2 values are significantly lower than other years. One potential reason for this is that the year 2016 saw a record high yield across nearly all the midwestern states [151]. The information extracted through the baseline methods fails to catch the causes that lead to such a disparate pattern. Our temporal self-attentive models, including the semi-supervised one and the supervised-only one, instead, are able to extract necessary signals from past data to generalize well in this case.

Another interesting observation is that the Gaussian process does not bring much improvement compared with its counterpart as in [13], especially for the CNNs model. It might be due to the geographically localized factors, such as elevation and soil properties, can capture part of the spatial correlation existing in the real values and thus leave less work to do for the Gaussian process.

To further test the performances of all the approaches under the stress of label scarcity, we decrease the number of years to be used as the training data from 10 to 7, then 3, e.g.,

Table 4.2: R^2 comparison of various methods when 10-year data are used for training.

Method	Year					Avg
	2014	2015	2016	2017	2018	
LR	0.495	0.231	0.002	0.458	0.361	0.309
RF	0.657	0.428	-0.15	0.532	0.432	0.380
MLP	0.669	0.528	0.209	0.579	0.347	0.466
LSTM	0.622	0.467	-0.220	0.450	0.693	0.402
LSTM+GP	0.696	0.605	0.207	0.694	0.707	0.582
CNN	0.717	0.595	-0.154	0.674	0.617	0.490
CNN+GP	0.719	0.607	-0.148	0.622	0.627	0.485
C3D	0.7	0.556	0.481	0.618	0.704	0.612
CNN-LSTM	0.701	0.582	0.224	0.637	0.674	0.564
S-Attention	0.724	0.634	**0.678**	0.739	0.714	0.698
Semi-Attention	**0.746**	**0.658**	0.641	**0.769**	**0.748**	**0.712**

when $N_y = 3$, there are only 3 years of data used in training, 1 year for validation, and 1 year for test. This is a relatively extreme case but can shed some light on practical cases where input labeled data are less available or expensive to obtain, such as in developing counties or when a manual survey has to be adopted to collect the data.

Figure 4.4 shows the results averaged from 2014 to 2018 with the x-axis representing the number of labeled training instances. To have better readability, we separate the comparisons with the baseline methods based on the dimensionality of the input data. It can be seen that our models again outperform all the baseline methods consistently. There are two observations that we would like to mention here. First is that the advantage of the semi-supervised model over the supervised-only model becomes more obvious when the size of training data decreases. The improvements in terms of RMSE/R^2 are increased to 3.17%/3.75% when 7 years of data are used for training and 8.55%/14.39% when 3 years of data are used for training. The second observation is that the performance gap between our models which take full advantage of the data and the models with dimensionality reduction techniques, shrinks, while still remains, when less data are used for training.

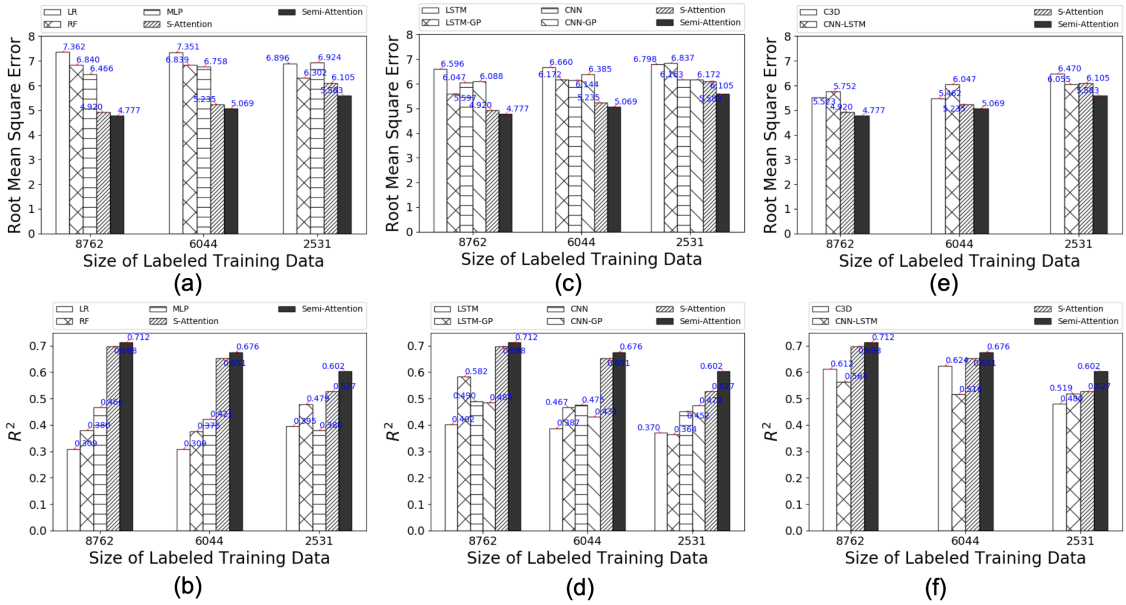

Figure 4.4: Model performance with varying numbers of labeled training instances. (a)-(b) Comparison between self-attentive method and conventional machine learning methods. (c)-(d) Comparison between self-attentive models and deep models (with three-dimensional input). (e)-(f) Comparison between self-attentive models and deep models (with four-dimensional input). All numbers are averaged from 2014 to 2018.

However, we still see a 16.0% R^2 improvement comparing our semi-supervised model with the best-performing baseline model CNN-LSTM when 3 years of labeled data are used for training.

To better visualize the trends of the model performances with the decreasing of training data, we plot the averaged RMSE and R^2 from 2014 and 2018 for our self-attentive models and the top four baselines, C3D, CNN-LSTM, LSTM+GP and CNN, as shown in Figure 4.5. Comparing Semi-Attention with S-Attention, it can be found that the semi-supervised model has a smaller performance drop when less training data is used. While CNN shows the slowest performance drop, there is still big metrics gap between it and our semi-supervised model.

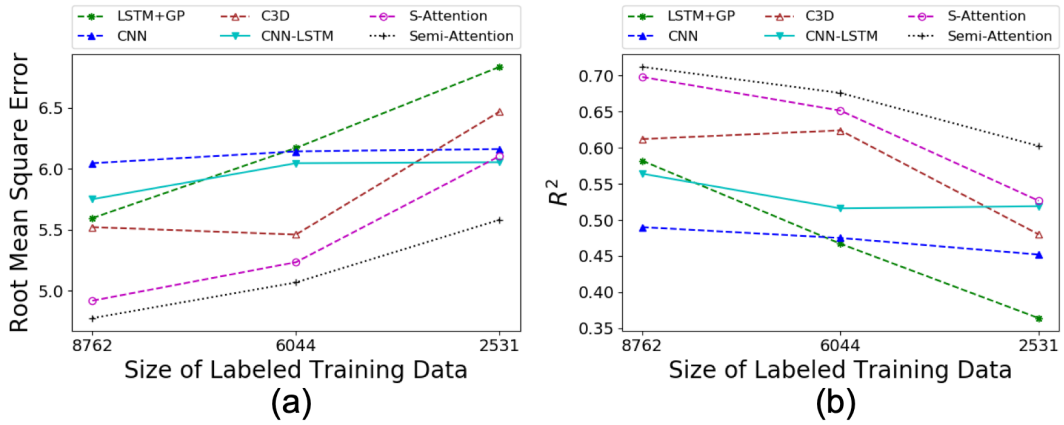

Figure 4.5: The trend of model performance with varying numbers of labeled training instances. (a) RMSE and (b) R^2. All numbers are averaged from 2014 to 2018.

4.4.7 Cross-border Model Generalization

In this section, we study how the semi-supervised model trained using input and crop yield data from one location performs for prediction tasks using input data from another location. This is to examine the generalization ability of the model across borders and is especially crucial for many computational sustainability tasks, of which the collection of ground-truth labels are expensive [152]. Here we first divide data into small groups based on the state of the location. Then data from different states are grouped based on location proximity. Specifically, we have group 1 for states North Dakota, South Dakota, Minnesota, and Wisconsin, group 2 for Nebraska, Kansas, Iowa, and Missouri, and group 3 for Illinois, Indiana, Ohio, and Kentucky. Group 1 has around 200 data records per year while group 2 and group 3 have around 300 data records per year. Figure 4.6(a) and 4.6(b) show the averaged results from 2014 to 2018 when 10-year training data are used. It can be found from the figure that the best model for one group is trained using data from that group, as would be expected. When considering cross-border predictions, it shows that model trained using group 2 or group 3 data can achieve reasonable accuracy

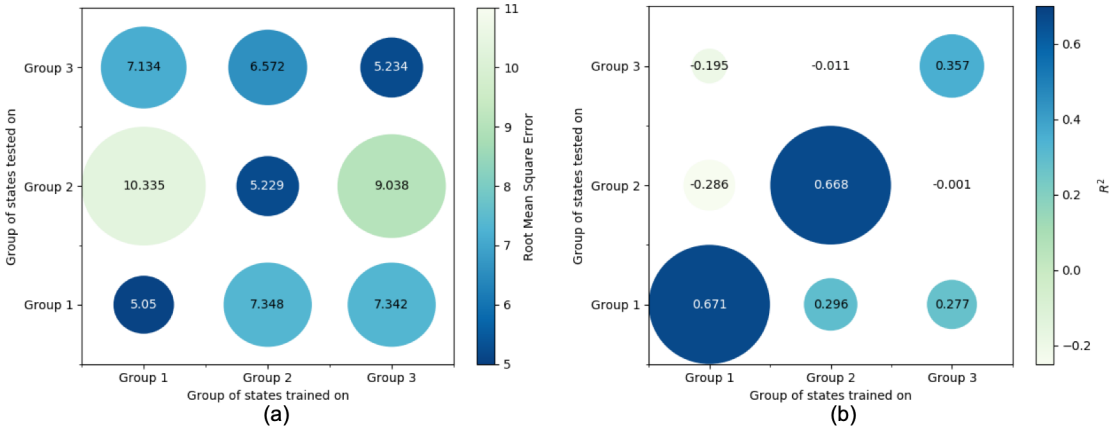

Figure 4.6: Cross-border model generalization. (a) RMSE and (b) R^2 values for semi-supervised model trained in one group of states and applied in another group of states. States are grouped based on location proximity. 10-year training data are used. All numbers are averaged from 2014 to 2018.

for prediction task on group 1 input. It also indicates that this cross-border model generalization ability needs to be further checked and verified for specific tasks as relatively low R^2 values can be found for some of the groups, e.g. model trained on data of group 1 and tested with input from group 2.

4.4.8 In-Season Crop Yield Prediction

Early in-season crop yield prediction benefits farmers, policy makers but poses extra challenges to the model than the yield prediction task right before harvesting. We evaluate in this section the ability of our semi-supervised model to predict soybean yield using input data collected in early growing months. In particular, we collect monthly data starting from March and ending in an in-season month between May and September. Experiments are conducted for each collection of data separately resulting in an online-learning fashion. Figure 4.7 shows the results averaged from 2014 to 2018 when 10-year training data are used. Again, we separate the comparisons with baseline methods based on the input data dimensionality. It is shown that both the self-attentive models,

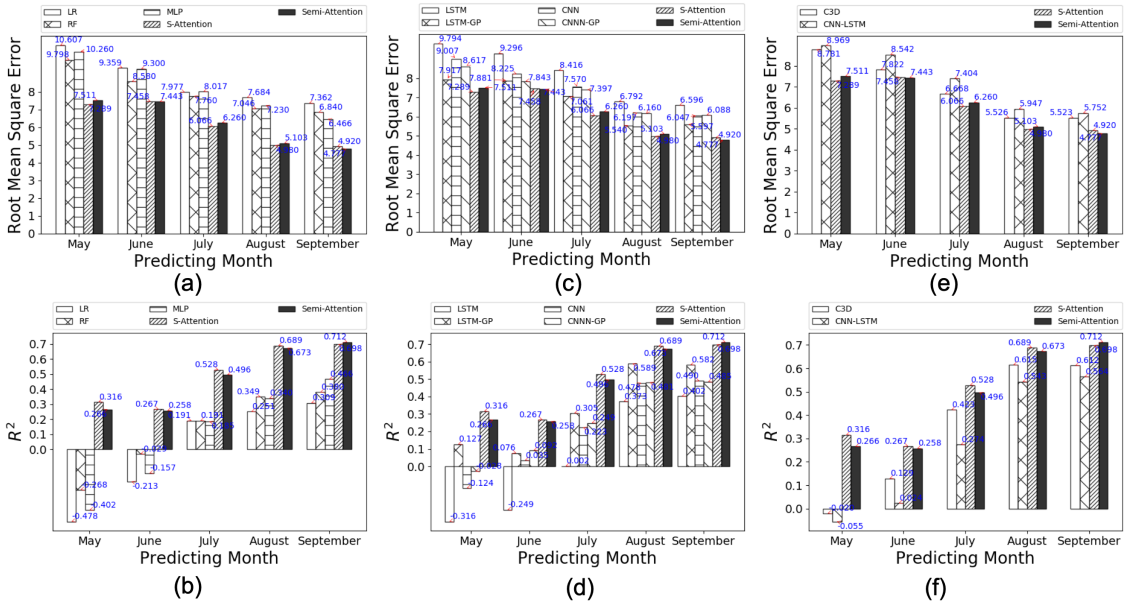

Figure 4.7: Model performance measured in in-season months. (a)-(b) Comparison between self-attentive method and conventional machine learning methods. (c)-(d) Comparison between self-attentive models and deep models (with three-dimensional input). (e)-(f) Comparison between self-attentive models and deep models (with four-dimensional input). All numbers are averaged from 2014 to 2018.

S-Attention and Semi-Attention, perform consistently better than the baseline methods. The predictions produced by the self-attentive methods in May achieve comparable or better performance compared to the predictions of baseline methods in July. The advantage of the semi-supervised model over the supervised-only model disappears for predictions between May and August though. One potential reason for this may be that the hyperparameter combination tuned for prediction in September needs to be modified for in-season predictions in earlier months. Another observation we have is that there is one performance jump happening in predictions in July compared to predictions in May and June. This might indicate that July is the month where more meaningful signals start to be contained in the input data, from which more accurate predictions can be produced.

4.4.9 Feature Importance Analysis

To understand how our semi-supervised model exploits the geo-spatiotemporal data fused from various sources, we provide a feature importance analysis here by excluding one type of environmental factors at a time from the input. The importance of a feature or a set of features can be demonstrated through the performance drop when they are not included in the input for training. The average performances from 2014 to 2018 of the semi-supervised model for both early in-season prediction (from May to August) and right-before harvesting prediction (September) is shown in Figure 4.8. Ten years of training data are used here. Specifically, we group the environmental features into six groups. Group VI is for the input when two vegetation indices are excluded. Group LST is for the land surface temperatures during the day and night. Group PPT, SOIL, and ELE are for the precipitation, soil properties and elevation, respectively. Group CLIMATE only uses two vegetation indices as input while all other climate variables are removed. We add group ALL for comparison which uses all available factors. For predictions in August and September, it can be seen from the figure that all groups see performance drops to different extents, which means that our model is capturing information from all environmental factors. Another interesting observation is that our model can still achieve satisfying performance when the two vegetation indices are not included and only climate data are used (group VI). It has been reported in previous studies [38, 30] that either vegetation indices or the source of their data, i.e., the multi-spectral remote sensing images, are necessary to achieve a satisfying crop yield prediction performance. Our self-attentive semi-supervised model, however, is able to achieve comparable performance solely using the climate data. This achievement is established because our model is more capable of extracting spatiotemporal signals in the climate data for producing accurate predictions.

Figure 4.8(c) and 4.8(d) show the RMSE and R^2 changes of each group compared

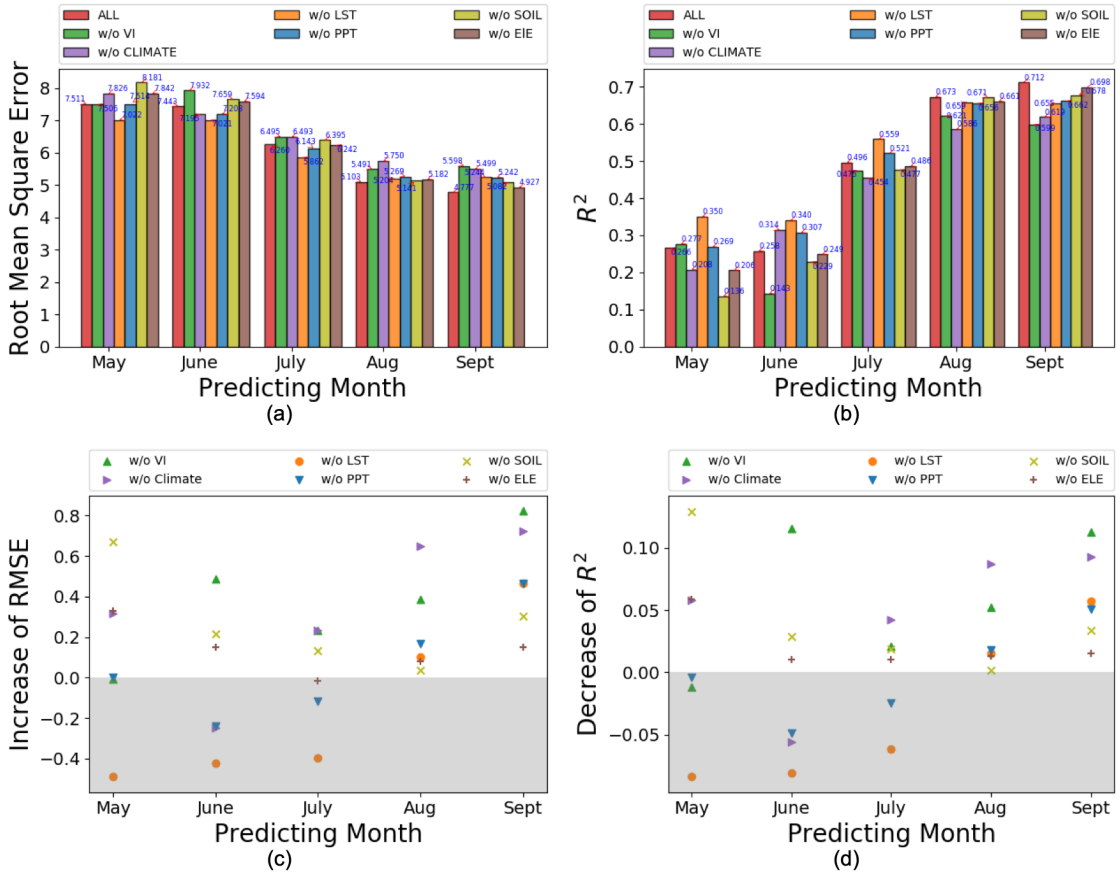

Figure 4.8: (a) RMSE and (b) R^2 values when one type of environmental factor is excluded from the input. The ALL group uses all available factors. (c) Increase of RMSE and (d) decrease of R^2 comparing each group with group ALL. All numbers are averaged from 2014 to 2018.

to group ALL. It can be seen from the figures that vegetation indices start to play a role in model prediction as early as in June, while precipitation and land surface temperature start to bring positive impact on model performance from August. Removing precipitation and land surface temperature from the input result in performance improvement in May, June, and July. This is consistent to researches in agricultural field [35, 41] that crop growing processes in late grain-filling period are more sensitive to environmental stress. Another interesting observation is that static factors like soil and elevation show much

importance for predictions in May and June when the soybeans are still in early-growing stages. Their importances drop, compared to other factors, when the soybeans are close to maturity, especially in August and September.

4.5 Conclusion

Machine learning approaches that exploit geo-spatiotemporal data have played a crucial role in many environmental sustainability applications. Unfortunately, the data heterogeneity and label scarcity often pose fundamental challenges to such applications. To tackle these challenges, we propose a novel semi-supervised self-attentive model. This model takes full advantage of the multi-channel spatiotemporal data and is able to learn global spatiotemporal representations for downstream prediction tasks. To overcome the limitation in label scarcity, we introduce unsupervised representation learning where spatiotemporal context sampled from unlabeled data are utilized. Our model jointly minimizes the unsupervised loss along with the supervised loss for the learning tasks. To evaluate the effectiveness of the proposed methods, we compare its performance with many state-of-the-art baselines in a regression task over large-scale real-world data. The experimental results clearly demonstrate the advantages of the proposed methods. We observe that the advantages become more obvious when less data are utilized in the training phase. To verify the impacts of this unsupervised loss, we compare the performance of the model with vs. without the unsupervised part. The results justify the positive impacts of the unsupervised loss in improving the overall performance of the proposed method.

Although the experiments mainly focus on regression tasks, it is worth emphasizing that the proposed framework generally works for other environmental sustainability applications, such as land cover classification or health index prediction. Since the inputs of our model only require the raw features of instances, the timestamps and positions the instances represent, and the labels of the instances. There are no restrictions on the type

of the labels to be categorical or numerical.

One promising direction for future work is to work on the convolutional neural networks part of our proposed model, recognizing that the conflated geographic data coming from different sources have varying spatial resolutions and characteristics, such as in spatial correlation, noise, and accuracy. Relying on one unifying convolutional neural networks with fixed-size filters, while as simple and scalable as it is, might not be able to extract the full spatial variations from the data. The same idea can also be applied to the temporal model part of the proposed model. Different environmental factors show different temporal patterns which ultimately have varying impacts on the target labels. Mining more spatial and temporal variation signals from the data would aid producing more accurate predictions, and is worth exploring in the future.

5. SUMMARY AND FUTURE WORK

We present in this chapter a summary of the contributions and highlight future directions for continued research.

5.1 Summary

Recent years see a proliferation of studies solving environmental sustainability problems through machine learning techniques applying on geo-spatiotemporal data. While this is still an emerging topic, this dissertation attempts to tackle two specific tasks driven by different observations and challenges. The proposed algorithms, however, can generally be applied to other computational environmental sustainability tasks as well.

We first address a challenging task falling into the hydrology field, the gap filling of high-resolution soil moisture problem. Different from existing studies that test the generalization ability of machine learning models at locations with the same hydroclimate conditions, the gap filling task examines how machine learning models learned at a location would perform at spatiotemporal neighboring locations with similar hydroclimate conditions. Traditional machine learning methods that learn relationship between coarse- and fine-resolution soil moisture, with the aid of auxiliary environmental factors, cannot produce satisfactory predictions in this case. We propose in this dissertation a two-layer machine learning framework, whose design is motivated by the soil moisture retrieval algorithm. The input features are also decided with the guidance of the retrieval algorithm. We find through extensive experiments that the proposed two-layer machine learning framework outperforms existing machine learning methods. The predictions generated are more consistent with the available high-resolution soil moisture product. Additional validations against monitoring networks data and airborne data are also satisfactory.

We continue investigating the crop yield prediction task, which is of great importance

especially with the increasing food demand and population growth. For each ground-truth label at a geographical unit, multi-channel geo-spatiotemporal data are taken as input. The design of this feature space is common in many computational sustainability tasks, such as wildfire prediction in field-scale and economic index for counties. Most existing machine learning methods however do not handle the spatiotemporal variations existing in the input features, which results in unnecessary information loss. Moreover, the lack of ground-truth labels hampers the performance of many machine learning methods, especially of the deep learning models. We propose in this dissertation a semi-supervised self-attentive model. It extracts spatiotemporal variations of the input data through the combination of convolution neural networks and the temporal self-attentive model. To alleviate the issue of training data scarcity, we introduce the spatiotemporal context signals that naturally reside in the largely-available unlabelled data. We train the models with the unsupervised loss and the supervised prediction loss to produce context-aware representations that adapt to the prediction task at the same time. We conduct experiments on a large-scale real-world crop yield prediction task. The results show that our semi-supervised self-attentive model outperforms existing state-of-the-art yield prediction methods and its counterpart, the supervised-only self-attentive model, especially under the stress of training data scarcity.

5.2 Future Work

We present here two promising directions for future research:

- The semi-supervised self-attentive model proposed in the third chapter processes multi-channel spatial images at each time step using the same convolution neural networks. While this is a simpler way to preprocess the data, there are two potential issues. First, spatial images of all environmental factors are required to be rescaled to the same resolution. This might incur information loss for spatial im-

ages which originally have higher resolutions. Second, fixed-size filters are applied to each channel of the input to extract spatial information from the data. This is less appropriate for conflated geographical data that have different spatial correlations. Instead, multiple convolutional neural networks can be utilized to process different types of input environmental factors separately. Moreover, a hierarchical self-attentive model can be adopted. First, a lower-layer self-attentive model is used to enable knowledge transferring among different types of environmental variables, followed by a second-layer temporal self-attentive model for knowledge transferring among different time steps. This architecture however is more complex than the semi-supervised self-attentive model proposed in this dissertation and requires more training data for model learning.

- The second direction is to improve prediction accuracy by investigating temporal variations of the environmental factors. Different environmental factors might show different patterns of temporal variations which consecutively play different roles determining the values of the target variable. This observation might not help the yearly crop yield prediction task but is crucial for real-time prediction tasks that follow a streaming mode, such as the real-time water flow prediction problem [153]. A similar observation is made in [3] for the land cover prediction task that both long-term and short-term temporal variation patterns exist in the input features. A dual-memory LSTM model is thus proposed to address this issue. Their model is however not flexible and scalable when taking multiple environmental factors as input. Another interesting problem that worth to be solved for the real-time prediction task is to take advantage of newly available data to augment the performance of the machine learning models.

REFERENCES

[1] S. I. Seneviratne, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Science Reviews*, vol. 99, no. 3-4, pp. 125–161, 2010.

[2] C. Justice, J. Townshend, E. Vermote, E. Masuoka, R. Wolfe, N. Saleous, D. Roy, and J. Morisette, "An overview of modis land data processing and product status," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 3–15, 2002.

[3] X. Jia, A. Khandelwal, G. Nayak, J. Gerber, K. Carlson, P. West, and V. Kumar, "Incremental dual-memory lstm in land cover prediction," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 867–876, 2017.

[4] S. G. Subramanian and M. Crowley, "Learning forest wildfire dynamics from satellite images using reinforcement learning," in *Conference on Reinforcement Learning and Decision Making*, 2017.

[5] X. He, N. W. Chaney, M. Schleiss, and J. Sheffield, "Spatial downscaling of precipitation using adaptable random forests," *Water Resources Research*, vol. 52, pp. 8217–8237, oct 2016.

[6] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, *et al.*, "Tackling climate change with machine learning," *arXiv preprint arXiv:1906.05433*, 2019.

[7] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[8] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*, pp. 160–167, 2008.

[11] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–38, 2019.

[12] A. Karpatne, I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar, "Machine learning for the geosciences: Challenges and opportunities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 8, pp. 1544–1554, 2018.

[13] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[14] Y. Du, W. Song, Q. He, D. Huang, A. Liotta, and C. Su, "Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection," *Information Fusion*, vol. 49, pp. 89–99, 2019.

[15] N. Jean, S. Wang, A. Samar, G. Azzari, D. Lobell, and S. Ermon, "Tile2vec: Unsupervised representation learning for spatially distributed data," in *Thirty-Third AAAI Conference on Artificial Intelligence*, pp. 3967–3974, 2019.

[16] J. D. Bolten, W. T. Crow, X. Zhan, T. J. Jackson, and C. A. Reynolds, "Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural

drought monitoring," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, pp. 57–66, mar 2010.

[17] W. Cai, T. Cowan, P. Briggs, and M. Raupach, "Rising temperature depletes soil moisture and exacerbates severe drought conditions across southeast australia," *Geophysical Research Letters*, vol. 36, nov 2009.

[18] M. Drusch, "Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *Journal of Geophysical Research*, vol. 112, feb 2007.

[19] E. G. Njoku and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *Journal of Hydrology*, vol. 184, pp. 101–129, oct 1996.

[20] E. G. Njoku, W. J. Wilson, S. H. Yueh, S. J. Dinardo, F. K. Li, T. Jackson, V. Lakshmi, and J. Bolten, "Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 2659–2673, dec 2002.

[21] N. N. Das, D. Entekhabi, E. G. Njoku, J. J. C. Shi, J. T. Johnson, and A. Colliander, "Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 2018–2028, apr 2014.

[22] T. Jagdhuber, M. Baur, R. Akbar, N. N. Das, M. Link, L. He, and D. Entekhabi, "Estimation of active-passive microwave covariation using SMAP and sentinel-1 data," *Remote Sensing of Environment*, vol. 225, pp. 458–468, may 2019.

[23] J. Dong and W. T. Crow, "The added value of assimilating remotely sensed soil moisture for estimating summertime soil moisture-air temperature coupling strength," *Water Resources Research*, jul 2018.

[24] S. Chai, J. Walker, B. Veenendaal, and G. West, "An artificial neural network model for downscaling of passive microwave soil moisture," *Recent Researches in Hydrology, Geology and Continuum Mechanics*, 2011.

[25] P. K. Srivastava, D. Han, M. R. Ramirez, and T. Islam, "Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application," *Water Resources Management*, vol. 27, pp. 3127–3144, jun 2013.

[26] P. Abbaszadeh, H. Moradkhani, and X. Zhan, "Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method," *Water Resources Research*, dec 2018.

[27] N. Gaur and B. P. Mohanty, "Land-surface controls on near-surface soil moisture dynamics: Traversing remote sensing footprints," *Water Resources Research*, vol. 52, pp. 6365–6385, aug 2016.

[28] T. Wang, T. E. Franz, R. Li, J. You, M. D. Shulski, and C. Ray, "Evaluating climate and soil effects on regional soil moisture spatial variability using EOFs," *Water Resources Research*, vol. 53, pp. 4022–4035, may 2017.

[29] D. Entekhabi, N. Das, E. Njoku, J. Johnson, and J. Shi, "Algorithm theoretical basis document l2 & l3 radar/radiometer soil moisture (active/passive) data products," *Jet Propulsion Lab. Report*, vol. 1, 2014.

[30] H. Jiang, H. Hu, R. Zhong, J. Xu, J. Xu, J. Huang, S. Wang, Y. Ying, and T. Lin, "A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US corn belt at the county level.," *Global Change Biology*, 2019.

[31] R. Bongiovanni and J. Lowenberg-Deboer, "Precision agriculture and sustainability," *Precision Agriculture*, vol. 5, no. 4, pp. 359–387, 2004.

[32] M. Parry, C. Rosenzweig, and M. Livermore, "Climate change, global food supply and risk of hunger.," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 360, no. 1463, pp. 2125–2138, 2005.

[33] H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin, "Food security: the challenge of feeding 9 billion people.," *Science*, vol. 327, no. 5967, pp. 812–818, 2010.

[34] A. Mateo-Sanchis, M. Piles, J. Muñoz-Marí, J. E. Adsuara, A. Pérez-Suay, and G. Camps-Valls, "Synergistic integration of optical and microwave satellite data for crop yield estimation," *Remote Sensing of Environment*, vol. 234, p. 111460, 2019.

[35] K. Guan, J. Wu, J. S. Kimball, M. C. Anderson, S. Frolking, B. Li, C. R. Hain, and D. B. Lobell, "The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields," *Remote Sensing of Environment*, vol. 199, pp. 333–349, 2017.

[36] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model.," *Sensors*, vol. 19, no. 20, 2019.

[37] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction," *Frontiers in Plant Science*, vol. 10, 2020.

[38] B. Peng, K. Guan, M. Pan, and Y. Li, "Benefits of seasonal climate prediction and satellite data for forecasting u.s. maize yield," *Geophysical Research Letters*, vol. 45, no. 18, pp. 9662–9671, 2018.

[39] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 163, p. 104859, 2019.

[40] Q. Yang, L. Shi, J. Han, Y. Zha, and P. Zhu, "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images," *Field Crops Research*, vol. 235, pp. 142–153, 2019.

[41] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," *Remote Sensing of Environment*, vol. 237, p. 111599, 2020.

[42] D.-J. Seo, "Real-time estimation of rainfall fields using radar rainfall and rain gage data," *Journal of Hydrology*, vol. 208, no. 1-2, pp. 37–52, 1998.

[43] D. L. Williams, S. Goward, and T. Arvidson, "Landsat," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 10, pp. 1171–1178, 2006.

[44] Z. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from lidar remote sensing data," *Landscape and Urban Planning*, vol. 130, pp. 134–148, 2014.

[45] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[46] A. Perez, C. Yeh, G. Azzari, M. Burke, D. Lobell, and S. Ermon, "Poverty prediction with public landsat 7 satellite imagery and machine learning," *arXiv preprint arXiv:1711.03654*, 2017.

[47] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of Sensors*, vol. 2017, 2017.

[48] C. Vittucci, L. Guerriero, P. Ferrazzoli, R. Rahmoune, M. Tanase, R. Panciera, A. Monerris, C. Rüdiger, and J. P. Walker, "Airborne forest monitoring during SMAPEx-3 campaign," in *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, pp. 987–990, IEEE, 2013.

[49] C. Daly, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states," *International Journal of Climatology: a Journal of the Royal Meteorological Society*, vol. 28, no. 15, pp. 2031–2064, 2008.

[50] C. Daly, D. R. Conklin, and M. H. Unsworth, "Local atmospheric decoupling in complex topography alters climate change impacts," *International Journal of Climatology*, vol. 30, no. 12, pp. 1857–1864, 2010.

[51] C. Gomes, T. Dietterich, C. Barrett, J. Conrad, B. Dilkina, S. Ermon, F. Fang, A. Farnsworth, A. Fern, X. Fern, *et al.*, "Computational sustainability: Computing for a better world and a sustainable future," *Communications of the ACM*, vol. 62, no. 9, pp. 56–65, 2019.

[52] F. A. Graybill, *Theory and application of the linear model*, vol. 183. Duxbury press North Scituate, MA, 1976.

[53] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, 2019.

[54] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[55] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[56] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[57] X. Jia, G. Nayak, A. Khandelwal, A. Karpatne, and V. Kumar, "Classifying heterogeneous sequential data by cyclic domain adaptation: An application in land

cover detection," in *Proceedings of the 2019 SIAM International Conference on Data Mining*, pp. 540–548, SIAM, 2019.

[58] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, pp. 802–810, 2015.

[59] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2012.

[60] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.

[61] M. A. Bhimra, U. Nazir, and M. Taj, "Using 3D residual network for spatiotemporal analysis of remote sensing data," in *ICA 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICA)*, pp. 1403–1407, IEEE, may 2019.

[62] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.

[63] K. Fang, C. Shen, D. Kifer, and X. Yang, "Prolongation of SMAP to spatiotemporally seamless coverage of continental u.s. using a deep learning neural network," *Geophysical Research Letters*, vol. 44, pp. 11,030–11,039, nov 2017.

[64] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[65] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.

[66] H. Russello, "Convolutional neural networks for crop yield prediction using satellite images," *IBM Center for Advanced Studies*, 2018.

[67] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, pp. 3111–3119, 2013.

[68] J. Pennington, R. Socher, and C. Manning, "Glove: global vectors for word representation," (Stroudsburg, PA, USA), pp. 1532–1543, Association for Computational Linguistics, 2014.

[69] S. I. Seneviratne, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Science Reviews*, vol. 99, pp. 125–161, may 2010.

[70] J. Pastor and W. M. Post, "Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles," *Biogeochemistry*, vol. 2, pp. 3–27, mar 1986.

[71] N. Wanders, D. Karssenberg, A. de Roo, S. M. de Jong, and M. F. P. Bierkens, "The suitability of remotely sensed soil moisture for improving operational flood forecasting," *Hydrology and Earth System Sciences*, vol. 18, pp. 2343–2357, jun 2014.

[72] H. Bogena, J. Huisman, C. Oberdörster, and H. Vereecken, "Evaluation of a low-cost soil water content sensor for wireless network applications," *Journal of Hydrology*, vol. 344, pp. 32–42, sep 2007.

[73] D. A. Robinson, S. B. Jones, J. M. Wraith, D. Or, and S. P. Friedman, "A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry," *Vadose Zone Journal*, vol. 2, no. 4, p. 444, 2003.

[74] A. Robock, K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, "The global soil moisture data bank," *Bulletin of the American Meteorological Society*, vol. 81, no. 6, pp. 1281–1299, 2000.

[75] W. Wagner, S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa-Saldaña, P. de Rosnay, A. Jann, S. Schneider, J. Komma, G. Kubu, K. Brugger, C. Aubrecht, J. Züger, U. Gangkofner, S. Kienberger, L. Brocca, Y. Wang, G. Blöschl, J. Eitzinger, and K. Steinnocher, "The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications," *Meteorologische Zeitschrift*, vol. 22, pp. 5–33, feb 2013.

[76] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from AMSR-e," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 215–229, feb 2003.

[77] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martín-Neira, and S. Mecklenburg, "The SMOS mission: New tool for monitoring key elements ofthe global water cycle," *Proceedings of the IEEE*, vol. 98, pp. 666–687, may 2010.

[78] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. Van Zyl, "The

soil moisture active passive (SMAP) mission," *Proceedings of the IEEE*, vol. 98, pp. 704–716, may 2010.

[79] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan, "GMES sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, pp. 9–24, may 2012.

[80] A. AghaKouchak, A. Farahmand, F. S. Melton, J. Teixeira, M. C. Anderson, B. D. Wardlow, and C. R. Hain, "Remote sensing of drought: Progress, challenges and opportunities," *Reviews of Geophysics*, vol. 53, pp. 452–480, jun 2015.

[81] R. D. Koster, L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy, "Precipitation estimation using l-band and c-band soil moisture retrievals.," *Water Resources Research*, vol. 52, pp. 7213–7225, sep 2016.

[82] S. Tian, P. Tregoning, L. J. Renzullo, A. I. J. M. van Dijk, J. P. Walker, V. R. N. Pauwels, and S. Allgeyer, "Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals," *Water Resources Research*, vol. 53, pp. 1820–1840, mar 2017.

[83] F. T. Ulaby, P. C. Dubois, and J. van Zyl, "Radar mapping of surface soil moisture," *Journal of Hydrology*, vol. 184, pp. 57–84, oct 1996.

[84] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish, "SMAP L3 radiometer global daily 36 km EASE-grid soil moisture, version 5," 2018a.

[85] Y. Lu, S. C. Steele-Dunne, L. Farhadi, and N. van de Giesen, "Mapping surface heat fluxes by assimilating SMAP soil moisture and GOES land surface temperature data," *Water Resources Research*, dec 2017.

[86] R. C. Nijzink, S. Almeida, I. G. Pechlivanidis, R. Capell, D. Gustafssons, B. Arheimer, J. Parajka, J. Freer, D. Han, T. Wagener, R. R. P. Nooijen, H. H. G. Savenije, and M. Hrachowitz, "Constraining conceptual hydrological models with multiple information sources," *Water Resources Research*, vol. 54, pp. 8332–8362, oct 2018.

[87] J. Peng, A. Loew, O. Merlin, and N. E. C. Verhoest, "A review of spatial downscaling of satellite remotely sensed soil moisture," *Reviews of Geophysics*, vol. 55, pp. 341–366, jun 2017.

[88] S. Sabaghy, J. P. Walker, L. J. Renzullo, and T. J. Jackson, "Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities," *Remote Sensing of Environment*, vol. 209, pp. 551–580, may 2018.

[89] T. J. Jackson, "III. measuring surface soil moisture using passive microwave remote sensing," *Hydrological processes*, vol. 7, pp. 139–152, apr 1993.

[90] D. Entekhabi, N. Das, E. Njoku, J. Johnson, and J. Shi, "SMAP L3 radar/radiometer global daily 9 km EASE-grid soil moisture, version 3," 2016.

[91] N. N. Das, D. Entekhabi, S. Kim, S. Yueh, and P. O'Neill, "Combining SMAP and sentinel data for high-resolution soil moisture product," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 129–131, IEEE, jul 2016.

[92] S. Yueh, D. Entekhabi, P. O'Neill, E. Njoku, and J. Entin, "NASA soil moisture active passive mission status and science performance," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 116–119, IEEE, jul 2016.

[93] H. Lievens, R. H. Reichle, Q. Liu, G. J. M. De Lannoy, R. S. Dunbar, S. B. Kim, N. N. Das, M. Cosh, J. P. Walker, and W. Wagner, "Joint sentinel-1 and SMAP data

assimilation to improve soil moisture estimates.," *Geophysical Research Letters*, vol. 44, pp. 6145–6153, jun 2017.

[94] E. Santi, S. Paloscia, S. Pettinato, D. Entekhabi, S. H. Alemohammad, and A. G. Konings, "Integration of passive and active microwave data from SMAP, AMSR2 and sentinel-1 for soil moisture monitoring," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5252–5255, IEEE, jul 2016.

[95] C. Rudiger, C.-H. Su, D. Ryu, and W. Wagner, "Disaggregation of low-resolution l-band radiometry using c-band radar data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 1425–1429, oct 2016.

[96] N. N. Das, D. Entekhabi, R. S. Dunbar, S. Kim, S. Yueh, A. Colliander, P. E. O'Neill, and T. Jackson, "SMAP/Sentinel-1 L2 radiometer/radar 30-second scene 3 km EASE-grid soil moisture, version 2," 2018.

[97] European Space Agency, "Sentinel-1 User Handbook," tech. rep., 2013.

[98] J. P. Walker and P. R. Houser, "Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution," *Advances in Water Resources*, vol. 27, pp. 785–801, aug 2004.

[99] N. S. Chauhan, S. Miller, and P. Ardanuy, "Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach," *International Journal of Remote Sensing*, vol. 24, pp. 4599–4622, jan 2003.

[100] O. Merlin, C. Rudiger, A. Al Bitar, P. Richaume, J. P. Walker, and Y. H. Kerr, "Disaggregation of SMOS soil moisture in southeastern australia," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 1556–1571, may 2012.

[101] M. Piles, A. Camps, M. Vall-llossera, I. Corbella, R. Panciera, C. Rudiger, Y. H. Kerr, and J. Walker, "Downscaling SMOS-derived soil moisture using MODIS vis-

ible/infrared data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3156–3166, sep 2011.

[102] H. Lievens, S. Tomer, A. Al Bitar, G. De Lannoy, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y. Kerr, B. Martens, M. Pan, J. Roundy, H. Vereecken, J. Walker, E. Wood, N. Verhoest, and V. Pauwels, "SMOS soil moisture assimilation for improved hydrologic simulation in the murray darling basin, australia," *Remote Sensing of Environment*, vol. 168, pp. 146–162, oct 2015.

[103] R. H. Reichle, D. Entekhabi, and D. B. McLaughlin, "Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach," *Water Resources Research*, vol. 37, pp. 2353–2364, sep 2001.

[104] J. Im, S. Park, J. Rhee, J. Baik, and M. Choi, "Downscaling of AMSR-e soil moisture with MODIS products using machine learning approaches," *Environmental earth sciences*, vol. 75, p. 1120, aug 2016.

[105] J. Kolassa, R. H. Reichle, Q. Liu, S. H. Alemohammad, P. Gentine, K. Aida, J. Asanuma, S. Bircher, T. Caldwell, A. Colliander, M. Cosh, C. H. Collins, T. J. Jackson, J. Martínez-Fernández, H. McNairn, A. Pacheco, M. Thibeault, and J. P. Walker, "Estimating surface soil moisture from SMAP observations using a neural network technique.," *Remote Sensing of Environment*, vol. 204, pp. 43–59, jan 2018.

[106] P. Coulibaly, Y. B. Dibike, and F. Anctil, "Downscaling precipitation and temperature with temporal neural networks," *Journal of Hydrometeorology*, vol. 6, pp. 483–496, aug 2005.

[107] M. Basu and T. K. Ho, eds., *Data Complexity in Pattern Recognition*. Springer Science & Business Media, illustrated ed., 2006.

[108] S. Park, J. Im, S. Park, and J. Rhee, "AMSR2 soil moisture downscaling using multisensor products through machine learning approach," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1984–1987, IEEE, jul 2015.

[109] W. Zhao, N. Sánchez, H. Lu, and A. Li, "A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression," *Journal of Hydrology*, vol. 563, pp. 1009–1024, aug 2018.

[110] P. E. O'Neill, S. Chan, E. Njoku, T. Jackson, and R. Bindlish, "Algorithm theoretical basis document L2 & L3 soil moisture (passive) data products," 2015.

[111] M. J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, "EASE-grid 2.0: Incremental but significant improvements for earth-gridded data sets," *ISPRS International Journal of Geo-information*, vol. 1, pp. 32–45, mar 2012.

[112] G. L. Schaefer, M. H. Cosh, and T. J. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *Journal of Atmospheric and Oceanic Technology*, vol. 24, pp. 2073–2077, dec 2007.

[113] J. E. Bell, M. A. Palecki, C. B. Baker, W. G. Collins, J. H. Lawrimore, R. D. Leeper, M. E. Hall, J. Kochendorfer, T. P. Meyers, T. Wilson, and H. J. Diamond, "U.S. climate reference network soil moisture and temperature observations," *Journal of Hydrometeorology*, vol. 14, pp. 977–988, jun 2013.

[114] A. Colliander, T. J. Jackson, M. Cosh, S. Misra, R. Bindlish, J. Powers, H. McNairn, P. Bullock, A. Berg, R. Magagi, P. O'Neill, and S. Yueh, "Soil moisture retrieval with airborne PALS instrument over agricultural areas in SMAPVEX16," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3949–3952, IEEE, jul 2017.

[115] D. A. Miller and R. A. White, "A conterminous united states multi-layer soil characteristics data set for regional climate and hydrology modeling," *Earth Interact*, vol. 2, 1998.

[116] S. Kim, "SMAP Ancillary Data Report on Landcover Classification," 2013.

[117] NASA JPL, "NASA shuttle radar topography mission global 3 arc second," 2013.

[118] Myneni, R and Knyazikhin, Y and Park, T, "MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006," tech. rep., 2015.

[119] Wan, Z and Hook, S and Hulley, G, "MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006," tech. rep., 2015.

[120] A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi, "The global precipitation measurement mission," *Bulletin of the American Meteorological Society*, vol. 95, pp. 701–722, may 2014.

[121] Z. Wan and J. Dozier, "A generalized split-window algorithm for retrieving land-surface temperature from space," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 892–905, jul 1996.

[122] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the köppen-geiger climate classification," *Hydrology and Earth System Sciences Discussions*, vol. 4, pp. 439–473, mar 2007.

[123] D. Entekhabi, N. Das, E. Njoku, J. Johnson, and J. Shi, "Algorithm theoretical basis document L2 & L3 radar/radiometer soil moisture (active/passive) data products," 2012.

[124] N. Gaur and B. P. Mohanty, "A nomograph to incorporate geo-physical heterogeneity in soil moisture downscaling," *Water Resources Research*, oct 2018.

[125] A. Loew and W. Mauser, "On the disaggregation of passive microwave soil moisture data usinga priori knowledge of temporally persistent soil moisture fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, pp. 819–834, mar 2008.

[126] J. F. Artiola and A. W. Warrick, "Sampling and data quality objectives for environmental monitoring," in *Environmental Monitoring and Characterization*, pp. 11–27, Elsevier Inc., 2004.

[127] G. M. Lovett, D. A. Burns, C. T. Driscoll, J. C. Jenkins, M. J. Mitchell, L. Rustad, J. B. Shanley, G. E. Likens, and R. Haeuber, "Who needs environmental monitoring?," *Frontiers in Ecology and the Environment*, vol. 5, no. 5, pp. 253–260, 2007.

[128] Y. Kang, S. Khan, and X. Ma, "Climate change impacts on crop yield, crop water productivity and food security–a review," *Progress in Natural Science*, vol. 19, no. 12, pp. 1665–1674, 2009.

[129] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi, "Exploratory spatio-temporal data mining and visualization," *Journal of Visual Languages & Computing*, vol. 18, no. 3, pp. 255–279, 2007.

[130] H. Mao, D. Kathuria, N. Duffield, and B. P. Mohanty, "Gap filling of high-resolution soil moisture for SMAP/Sentinel-1: A two-layer machine learning-based framework," *Water Resources Research*, vol. 55, no. 8, pp. 6986–7009, 2019.

[131] V. Masson, J.-L. Champeaux, F. Chauvin, C. Meriguet, and R. Lacaze, "A global database of land surface parameters at 1-km resolution in meteorological and climate models," *Journal of Climate*, vol. 16, no. 9, pp. 1261–1282, 2003.

[132] S. Rahimi, M. Cobb, D. Ali, M. Paprzycki, and F. Petry, "A knowledge-based multi-agent system for geospatial data conflation," *Journal of Geographic Information and Decision Analysis*, vol. 6, no. 2, pp. 67–81, 2002.

[133] H. J. Miller, "The data avalanche is here," *Shouldn't we be*, 2010.

[134] X. J. Zhu, "Semi-supervised learning literature survey," tech. rep., University of Wisconsin-Madison Department of Computer Sciences, 2005.

[135] J. Lässig, K. Kersting, and K. Morik, *Computational Sustainability*, vol. 645. Springer, 2016.

[136] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, 2017.

[137] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634, 2015.

[138] A. W. Yu, D. Dohan, M.-T. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, "Qanet: Combining local convolution with global self-attention for reading comprehension," *arXiv preprint arXiv:1804.09541*, 2018.

[139] J. Zhao, B. Du, L. Sun, F. Zhuang, W. Lv, and H. Xiong, "Multiple relational attention network for multi-task learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1123–1131, 2019.

[140] X. Jia, S. Li, A. Khandelwal, G. Nayak, A. Karpatne, and V. Kumar, "Spatial context-aware networks for mining temporal discriminative period in land cover

detection," in *Proceedings of the 2019 SIAM International Conference on Data Mining*, pp. 513–521, SIAM, 2019.

[141] X. Liu, M. Xie, X. Wen, R. Chen, Y. Ge, N. Duffield, and N. Wang, "A semi-supervised and inductive embedding model for churn prediction of large-scale mobile games," in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 277–286, IEEE, 2018.

[142] V. Ramanathan, K. Tang, G. Mori, and L. Fei-Fei, "Learning temporal embeddings for complex video analysis," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4471–4479, 2015.

[143] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[144] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[145] Y. Lu, Z. Li, D. He, Z. Sun, B. Dong, T. Qin, L. Wang, and T.-Y. Liu, "Understanding and improving transformer from a multi-particle dynamic system point of view," *arXiv preprint arXiv:1906.02762*, 2019.

[146] D. M. Johnson, "An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the united states," *Remote Sensing of Environment*, vol. 141, pp. 116–128, 2014.

[147] U. NASS, "Quick stats database. Washington, DC: USDA national agricultural statistics service," 2017.

[148] K. Didan, "Mod13a3 modis/terra vegetation indices monthly l3 global 1km sin grid v006," *NASA EOSDIS Land Processes DAAC*, 2015.

[149] D. B. Lobell, M. J. Roberts, W. Schlenker, N. Braun, B. B. Little, R. M. Rejesus, and G. L. Hammer, "Greater sensitivity to drought accompanies maize yield increase in the us midwest," *Science*, vol. 344, no. 6183, pp. 516–519, 2014.

[150] J. NASA, "Nasa shuttle radar topography mission global 1 arc second," *NASA LP DAAC*, vol. 15, 2013.

[151] S. Irwin and D. Good, "Uncertainty about 2016 us average corn and soybean yields persists," *Farmdoc Daily*, vol. 6, no. 132, 2016.

[152] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[153] K. J. Franz, H. C. Hartmann, S. Sorooshian, and R. Bales, "Verification of national weather service ensemble streamflow predictions for water supply forecasting in the colorado river basin," *Journal of Hydrometeorology*, vol. 4, no. 6, pp. 1105–1118, 2003.